Evaluation analytique des performanes des réseaux sans-fil par un processus de Markov spatial prenant en compte leur géométrie, leur dynamique et leurs algorithmes de contrôle

Mohamed Kadhem Karray

École Nationale Supérieure des Télécommunications

École Doctorale d'Informatique, Télécommunications et Électronique

Thesis: Analytic evaluation of wireless cellular networks performance by a spatial Markov process accounting for their geometry, dynamics and control schemes

Informatique et Réseaux

Mohamed Kadhem KARRAY
Jury: Zwi Altman (HDR, France Telecom R&D) examiner
François Baccelli (Professor, INRIA)
Bartłomiej Błaszczyszyn (Doctor, INRIA) thesis supervisor
David MacDonald (Professor, University of Ottawa) examiner
Laurent Massoulié (Doctor, Thomson Research) examiner
–contresigned by Laurent Viennot (HDR, INRIA)–
Eric Moulines (Professor, ENST) thesis supervisor

Thesis defence: 10 septembre 2007

Contents

Preface

I wish to thank my family, my parents, my brothers and my sister, my wife and my daughter, and also my friends who support and encourage me throughout the duration of my thesis.

I am particularly honored to have been able to collaborate with professor François Baccelli and doctor Bartek Błaszczyszyn, from INRIA[1] & ENS[2], while writing my thesis; I hope to continue this fruitful collaboration. I wish to thank my supervisors professor Eric Moulines, from ENST[3] and doctor Bartek Błaszczyszyn, for their encouragement and advise. I am particularly grateful to professor Pierre Brémaud for his rich lectures on "Signal, Information and Communication" at ENS. I wish also to thank the referee and examiners for reading the thesis and their suggestions. I particularly thank professor David McDonald and doctor Laurent Massoulié who accept to review my report, doctor Zwi Altman for his encouragements throughout the duration of my thesis and mister Omar Soufit for his fruitful collaboration particularly in the simulations. Thanks also to my colleagues at France Telecom who supplied me with interesting problems and relevant information.

[1]Institut National de Recherche en Informatique et en Automatique
[2]Ecole Normale Supérieure
[3]Ecole Nationale Supérieure des Télécommunications

Short abstract: We build load control schemes for wireless cellular networks and develop analytic methods for the performance evaluation of these networks by a spatial Markov process accounting for their geometry, dynamics and control schemes. First, we characterize the single link performance by using the digital communication techniques. Then the interactions between the links are taken into account by formulating a power allocation problem. We propose decentralized load control schemes which take into account the influence of geometry on the combination of inter-cell and intra-cell interferences. In order to study the performance of these schemes, we analyze a pure-jump Markov generator that can be seen as a generalization of the spatial birth-and-death generator, which allows for mobility of particles. We give sufficient conditions for the regularity of the generator (i.e., uniqueness of the associated Markov process) as well as for its ergodicity. Finally we apply our spatial Markov process model to evaluate the performance of wireless cellular networks using the feasibility based load control schemes.

Keywords: wireless, cellular, performance, load control, power allocation feasibility, birth-and-death, mobility, regularity, ergodicity, invariant measure, Gibbs, spatial Erlang formula, blocking, cut probability, delay, throughput.

Titre (in frensh): Evaluation analytique des performances des réseaux cellulaires sans fil par un processus de Markov spatial prenant en compte leur géométrie, dynamique et algorithmes de contrôle

Cour résumé: Nous proposons des algorithmes de contrôle de charge pour les réseaux cellulaires sans fil et développons des méthodes analytiques pour l'évaluation des performances de ces réseaux par un processus de Markov spatial prenant en compte leur géométrie, dynamique et algorithmes de contrôle. D'abord, nous caractérisons la performnace d'un lien unique en utilisant les techniques de communication numérique. Ensuite les interactions entre les liens sont prises en compte en formulant un problème d'allocation de puissances. Nous proposons des algorithmes de contrôle de charge décentralisés qui tiennent compte de l'influence de la géométrie sur la combinaison des interférences inter-cellules et intra-cellules. Afin d'étudier les performances de ces algorithmes, nous analysons un générateur d'un processus Markovien de saut qui peut être vu comme une généralisation du générateur de naissance-et-mort spatial, qui tient compte de la mobilité des particules. Nous donnons des conditions suffisantes pour la régularité du générateur (c.-à-d., unicité du processus de Markov associé) aussi bien que pour son ergodicité. Enfin nous appliquons notre processus de Markov spatial pour évaluer les performances des réseaux cellulaires sans fil utilisant les algorithmes de contrôle de charge basés sur la faisabilité de l'allocation de puissance.

Mots-clés: sans fil, cellulaire, performance, contrôle de charge, faisabilité de l'allocation de puissance, naissance-et-mort, mobilité, régularité, ergodicité, mesure invariable, Gibbs, formule d'Erlang spatiale, blocage, probabilité de coupure, delai, débit.

Abstract

We build load control schemes for wireless cellular networks which are rapid and efficient and develop analytic methods for the performance evaluation of these networks by a spatial Markov process accounting for their geometry, dynamics and control schemes. We show that the analytic evaluation of the performance of wireless cellular networks is possible, but it requires the use of tools coming from several disciplines.

The first step is to characterize the single link performance by using the digital communication techniques. Then the interactions between the links are taken into account by formulating a power allocation problem. Optimal load control schemes based on the necessary and sufficient condition for the feasibility of the power allocation problem are unpractical because they are centralized. Extending the work in [16], we propose load control schemes that allow each base station to decide independently of the others what set of voice calls to serve and/or what bit rates to offer to elastic calls competing for bandwidth. These control schemes are primarily meant for large CDMA networks. They take into account in an exact way the influence of geometry on the combination of inter-cell and intra-cell interferences as well as the existence of maximal power constraints of the base stations and users. We also evaluate the performance of these control schemes in terms of the infeasibility probability (the probability that the admission's condition doesn't hold for a given cell when the calls are modelled as a Poisson process).

From the user's point of view, the performance is more suitably evaluated by the mean of the blocking and cut (drop) probabilities of streaming (real-time as voice) calls and the delay and throughput of elastic (non-real-time as web surfing) calls in the long run of the network. In order to build analytic methods for evaluating these performance indicators we build and analyze a pure-jump Markov generator that can be seen as a generalization of the spatial birth-and-death generator, which allows for mobility of particles. We give sufficient conditions for the regularity of the generator (i.e., uniqueness of the associated Markov process) as well as for its ergodicity. We show when the stationary distribution is a Gibbs measure. This extends previous work in [97] by allowing for mobility of particles. Our spatial birth-mobility-and-death process can be seen also as a generalization of the spatial queueing process considered in [106, 67]. This way our approach yields regularity conditions and alternative conditions for ergodicity of spatial open Whittle networks, complementing

works in [106, 67].

Next we apply our spatial Markov queueing process model to build analytical methods for evaluating the performance of wireless cellular networks controlled by feasibility based load control schemes. This evaluation is made by the mean of indicators which are relevant from the user's point of view rather than the classical outage probability [47]. Our formula for the blocking probability of streaming traffic might be seen as a spatial extension of the well-known Erlang loss formula. In the case of elastic traffic, we build explicit analytic expressions for the user throughput.

The analytic performance evaluation permit to build a new class of coherent methods for the different fundamental problems in wireless cellular systems: quality of service, capacity, dimensioning and cost. The ease of use of the analytical expressions makes this type of approach more effective than simulations for macroscopic evaluation and optimization.

Introduction

Our research stems from wireless cellular communications which are in permanent and rapid evolution. In few years, wireless cellular systems have evolved through several generations with completely different characteristics (for example, different multiple access schemes: FDMA[4] for GSM[5], CDMA[6] for UMTS[7], TDMA[8] for HSDPA[9]). This rapid evolution explains perhaps why much of the global performance analysis of these networks is made by simulations. Unfortunately, simulations give numerical result for a given situation but don't give a global comprehension of the key parameters and relationships.

0.1 Problem statement

The performance evaluation of wireless cellular networks, is a hard task. First, the performance evaluation of a single radio link is difficult because we should take into account the radio signal variations due to multi-path fading. Moreover the signal processing techniques, such as modulation, spreading and power control, etc., used to counteract the harmful effects of the radio signal variations are often complex (cf. for example [51, Chapter 9], [118, §3.4.3]).

Once the single link performance evaluation is carried, we should take into account the interference between the different links which depends on the relative geographic positions of the users. This process is usual in the engineering of the wireless cellular networks; first we characterize the performance of a single link and in a second step we consider the interactions between the different links. In doing so, we should consider carefully the separation of the time scales between phenomena such as multi-path fading and variations due to the geometry of the problem. (We will see an example of this issue when we study HSDPA later.) This is not an easy task as it depends not only on the phenomenon itself but also on the control algorithms (e.g. power control) in the network.

Wireless networks have to offer service for CALLS (USERS) which have different requirements and which may be roughly classified in two classes:

[4] Frequency Division Multiple Access
[5] Global System for Mobile
[6] Code Division Multiple Access
[7] Universal Mobile Telecommunications System
[8] Time Division Multiple Access
[9] High Speed Downlink Packet Access

- Real-time (or streaming) calls (voice calls, real-time audio-video streaming) require a fixed bit-rate on each link, and they are blocked if momentarily it is not possible to satisfy their requirement. To analyze the quality offred to such calls, one constructs loss models and studies blocking and cut (drop) probabilities.

- Non-real-time (or elastic) calls (data traffic) can be served at an arbitrary low bit-rate, for the price of large delays. To analyze the quality offred to such calls, one uses typically queueing models and studies sojourn times (delays) and average throughput.

The interaction between all the users is taken into account in specific control schemes called load control schemes, which may be roughly classified in two types: admission control for streaming calls and congestion control for elastic ones. These load control schemes attempt to assure that the required performance specific to each single radio link is satisfied. More precisely they manage traffic in order to ensure required quality for both incoming calls as well as previously admitted ones, and to reject incoming calls only when necessary. The essential problem for the load control decisions may be formulated as follows: may the network allocate a power to each user large enough for him to get his required link performance and small enough for the other users to get their required link performance. We say that load control attempt to ensure the *feasibility* of the *power allocation* problem, and if so, the user powers may be found by an iterative process called power control. Solving the power control problem isn't in the scope of the present work (cf. for example [55] and the references therein). We will focus only on criteria which indicate if the power control problem is feasible or not, without trying to solve it.

One may consider a load control scheme based on the necessary and sufficient condition (denoted NSFC) of the feasibility of the power allocation problem. Such load control scheme is optimal (i.e. offers the maximal capacity) but is unfortunately difficult to implement in real networks as the admission decision of a new call requires the collection and treatment of information from all the calls in the network. We call such load control scheme centralized. Moreover the performance evaluation of the optimal load control scheme is time consuming numerically, and hard analytically (no explicit formulae exist to our knowledge). The load control schemes implemented in the real networks are proposed by constructors (manufacturers) [64], [81]. They are decentralized (i.e. depend on parameters which are local to the cell in which the new call request for admission) but they don't assure the power allocation feasibility. Moreover no analytic evaluation method of the constructor schemes exist to our knowledge.

There is a rich literature on the performance evaluation of load control schemes in cellular networks. Unfortunately it is often difficult to find the relationship between the indicators calculated by the authors because they consider different traffic models. The distinction between the following four classes of traffic models allows a first classification.

- STATIC MODEL: Number and positioning of ACTIVE (i.e. currently being served) calls are fixed.

- SEMI-STATIC MODEL: Active calls are modelled by a spatial Poisson point process. In other words, "snapshots" of active calls are seen as realizations of spatial Poisson processes; these snapshots are used as the non-constrained traffic process on which we will define and evaluate the (in)feasibility probabilities.

- SEMI-DYNAMIC MODEL: Users (or calls) arrive at a random location and last for some random duration; each user is motionless during its call; this is the "minimal" dynamic model where an admission control can be specified, and where blocking probabilities can be considered.

- DYNAMIC MODEL: We have the same as above but users may move during their calls; an admission and motion (or handoff) control can then be specified. Blocking and motion-cut probabilities can be evaluated.

The load control scheme performance may be evaluated by modelling the users as a planer Poisson process. This lead in the classical literature to the notion of *outage* probability [47], which is roughly the probability that a given user doesn't attain his required link performance. The outage probability is not a satisfying performance indicator because it relies on some simplifying assumptions (especially on the powers of the users) which makes its meaning unclear. Another approximate method consists of making some average calculus leading in the classical literature to the notion of *pole capacity.* Both the outage probability and pole capacity are heuristics, and we consider in the present work more relevent performance indicators.

From the user's point of view, the PERFORMANCE is more suitably evaluated by means of the long run *blocking* and *cut* probabilities for streaming calls and the *delay* and *throughput* for elastic calls. Recall that the BLOCKING PROBABILITY is defined as the fraction of calls that are rejected by the admission control scheme in the long run, a notion of central practical importance. (Analogous definitions may be formulated for the other indicators.) In order to calculate these DYNAMIC PERFORMANCE INDICATORS (blocking, cut, delay and throughput) we should consider the *temporal* dynamics and the GEOMETRY (localizations) of the call arrivals, mobility and departures from the network. The temporal dynamics of the call arrivals and departures are well studied in wired communication networks, which led in particular to the famous (and widely used) Erlang's formula [44]. This formula is often used for wireless cellular networks by eliminating the spatial component of the problem. In fact no analytic methods for calculating performance of wireless cellular networks accounting for the spatial component (geometry of interference) exist in previous literature to our knowledge. Most work done in this field involves time consuming and complex simulations which are not suitable for dimensioning and global cost and capacity optimization of wireless networks. Analytic performance evaluation methods are not only suitable for the dimensioning and optimization of wireless cellular networks which are crucial tasks for the network operators, but also of great interest for the scientists how attempt to understand the current system performances and propose modifications to ameliorate them. Till now the work of operators and scientists is done either by considering only the single link performance; or by heuristically accounting for the geometry of interference.

Classical queueing and loss models (see e.g. [74]) are well adapted to wired networks, where the spatial component of the model is typically represented by some *graph of links*, and where the coexistence of calls on a common link is modeled by the occupancy of a *discrete number of circuits* available on this link. In wireless cellular communications, one needs to take into account the spatial characteristics of the network in a more thorough way because *it is the relative location of all the radio channels that determines their joint feasibility.* One of the additional difficulties then stems from the fact that the spatial component of the model is subject to changes due to the mobility of users and instantaneous changes of radio conditions. All this makes spatial models more suitable for analysis of wireless communications.

0.2 Objectives

We aim to solve the problem stated above. More specifically, three objectives are particularly relevant:

1. Firstly we aim to build rapid, accurate, and efficient *load control* schemes for wireless cellular networks. (We say that a load control scheme is ACCURATE if it assures the feasibility of the power allocation problem; and we say that it is EFFICIENT if it offers a capacity close to the optimum –corresponding to NSFC–.)

2. Secondly we aim to develop a *stochastic model* for wireless cellular networks accounting for their geometry, dynamics and control schemes and permitting to evaluate analytically their performance. This model should be general enough such that different cases as: streaming or elastic traffic; with or without mobility; CDMA, FDMA or TDMA; etc. would be particular cases of the general model.

3. Thirdly we aim to apply the above model to the *performance evaluation* of real wireless cellular networks: UMTS, HSDPA, GSM.

In fact the three objectives above are closely related since we begin by building load control schemes (objective 1), then we develop mathematical models (objective 2) permitting to analytically evaluate the performance of these schemes which are precisely the performance of the wireless cellular networks (objective 3).

0.3 Organization

The report comprises three parts I, II and III corresponding to the three above objectives respectively.

The introductions of each part (and some chapters) give a detailed state of the art and a description of its novelty as well as its organization.

The thesis report is long, but the reader may read the parts I and II in any order he wants. The reader interested in the mathematics of the stochastic tools developed in the thesis may read part II, whereas the reader interested in the application of these tools to cellular networks may read parts I and III. The appendices (Part V) are long because we gather some basic results scattered in the litterature, and present some useful complementary numerical results.

0.4 Publications

Journals. Our papers [13] and [15] contain some material from Parts I and II respectively.

Conferences. Our papers [14] and [21] contain some material from Parts II and III.

Patents. The load control algorithms proposed in Part I are patented in [11] and [12].

Part I

Feasibility based load control

Chapter 1

Introduction

The present part focuses on the first objective described in §0.2, i.e. to build load control schemes for wireless cellular networks which are rapid, accurate, and efficient. Recall that the load control schemes attempt to assure that the required performance of each single radio link is satisfied while taking into account the interactions between all the users.

1.1 Related works

Load control algorithms. The most largely proposed load control algorithms for CDMA networks are based on the total interference received at the base station for the uplink [120] and on the power transmitted by the base station for the downlink [77]. Constructors of UMTS infrastructure implemented load control schemes based on these indicators as described in [64], [81]. Many other load indicators are proposed: signal to interference ratio [84], throughput, effective bandwidth, number of active connections. We call this class of algorithms DIRECT algorithms. The direct algorithms usually do not guarantee the quality requirements to all calls and call dropping can occur even instantaneously as a call is admitted. In order to avoid this call dropping, security margins are applied which may decrease the offered capacity if the margins are too large.

Some authors propose to temporarily admit new calls with a low power level and to evaluate if a new feasible power allocation can be found (cf. for example [7]). We call this class of algorithms TRIAL algorithms. The duration of the admission process of these algorithms may be too long which makes them impractical.

An emergent method is based on a criterion which indicates if the power control problem is feasible or not, without trying to solve it. A decentralized version of such criterion is proposed in [16] for the downlink case without power limit. We will call the load control algorithms based on this idea FEASIBILITY load control algorithms.

Performance. The performance indicators introduced in [52, 121, 84, 47] correspond to the probability that the signal-to-interference-and-noise ratio (SINR) threshold is less than some threshold, when users, modeled as a Poisson point process, are all accepted. In [84] and [47] this indicator is called the OUTAGE PROBABILITY. The authors of [121] call it the blocking probability, but as mentioned in [84], the term outage probability is more appropriate.

The authors of [16] introduce the notion of INFEASIBILITY PROBABILITY which designates the probability that *there is no solution to the power allocation problem* when the users are modelled as a Poisson point process. Observe that the *infeasibility* probability is different from the *outage* probability which is related to the event that the transmission quality of service is not attained for *given transmission powers*. Hence both the outage and the infeasibility probabilities are related to "the probability that the transmission quality of service is not attained". But the outage probability depends on the transmission powers of the users and the base stations; whereas the infeasibility probability corresponds to an intrinsic characterization of power allocation feasibility, and consequently doesn't depend on transmission powers. The infeasibility probability is then a more appropriate performance indicator.

Power allocation related work. Load control is closely related to the power allocation problem. In fact the latter has already been considered by several authors more for estimating the capacity of wireless cellular networks than for building load control schemes. Nettleton and Alavi [1] first considered the power allocation problem in the cellular spread spectrum context.

Gilhousen et al [52], pose the problem the following way. Suppose Base Station number 1 emits at the total power P_1 in the presence of $K-1$ other base stations, which emit at power $P_2, \ldots, P_K$ respectively. How many users N_1 can then base station 1 accommodate assuming that the load of the network is only interference-limited and that each user has some required bit rate? In [52] a sufficient condition is proposed which allows for the determination of N_1. But this condition comprises $P_1, \ldots, P_K$, hence it does not reflect a key feature, that in reality the total power emitted by the base station should depend on the number of users (and even on their locations), namely P_k should be a function $P_k(N_1, \ldots, N_K)$.

In order to address this issue, Zander [126, 125] expresses the global power allocation problem by the multidimensional linear inequality

$$\mathbb{A}\mathbb{P} \leq \frac{1+\xi}{\xi}\mathbb{P} \tag{1.1}$$

with unknown vector $\mathbb{P}$ of emitted powers; here one assumes the required signal-to-interference power ratio ξ (or equivalently the required user bit rate) to be given and one assumes the matrix $\mathbb{A}$, the i,k-th entry of which gives the normalized path-losses between user i and base station k, to be given too. The main result is then that the power allocation is *feasible* (i.e. there exists a non-negative, finite solution to (1.1)) if and only if $\xi < 1/(\boldsymbol{\rho}(\mathbb{A}) - 1)$, where $\boldsymbol{\rho}(\mathbb{A})$ is the spectral radius of the matrix $\mathbb{A}$. (The SPECTRAL RADIUS of a matrix $\mathbb{A}$ is defined as the maximum of the absolute values of the eigenvalues of $\mathbb{A}$.) In order to simplify the problem, all same-cell channels are assumed to be completely orthogonal and the external noise is suppressed.

Foschini and Miljanic [49] and Hanly [56] introduced external noise to the model: Foschini considered a narrow-band cellular network and Hanly a two-cell spread spectrum network. On the basis of the previous works, Hanly extended the model in several articles. Hanly [59] extends this approach to the case with intra-cell interference and external noise (essentially for the uplink). Using the block structure of $\mathbb{A}$, he solves the problem in two steps: first the own-cell power allocation conditions are studied (microscopic view) and then the macroscopic view considers some aggregated cell-powers. He also interprets $\boldsymbol{\rho}(\mathbb{A})$ as a measure of the traffic congestion in the network.

The evaluation of $\boldsymbol{\rho}(\mathbb{A})$ can be done either from a centralized knowledge of the state of the network, which is non practical in large networks, or by channel probing as suggested in [59, §VIII] and described in [128]. When it exists, the minimal finite solution of inequality (1.1) can also be evaluated in a decentralized way (using Picard's iteration of operator $\mathbb{A}$, cf. the discussion in [58, §IX]). However this does not provide decentralized admission or congestion control algorithms, namely decentralized ways of controlling the network population or bit rates in such a way that the power allocation problem remains feasible, namely that $\boldsymbol{\rho}(\mathbb{A})$ remains less than $1 + 1/\xi$.

The approach in [126, 59] is continued in [16], where decentralized admission/congestion control protocols are proposed for the downlink, without maximal power constraints. These protocols are based on the simple mathematical fact that the maximal eigenvalue of any sub-stochastic matrix (i.e. matrix with non-negative entries, whose row sums are less than 1) is less than 1.

1.2 Our contribution

The single link performance requirement may be expressed as the signal-to-interference power ratio larger than a given threshold. If a power allocation satisfying these constraints and maximal power limit exists, then we say that power allocation with power limitations is FEASIBLE.

Our work continues [16] by building decentralized power allocation FEASIBILITY CONDITIONS (denoted FC) taking into account the power limits and the uplink. We build admission control algorithms based on these conditions for a given user positions (static model).

Moreover we build explicit approximate formulae for the performance evaluation of FC load control in a hexagonal network in terms of the INFEASIBILITY PROBABILITY (defined as the probability that FC doesn't hold for a given cell when the users are modelled as a Poisson process –semi-static model).

1.3 Organization

The present part is organized as follows.

In the premilinary chapter 2 we characterize the performance of each single link, describe the model and present the notation.

In Chapter 3 we build decentralized conditions for the feasibility of the power allocation problem in a *static* traffic model.

In Chapter 4 we evaluate the performance of these conditions in a *semi-static* traffic model.

Chapter 2

Preliminaries

2.1 Single link performance

The present section is just a collection of the relevant results from the literature to which we refer the reader for more details. The MULTI-PATH FADING channel may be modelled as a linear time varying Input/output model [51, Chapter 9]

$$v_m = \sum_{k=1}^{L} g_{k,m} u_{m-k} + z_m$$

where $\{u_m\}$ is the input, $\{v_m\}$ is the output, $\{g_{k,m}\}$ is the channel filter and $\{z_m\}$ designates the noise. The multi-path channel is characterized by the statistics of the channel filter. The noise is assumed to be additive white Gaussian (AWGN) with (power-spectral) density N_0.

The multi-path fading channel may be seen as a random channel with AWGN noise as that analyzed in [31, §III.4.1]. The performance of a given *modulation* scheme may be expressed by its bit-rate, r, and a curve giving the error-probability as function of the energy-per-bit to noise-density ratio, E_b/N_0, at the *input of the receiver*. It is usual to fix some error-probability threshold, and deduce the corresponding E_b/N_0 threshold. In general, we have several modulations and a specific E_b/N_0 threshold for each one.

[118, §3.4.3] studies the modulation performance in a CDMA network such as UMTS Release 99 where active users use *simultaneously* the *entire* system bandwidth. Using the arguments in [53, §6.3], [118, §3.4.3], we deduce that for this system the energy-per-bit E_b is averaged over fading.

The signal and noise powers are given respectively by

$$\mathbf{S} = rE_\mathrm{b}, \quad N = WN_0$$

where r is the bit-rate, E_b is the energy-per-bit and W is the bandwidth (5 MHz for UMTS). Hence the SIGNAL-TO-NOISE RATIO, denoted $\mathbf{S}/N$, equals

$$\frac{\mathbf{S}}{N} = \frac{r}{W}\frac{E_\mathrm{b}}{N_0}$$

In practice, we have for UMTS the following relation

$$\frac{\mathbf{S}}{N} = \frac{r}{W'}\frac{E_{\mathrm{b}}}{N_0} \tag{2.1}$$

where W' designates the chip-rate ($W' = 3.84\,\mathrm{MHz}$ for UMTS). For each modulation, given the error-probability threshold we get the E_{b}/N_0 and $\mathbf{S}/N$ thresholds.

We apply now the above link performance characterization to streaming and elastic traffic. A streaming call requires to transmit for some *duration* with a given modulation, i.e. a *given* bit-rate and energy-per-bit to noise-density ratio E_{b}/N_0, and thus a given $\mathbf{S}/N$ threshold. It is served by a (fixed bit-rate) DEDICATED CHANNEL (DCH) in UMTS. If the required rate may not be offered, then the call is blocked (by the ADMISSION CONTROL scheme). We may consider different streaming classes, each characterized by a specific $\mathbf{S}/N$ threshold.

An elastic call has an *amount of data* to transmit at a bit-rate among a (finite) set of possible rates which *may be adjusted* by the network (by the CONGESTION CONTROL scheme). Elastic calls may be served by the DOWNLINK SHARED CHANNEL (DSCH) in UMTS. The E_{b}/N_0 thresholds of the various modulations used on DSCH (called DSCH MODULATIONS) are close [64, §12.5.1], so we may take a single representative E_{b}/N_0 and assume that the set of possible rates is *continuous*: $\mathbb{R}_+$. With this assumption we get the linear relation (2.1) between the $\mathbf{S}/N$ ratio and the bit-rate r.

We may also consider the SHANNON'S BOUND which gives the *theoretical maximal bit-rate* over an AWGN channel

$$r = W \log_2\left(1 + \frac{\mathbf{S}}{N}\right) \tag{2.2}$$

where the parameters are the same as for the previous display. For a steaming call, the bit-rate is fixed, hence we may deduce from (2.2) the corresponding $\mathbf{S}/N$ threshold. For an elastic call, Equation (2.2) is a non linear relation between the $\mathbf{S}/N$ ratio and bit-rate. Using the property of the log function we have the bound

$$\frac{\mathbf{S}}{N} \geq \frac{r}{W}\ln(2) \tag{2.3}$$

and, if $\mathbf{S}/N \ll 1$ then we have the approximation

$$\frac{\mathbf{S}}{N} \simeq \frac{r}{W}\ln(2) \tag{2.4}$$

hence we get also in this case a linear relation between the $\mathbf{S}/N$ ratio and the bit-rate.

Once the single link performance is characterized, we should take into account the interference between the different links. To this end, we make the approximation that the interference observed by some user may be *approximated by a AWGN* of power equal to the sum of the noise and the interference powers (averaged over the multipath fading). This approximation is justified by the large number of interferers (Central Limit Theorem) in [118, §4.3.1]. Hence the $\mathbf{S}/N$ threshold may be called the signal-to-interference-and-noise ratio (SINR) threshold.

2.2 Model

We now describe the considered multiple access scheme, the cell patterns (base station positions), the propagation model, the antenna types and the typical numerical values for UMTS system used in the numerical applications throughout the present part.

2.2.1 Multiple access

In order to simplify the presentation, we focus in the present part on CDMA networks such as UMTS Release 99 (recall that in such system the active users use *simultaneously* the *entire* system bandwidth). However, our approch is sufficiently general and may be extended to wireless cellular networks with other multiple access schemes (cf., for example, Chapter 11 for TDMA and Chapter 12 for FDMA).

2.2.2 Cell pattern

The radio part of a wireless cellular network comprizes some base stations. Each BASE STATION has an antenna and serves a geographic zone called cell. The CELL is defined as the set of locations in the plane which receive a signal from a base station which is stronger than the signal from any other base station. We assume in the present study that each user is served by a single base station (no macrodiversity). The effect of MACRODIVERSITY (that is a user may be served by several base stations) on the stability of a wireless cellular network serving elastic traffic is studied for example in [28].

We consider a HEXAGONAL MODEL where the base stations are placed on a regular grid denoted on the complex plane by $\{\Delta(p+qe^{i\pi/3});(p,q)\in\mathbb{Z}^2\}$ where Δ is the distance between two adjacent base stations and $\mathbb{Z}$ designates the set of all the integers, both positive and non-positive. Denote by λ_{S} the mean number of base stations per km^2. Let R be defined by the formula

$$\lambda_{\mathrm{S}} = 1/(\pi R^2) \tag{2.5}$$

Bearing this definition in mind, we call R the CELL RADIUS. The cell radius is in fact the radius of the disc whose area is equal to that of the hexagon. In order to simplify some calculations, we make the following approximation (as in [80]).

Approximation 1 FROM HEXAGON TO DISC. *We approximate the hexagonal cell with the (virtual) disc of radius R. This is illustrated in Figure 2.1.*

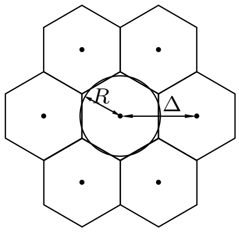

Figure 2.1: Hexagon to disc approximation

Lemma 1 *The cell radius R is related to the distance Δ between two adjacent hexagons by $\Delta^2 = 2\pi R^2/\sqrt{3}$, or equivalently*

$$R = \Delta\sqrt{\frac{\sqrt{3}}{2\pi}} \tag{2.6}$$

Numerically this gives $R \simeq 0.525\ \Delta$.

Proof. Let $\mathbf{r}$ be the radius of the circumscribed circle to a hexagon. It is easy to see that $\Delta = \mathbf{r}\sqrt{3}$. The surface of a hexagon is $6\frac{1}{2}\mathbf{r}\frac{\Delta}{2} = \frac{\sqrt{3}}{2}\Delta^2$. Hence we should have $\frac{\sqrt{3}}{2}\Delta^2 = \pi R^2$ which gives the desired relation. ■

We consider sometimes a pattern where base station positions constitute a Poisson process in the plane, with intensity λ_S. The cell pattern in this model is called POISSON-VORONOI. The hexagonal and Poisson-Voronoi models are illustrated in Figure 2.2. Note that the base station locations as well as the cells are *random* in the Poisson-Voronoi model. We always assume that the Poisson process of base stations is independent from all other considered random elements.

The hexagonal and Poisson-Voronoi models are extreme and complementary architectures: The hexagonal model represents perfectly structured networks, whereas the Poisson-Voronoi model takes into account irregularities of real networks in a statistical way. We shall treat in details the hexagonal model and just recall the results for the Poisson-Voronoi model from [16] for the purpose of comparison.

We consider large networks where the number of base stations and the area covered by the network may be very large. We consider both the DOWNLINK (from the base stations to the users) and the UPLINK (from the users to the base stations).

If the network is modelled on some bounded zone of $\mathbb{R}^2$, then the cells at the frontier of the zone don't suffer the same interference as the cells in the center of the zone. In this case we model the network on a TORUS in order to avoid the border effect.

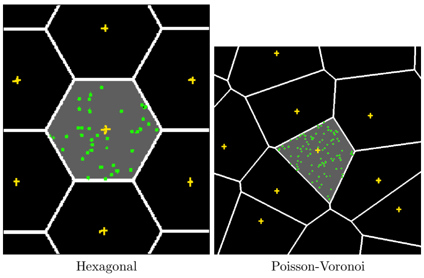

Figure 2.2: Hexagonal and Poisson-Voronoi models

2.2.3 Propagation

We model PROPAGATION-LOSS on distance $\mathbf{r}$ by

$$L(\mathbf{r}) = (K\mathbf{r})^{\eta} \tag{2.7}$$

where $\eta > 2$ is the so-called PROPAGATION EXPONENT and $K > 0$ is a multiplicative constant. The above formula represents the effect of the distance which is the principal cause of the received signal variations.

In order to simplify some formulae we introduce the NORMALIZED PROPAGATION-LOSS

$$l(\mathbf{r}) = L(\mathbf{r})/L(R)$$

where R designates the cell radius.

The objects in the path between the antenna and the user (as for e.g. hills, buildings, trees, etc.) affect also the received signal. We may distinguish two scales of these variations: *fast fading* and *shadowing* (called also slow fading). FAST FADING is related to multi-path propagation [51, Chapter 9] and induces variations over about half a wavelength. The effect of fast fading on the performance of a single link is studied in [118, §3.4.3] and recalled briefly in Section 2.1. SHADOWING is related to diffraction over the obstacles in the path between the antenna and the user and induces variations over several wavelengths. Measurements have shown that the shadowing factor is log-normal distributed. We will not take into account the shadowing effect in the present work.

2.2.4 Antennas

We will consider the following two versions of the hexagonal network (see Figure 2.3):

- OMNI**:** each base station is equipped with an *omni* antenna; this antenna serves users in the whole disc around the base station.

- DIRECTIONAL**:** each base station is equipped with a *directional* antenna; each antenna serves users in its SECTOR defined as the set of locations in the disc within the cone of 120° around the antenna's azimuth.

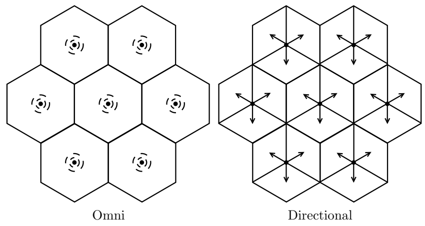

Figure 2.3: Illustration of the two versions of the hexagonal network: omni and directional.

In real network the power transmitted by the antennas and by the users is limited by some maximal-power constraints. We consider the cases WITH and WITHOUT POWER LIMIT. Even if the case "without power limit" isn't realistic, we will see that it is a limit of the case "with power limit" when the cell radius is small which is the case for dense urban zones.

We denote by $\tilde{P}$ the maximum power including the transmitting and receiving antenna gains, denoted $\mathbf{G}$, and losses, denoted $\mathbf{L}$ (which may include transmitting and receiving antenna loss, body loss, indoor loss, etc.). We denote by $\hat{P}$ the maximum power which doesn't account for $\mathbf{G}$ and $\mathbf{L}$, i.e.

$$\hat{P} = \tilde{P} - \mathbf{G} + \mathbf{L}$$

2.2.5 UMTS numerical parameters

Unless otherwise specified, the following values are used for the numerical applications throughout the present part.

Propagation

We consider the following propagation-loss parameters $\eta = 3.38$, $K = 8667$ (which correspods to the so-called Cost-Hata propagation model in an urban area [45]).

Cell radius

We consider the following cell radii

$$R = \begin{cases} 0.525, 1, 2, \ldots, 5\text{km} & \text{for the downlink} \\ 0.525, 1, 2, 3\text{km} & \text{for the uplink} \end{cases}$$

Antenna parameters

By default, we consider omni antennas with loss $\mathbf{L} = 0$ and gain([1])

$$\mathbf{G} = \begin{cases} 9\text{dBi} & \text{for the downlink} \\ 12\text{dBi} & \text{for the uplink} \end{cases}$$

(The antenna gain isn't null because the energy is focused on a plan). For directional antennas we take

$$\mathbf{G} = \begin{cases} 12\text{dBi} & \text{for the downlink} \\ 15\text{dBi} & \text{for the uplink} \end{cases}$$

The maximum power (without antenna gain and loss) equals $\hat{P} = 43dBm$([2]) for the base stations and $\hat{P} = 21dBm$ for the mobiles.

Downlink specific parameters

A supplementary limit on the power transmitted to each user is sometimes imposed, typically $36dBm$ (without antenna gain and loss). In the present work, we have only considered the limit on the total power transmitted by the base station.

The COMMON CHANNELS (CCH), including pilot, synchronization and paging channels, have a constant power, denoted P'. We assume that P' is a fraction of the maximal power

$$P' = \epsilon \tilde{P}$$

where $\epsilon = P'/\tilde{P} = 0.12$.

Orthogonality factor

The ORTHOGONALITY FACTOR, denoted α, affects the intra-cell interference. Typically

$$\alpha = \begin{cases} 0.4 & \text{in the downlink} \\ 1 & \text{in the uplink} \end{cases}$$

The orthogonality factor takes into account approximately the loss of orthogonality of the spreading sequences within a cell due to the multi-path. Hence $\alpha = 0$ for perfectly orthogonal. (Therefore, we should call α the *non*-orthogonality factor, but it is the usage to call it the orthogonality factor.)

Noise power

Typically, the noise power equals

$$N = \begin{cases} -103\text{dBm} & \text{in the downlink} \\ -105\text{dBm} & \text{in the uplink} \end{cases}$$

[1] The antenna gain $\mathbf{G}$ is a ratio of two powers, hence it has not a unit. The logarithmic representation, $10\log_{10}(\cdot)$, may be expressed in dB, which is denoted "dBi" in the particular case of antenna gains.

[2] The abbreviation dBm designates "dB milli-Watt", i.e. $10\log_{10}$ of the value in milli-Watt.

*Link***service**	**voice**	**data** 64kbps(3)	**data** 144kbps	**data** 384kbps
Downlink	−16	−11	−9	−5
Uplink	−18	−14	−12	−8

Table 2.1: SINR thresholds in dB for UMTS vehicular-A channel.

SINR thresholds

Unless otherwise specified, calculations are made for voice with SINR threshold

$$\xi = \begin{cases} -16\text{dB} & \text{in the downlink} \\ -18\text{dB} & \text{in the uplink} \end{cases}$$

We consider sometimes other streaming classes whose SINR thresholds are given in Table 2.1 (from [81]).

E_{b}/N_0 threshold

We take for elastic services on the DSCH in UMTS a single representative value of the E_{b}/N_0 equal to 5dB [64, §12.5.1].

2.3 Notation

We will use the following notation.

2.3.1 Antenna locations and path loss

- u, v designate indexes for base stations.
- m, n designate indexes for users (mobiles). The letter designating a base station (or a user) is sometimes used to designate it geographic position.
- $\mathbf{U}$ is the set of base stations (which is assumed finite, but some results may be extended to the infinite case [16]).
- $\mathbf{M}$ is the set of users.
- We denote $m \in u$ to say that a user m is served by base station u. Hence we use the same letter to designate the base station and the set of users it serves.
- $L_{u,m}$ is the propagation-loss between base station u and user m. For the propagation-loss function (2.7) we get $L_{u,m} = L\left(d\left(u,m\right)\right)$ where $d\left(u,m\right)$ designates the Euclidian distance between a user at position m and a base station at position u.

2.3.2 Engineering parameters

- ξ_m is the signal-to-interference-and-noise ratio threshold for user m.
- N is the external noise power.
- α is the orthogonality factor.
- In order to simplify the formulae we introduce

$$\alpha_{uv} = \begin{cases} 1 & \text{if } v \neq u \in \mathbf{U} \\ \alpha & \text{if } v = u \in \mathbf{U} \end{cases}$$

  and the MODIFIED SINR

$$\xi'_m = \xi_m / (1 + \alpha \xi_m), \quad m \in \mathbf{M}$$

- In the downlink, we will use the following notation:
  - $\tilde{P}_u$ designates the maximal total power of base station u;
  - $P_{u,m}$ designates the power of dedicated channel (DCH) of user $m \in u$;
  - P'_u is the power of common channels of base station u which is a fraction of the maximal power

$$P'_u = \epsilon \tilde{P}_u, \quad u \in \mathbf{U}$$

    where ϵ is a given constant.
  - $P_u = P'_u + \sum_{m \in u} P_{u,m}$ is the total power transmitted by base station u.
- In the uplink, we will use the following notation:
  - $\tilde{P}_m$ designates the maximum power of user m.
  - P_m designates the power transmitted by user m.
  - $I_u = N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} P_v / L_{u,n}$ is the total power (sum of noise and powers from all the users) received at base station u. We will call I_u the TOTAL INTERFERENCE received at base station u.
- We shall see that the following parameter, called F-FACTOR, plays an important role in the analysis of the performance of cellular networks

$$f(m) = \sum_{v \neq u} L_{u,m} / L_{v,m}, \quad u \in \mathbf{U}$$

  (See Annex 13.B for the properties and in particular the moments of the f-factor.)

2.3.3 Mathematics

For a random variable Z we denote $\bar{Z}$ its expectation, i.e.

$$\bar{Z} = \mathbf{E}\left[\bar{Z}\right]$$

In particular for a function $f\left(m\right)$ of the user $m \in u$ we denote

$$\bar{f} = \mathbf{E}\left[f\left(n\right)\right]$$

where n is a random user in the cell u.

Chapter 3

Feasibility conditions

The objective of the present chapter is to build decentralized conditions for the feasibility of the power allocation problem in a *static* traffic model.

The present chapter is organized as follows. In Section 2.1 we characterize the single link performance. The following two Sections 3.1 and 3.2 correspond to the downlink and uplink respectively with similar contents.

Section 3.1 is composed of 2 subsections. In subsection 3.1.1 we state the power allocation feasibility problem which rises from the interactions between the different radio links. In subsection 3.1.2 we solve this problem by building decentralized feasibility conditions.

In Section 3.3 we describe admission control schemes based on the feasibility conditions.

3.1 Downlink

We consider here the downlink (DL), i.e. the link from the base stations to the users. The reverse or uplink (UL) is treated in the following section.

The downlink power allocation is studied in [126], [87], [112], [63] and [16].

A first difference from the uplink case is the orthogonality factor α which affects the intra-cell interference only in the downlink. Moreover in the downlink there are common channels (such as the pilot channel) which have to be taken into account as interferers. Finally, interference in the downlink comes from the base stations which have fixed locations whereas the interference in the uplink comes from the users which have variable locations. We will see that in spite of these differences, the algebras of the power allocation problems of the two links present some similarities.

The authors of [16] consider the downlink with neither power limit nor common channels. We extend their results to take into account these two features. Moreover we propose explicit approximate methods to evaluate the probability that the power allocation is infeasible (for a Poisson user population and a hexagonal base station architecture).

3.1.1 Power allocation problem

We aim to formulate the power allocation problem for a given base station positions and user population (fixed positions $\{m\}$ and SINR thresholds $\{\xi_m\}$). We will say that power allocation is FEASIBLE if there exist powers such that the SINR for each user m is larger than the SINR threshold ξ_m.

Proposition 1 MATRIX REPRESENTATION. *The downlink power allocation problem is feasible if there exist powers* $\{P_{u,m} \in \mathbb{R}^+; m \in u \in \mathbf{U}\}$ *such that*

$$\frac{P_{u,m}/L_{u,m}}{N - \alpha P_{u,m}/L_{u,m} + \sum_{v \in \mathbf{U}} \alpha_{uv} P_v / L_{v,m}} \geq \xi_m, \quad m \in u \in \mathbf{U} \tag{3.1}$$

The problem above may be written

$$\begin{cases} (\mathbf{1} - \mathbb{A})\,\mathbb{P} \geq \mathbf{a} \\ \mathbb{P} \geq 0 \end{cases} \tag{3.2}$$

where the matrix $\mathbb{A} = [\mathbb{A}_{m,n}]$ *is given by*

$$\mathbb{A}_{m,n} = \alpha_{uv} \xi'_m L_{u,m} / L_{v,m}, \quad m \in u \in \mathbf{U}, n \in v \in \mathbf{U} \tag{3.3}$$

the vector $\mathbf{a} = (\mathbf{a}_m)^T$ *(where* ^{T} *designates the transpose operation) is given by*

$$\mathbf{a}_m = \left(N + \sum_{v \in \mathbf{U}} \alpha_{uv} P'_v / L_{v,m} \right) L_{u,m} \xi'_m, \quad m \in u \in \mathbf{U}$$

and the vector $\mathbb{P} = (P_{u,m})^T$.

The power allocation problem above is feasible iff[1]

$$\boldsymbol{\rho}(\mathbb{A}) < 1$$

in which case

$$\mathbb{P}^* = (\mathbf{1} - \mathbb{A})^{-1} \mathbf{a} \tag{3.4}$$

is the minimal solution.

Proof. The power received by user m from its serving base station u is $P_{u,m}/L_{u,m}$. The interference due to another user $n \in u$ is $P_{u,n}/L_{u,m}$. Then the interference due to own cell, called INTRA-CELL interference, is

$$I^{(i)}_{u,m} = \alpha \left(P'_u + \sum_{n \in u \setminus \{m\}} P_{u,n} \right) / L_{u,m} = \alpha \left(P_u - P_{u,m} \right) / L_{u,m}, \quad m \in u$$

The interference due to other cells, called INTER-CELL interference, is

$$I^{(e)}_{u,m} = \sum_{v \in \mathbf{U} \setminus \{u\}} P_v / L_{v,m}, \quad m \in u$$

[1]The abbreviation iff means "if and only if".

Hence, the signal-to-interference-and-noise ratio equals

$$\frac{P_{u,m}/L_{u,m}}{N + \alpha\left(P_u - P_{u,m}\right)/L_{u,m} + \sum_{v\in\mathbf{U}\setminus\{u\}} P_v/L_{v,m}}$$

which may be rearranged to get the left hand side of the inequality (3.1). (When we neglect the noise term $N = 0$, Inequality (3.1) is similar to [126, Equation (3) and Definition §III]. The inequality (3.1) is slightly different from that given in [63] because we neglect here the synchronization channel specificity considered there.)

We rearrange the inequality (3.1) as follows

$$\frac{P_{u,m}/L_{u,m}}{N + \sum_{v\in\mathbf{U}} \alpha_{uv} P_v/L_{v,m}} \geq \xi'_m$$

Hence

$$\frac{P_{u,m}/L_{u,m}}{N + \sum_{v\in\mathbf{U}} \alpha_{uv} P'_v/L_{v,m} + \sum_{v\in\mathbf{U}} \sum_{n\in v} \alpha_{uv} P_{v,n}/L_{v,m}} \geq \xi'_m$$

Then it is easy to see that Problem (3.1) may be written in the form (3.2).

Corollary 12 gives the last part of the proposition. ■

Corollary 1 ACHIEVABLE SINR TARGETS. *Assume that the ξ_m are constant, i.e. $\xi_m = \xi$. We say that some* SINR TARGET *is* ACHIEVABLE *if the power allocation problem for that value of SINR is feasible. The set of achievable SINR targets is*

$$0 \leq \xi < \frac{1}{\boldsymbol{\rho}\left(\mathbb{A}'\right) - \alpha}$$

where the matrix $\mathbb{A}' = \left[\mathbb{A}'_{m,n}\right]$ *is given by*

$$\mathbb{A}'_{m,n} = \alpha_{uv} L_{u,m}/L_{v,m}, \quad m \in u \in \mathbf{U}, n \in v \in \mathbf{U}$$

Proof. We may write $\mathbb{A} = \xi'\mathbb{A}'$. By Proposition 8, the power allocation problem is feasible iff

$$\xi'\boldsymbol{\rho}\left(\mathbb{A}'\right) < 1$$

Note that $\mathbb{A}'_{m,m} = \alpha$ then $\boldsymbol{\rho}\left(\mathbb{A}'\right) \geq \sum_{n\in\mathbf{M}} \mathbb{A}'_{m,n} > \alpha$. Since $\xi' = \xi/\left(1 + \alpha\xi\right)$, we get the desired result. (The author [126] gives a similar result.) ■

Reduced problem. We aim to formulate the power allocation problem in terms of the powers $\{P_u; u \in \mathbf{U}\}$ transmitted by the base stations. To this end we rearrange the inequality (3.1) as follows

$$P_{u,m} \geq \left(N + \sum_{v\in\mathbf{U}} \alpha_{uv} P_v/L_{v,m}\right) L_{u,m}\xi'_m, \quad m \in u \in \mathbf{U}$$

We now add over the set $\{m \in u\}$

$$\sum_{m\in u} P_{u,m} \geq \sum_{m\in u} \left(N + \sum_{v\in\mathbf{U}} \alpha_{uv} P_v / L_{v,m} \right) L_{u,m}\xi'_m, \quad u \in \mathbf{U}$$

The left term equals $P_u - P'_u$.

We say that the REDUCED POWER ALLOCATION PROBLEM is feasible if there exist powers $\{P_u \in \mathbb{R}^+; u \in \mathbf{U}\}$ of the base stations such that

$$P_u \geq N \sum_{m\in u} L_{u,m}\xi'_m + P'_u + \sum_{v\in\mathbf{U}} \alpha_{uv} \sum_{m\in u} L_{u,m}/L_{v,m}\xi'_m P_v, \quad u \in \mathbf{U} \tag{3.5}$$

Proposition 2 MATRIX REPRESENTATION OF THE REDUCED PROBLEM. *The power allocation problem (3.5) may be written*

$$\begin{cases} (\mathbf{1} - A)\, P \geq a \\ P \geq 0 \end{cases} \tag{3.6}$$

where

$$A_{uv} = \alpha_{uv} \sum_{m\in u} L_{u,m}/L_{v,m}\xi'_m, \quad u, v \in \mathbf{U} \tag{3.7}$$

$$a_u = P'_u + N \sum_{m\in u} L_{u,m}\xi'_m, \quad u \in \mathbf{U} \tag{3.8}$$

and $A = [A_{uv}]$, $a = [a_u]$ *and* $P = (P_u)^T$.

The power allocation problem above is feasible iff

$$\boldsymbol{\rho}\,(A) < 1 \tag{3.9}$$

in which case

$$P^* = (\mathbf{1} - A)^{-1}\, a \tag{3.10}$$

is the minimal solution.

The condition (3.9) is a NECESSARY AND SUFFICIENT FEASIBILITY CONDITION (abbreviated by NSFC).

Proof. The first part of the proposition is just a matrix representation. (The idea of this "reduced problem" is from [16] and we just extend it to account for the effect of common channels.)

Corollary 12 gives the last part of the proposition. ■

Original versus reduced problem.

Proposition 3 *The power allocation problem (3.2) is feasible iff the reduced problem (3.6) is feasible. In this case their respective minimal solutions* $\mathbb{P}^*$ *and* P^* *are related by*

$$P^*_u = P'_u + \sum_{m\in u} P^*_{u,m}, \quad u \in \mathbf{U} \tag{3.11}$$

and

$$P^*_{u,m} = \left(N + \sum_{v \in \mathbf{U}} \alpha_{uv} P^*_v / L_{v,m} \right) L_{u,m} \xi'_m, \quad m \in u \in \mathbf{U} \tag{3.12}$$

Proof. (i) Suppose that Problem (3.2) is feasible and let $\mathbb{P}^* = (\mathbf{1} - \mathbb{A})^{-1} \mathbf{a}$ be the minimal solution. Let $P^* = \{P^*_u; u \in \mathbf{U}\}$ be defined by (3.11). From the fact that $(\mathbf{1} - \mathbb{A}) \mathbb{P}^* = \mathbf{a}$ which may be written $\mathbb{P}^* = \mathbf{a} + \mathbb{A}\mathbb{P}^*$, we get (3.12). Adding the equalities (3.12) over the set $\{m \in u\}$, we get

$$\sum_{m \in u} P^*_{u,m} = \sum_{m \in u} \left(N + \sum_{v \in \mathbf{U}} \alpha_{uv} P^*_v / L_{v,m} \right) L_{u,m} \xi'_m, \quad u \in \mathbf{U} \tag{3.13}$$

The left term equals $P^*_u - P'_u$. Then P^* is a solution of the reduced problem (3.6). It is easy to see that in fact it is the minimal one.

(ii) Suppose now that the reduced problem (3.6) is feasible and let $P^* = (\mathbf{1} - A)^{-1} a$ be the minimal solution. Let $\mathbb{P}^* = \left\{ P^*_{u,m}; m \in u \in \mathbf{U} \right\}$ be defined by (3.12). Adding the equalities (3.12) over the set $\{m \in u\}$, we get (3.13). Observe that the right-hand side of (3.13) equals $P^*_u - P'_u$ since P^* satisfies $P^* = a + AP^*$. Hence we get $\sum_{m \in u} P^*_{u,m} = P^*_u - P'_u$. Replacing P^*_v by $P'_v + \sum_{n \in v} P^*_{v,n}$ in the right-hand side of (3.12) shows that $\mathbb{P}^*$ is the minimal solution of Problem (3.2). (Note that [16] defines a local and a global problem and shows that the power allocation feasibility is equivalent to the feasibility of both the local and global problems. We show in Proposition 3 that there is no need to introduce the local problem.) ■

The above proposition shows that the condition for the feasibility of the reduced problem should be the same as that for the original problem. Therefore

$$\boldsymbol{\rho}(\mathbb{A}) < 1 \Leftrightarrow \boldsymbol{\rho}(A) < 1$$

In fact we have the following stronger result.

Proposition 4 *We have*

$$\boldsymbol{\sigma}(\mathbb{A}) \setminus \{0\} = \boldsymbol{\sigma}(A) \setminus \{0\}$$

where $\boldsymbol{\sigma}(\mathbb{A})$ and $\boldsymbol{\sigma}(A)$ designate the sets of eigenvalues of $\mathbb{A}$ and A respectively; and in particular

$$\boldsymbol{\rho}(\mathbb{A}) = \boldsymbol{\rho}(A)$$

Proof. From Equation (3.3) we deduce that

$$\mathbb{A}^T_{m,n} = \alpha_{vu} L_{v,n} / L_{u,n} \xi'_n, \quad m \in u, n \in v$$

We know that $\mathbb{A}$ and $\mathbb{A}^T$ have the same eigenvalues. Consider a eigenvector x of $\mathbb{A}^T$ corresponding to a non zero eigenvalue λ. Then

$$\sum_{v \in \mathbf{U}} \alpha_{vu} \sum_{n \in v} L_{v,n} / L_{u,n} \xi'_n \mathbf{x}_n = \lambda \mathbf{x}_m, \quad m \in u$$

So x_m depends only on the base station u serving the user m. Denote the common value x_u. Then

$$\sum_{v\in\mathbf{U}}\alpha_{vu}\sum_{n\in v}L_{v,n}/L_{u,n}\xi'_n x_v=\lambda x_u$$

Recall that is given by (3.7), then the above equation may be written

$$\sum_{v\in\mathbf{U}}A_{vu}x_v=\lambda x_u$$

Hence $A^T x=\lambda x$, which means that $x=(x_u)^T$ is an eigenvector of A^T corresponding to the eigenvalue λ.

Inversely from an eigenvector $x=(x_u)^T$ of A^T corresponding to a non zero eigenvalue λ we construct a vector x= $(\mathbf{x}_m)^T$ by $x_m=x_u$. It is easy to see that x is an eigenvector of $\mathbb{A}^T$ corresponding to the eigenvalue λ.

Then $\mathbb{A}^T$and A^T have the same non zero eigenvalues. We deduce that $\mathbb{A}$ and A have the same non zero eigenvalues. ■

Power limitation. In the case where there is a power limitation constraint, the power allocation problem becomes

$$\begin{cases}(\mathbf{1}-A)\,P\geq a\\ 0\leq P\leq\tilde{P}\end{cases}\tag{3.14}$$

where $\tilde{P}=\left(\tilde{P}_u\right)^T$ designates the vector of the base station power limits.

Proposition 5 *The power allocation problem above is feasible iff*

$$\boldsymbol{\rho}(A)<1\quad\textit{and}\ (\mathbf{1}-A)^{-1}a\leq\tilde{P}\tag{3.15}$$

In this case, the minimal solution is $P^*=(\mathbf{1}-A)^{-1}a$.

The condition (3.15) is a NECESSARY AND SUFFICIENT FEASIBILITY CONDITION (abbreviated by NSFC).

Proof. Immediate from Proposition 2. ■

3.1.2 Feasibility conditions

Without power limit. We use the fact that the spectral radius of a matrix is lower than the maximum row sum [85, Exercice 8.2.7] to establish a sufficient condition for the feasibility of the power allocation problem (3.6).

Proposition 6 *If*

$$\sum_{m\in u}\xi'_m\sum_{v\in\mathbf{U}}\alpha_{uv}L_{u,m}/L_{v,m}<1,\quad u\in\mathbf{U}\tag{3.16}$$

then the power allocation problem (3.6) is feasible.

The inequality in the above proposition is called DOWNLINK FEASIBILITY CONDITION abbreviated by DFC (or simply FC when there is no need to precise "for the downlink without power limit").

Proof. The row sums of A may be written as follows

$$\sum_{v \in \mathbf{U}} A_{uv} = \sum_{m \in u} \xi'_m \sum_{v \in \mathbf{U}} \alpha_{uv} L_{u,m} / L_{v,m}, \quad u \in \mathbf{U}$$

The spectral radius $\boldsymbol{\rho}(A)$ is less than the largest row sum of A which is less than 1 under DFC. Then DFC implies $\boldsymbol{\rho}(A) < 1$ which implies that power allocation problem is feasible by Proposition 2. (The idea of this sufficient feasibility condition is from [16].) ■

With power limit. We shall now establish a sufficient condition for the feasibility of the power allocation problem with power limit.

Proposition 7 *If*

$$(\mathbf{1} - A) \tilde{P} \geq a \tag{3.17}$$

then the power allocation problem (3.14) is feasible and admits $P^* = (\mathbf{1} - A)^{-1} a$ *as the minimal solution. The above inequality is equivalent to*

$$\sum_{m \in u} \xi'_m L_{u,m} \left(\sum_{v \in \mathbf{U}} \alpha_{uv} \tilde{P}_v / L_{v,m} + N \right) \leq \tilde{P}_u - P'_u, \quad u \in \mathbf{U} \tag{3.18}$$

The inequality in the proposition above which is called EXTENDED DOWNLINK FEASIBILITY CONDITION, abbreviated by EDFC or simply FC when there is no need to precise "for the downlink with power limit".

Proof. Since the vector $\tilde{P}$ is non-negative and satisfies (3.17) we deduce that $\boldsymbol{\rho}(A) < 1$ and hence $(\mathbf{1} - A)^{-1}$ is non-negative. Then $\tilde{P} = (\mathbf{1} - A)^{-1} (\mathbf{1} - A) \tilde{P} \geq (\mathbf{1} - A)^{-1} a$. Hence (3.15) is satisfied which finishes the first claim of the proposition.

Condition (3.17) may be written as follows

$$a_u \leq \left(\tilde{P}_u - \sum_{v \in \mathbf{U}} A_{uv} \tilde{P}_v \right), \quad u \in \mathbf{U}$$

which is equivalent to EDFC. ■

In the case where $\tilde{P}_u = \tilde{P}$ is the same for all base stations, EDFC becomes

$$\sum_{m \in u} \left(\sum_{v \in \mathbf{U}} \alpha_{uv} L_{u,m} / L_{v,m} + N L_{u,m} / \tilde{P} \right) \xi'_m \leq 1 - P'_u / \tilde{P}, \quad u \in \mathbf{U} \tag{3.19}$$

If $\tilde{P} \to \infty$ we find

$$\sum_{m \in u} \sum_{v \in \mathbf{U}} \alpha_{uv} L_{u,m} / L_{v,m} \xi'_m \leq 1, \quad u \in \mathbf{U}$$

analogous to DFC (the inequality is strict there).

Remark 1 *We will build in Section 3.3 load control schemes based on DFC and EDFC. Since DFC and EDFC include the users served by the base station but not the users served by other base stations, we say that they are* DECENTRALIZED *(whereas the NSFC is* CENTRALIZED*).*

3.2 Uplink

We consider here the uplink, i.e. the link from the users to the base stations.

The uplink is extensively addressed in literature [126], [57], [59], [46], [90].

3.2.1 Power allocation problem

Similarly to the downlink, we formulate in the following proposition the power allocation problem in the uplink for given user population (fixed positions $\{m\}$ and SINR thresholds $\{\xi_m\}$).

Proposition 8 MATRIX REPRESENTATION. *The power allocation problem is feasible if there exist powers $\{P_m \in \mathbb{R}^+; m \in \mathbf{M}\}$ such that*

$$\frac{P_m/L_{u,m}}{N + \sum_{v\in\mathbf{U}} \alpha_{uv} \sum_{n\in v\setminus\{m\}} P_n/L_{u,n}} \geq \xi_m, \quad m \in u \in \mathbf{U} \tag{3.20}$$

The above problem may be written

$$\begin{cases} (\mathbf{1} - \mathbb{B})\,\mathbb{P} \geq \mathbf{b} \\ \mathbb{P} \geq 0 \end{cases} \tag{3.21}$$

where the matrix $\mathbb{B}=[\mathbb{B}_{m,n}]$ is given by

$$\mathbb{B}_{m,n} = \alpha_{uv}\xi'_m L_{u,m}/L_{u,n}, \quad m \in u \in \mathbf{U}, n \in v \in \mathbf{U} \tag{3.22}$$

the vector $\mathbf{b} = (\mathbf{b}_m)^T$ is given by

$$\mathbf{b}_m = \xi'_m N L_{m,u}, \quad m \in u \in \mathbf{U}$$

and the vector $\mathbb{P} = (P_m)^T$.

The above power allocation problem is feasible iff

$$\boldsymbol{\rho}(\mathbb{B}) < 1$$

in which case

$$\mathbb{P}^* = (\mathbf{1} - \mathbb{B})^{-1}\mathbf{b} \tag{3.23}$$

is the minimal solution.

Proof. The signal transmitted by a user n is received at base station u with power

$$\mathbf{S}_{n,u} = P_n/L_{u,n}$$

Hence for a user $m \in u$, the usefull signal power at u is

$$\mathbf{S}_{m,u} = P_m / L_{u,m}$$

and the interference and noise power is given by

$$N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v \setminus \{m\}} P_n / L_{u,n}$$

Hence, the signal-to-interference-and-noise ratio equals

$$\frac{P_m / L_{u,m}}{N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v \setminus \{m\}} P_n / L_{u,n}}$$

which finishes the proof of (3.20). (When we neglect the noise term $N = 0$, (3.20) is similar to [126, Equation (3) and Definition §III].)

We rearrange the inequality (3.20) as follows

$$P_m \geq N L_{m,u} \xi_m + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v \setminus \{m\}} L_{u,m} / L_{u,n} \xi_m P_n$$

which may be written as

$$P_m \geq N L_{m,u} \xi'_m + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} L_{u,m} / L_{u,n} \xi'_m P_n$$

Then it is easy to see that Problem (3.20) may be written in the form (3.21).

Corollary 12 gives the last part of the proposition. ■

Corollary 2 Achievable SINR targets. *Assume that the ξ_m are constant, i.e. $\xi_m = \xi$. We say that some SINR target is achievable if the power allocation problem for that value of SINR is feasible. The set of achievable SINR targets is*

$$0 \leq \xi < \frac{1}{\boldsymbol{\rho}(\mathbb{B}') - \alpha}$$

where the matrix $\mathbb{B}' = \left[\mathbb{B}'_{m,n}\right]$ *is given by*

$$\mathbb{B}'_{m,n} = L_{u,m} / L_{u,n}, \quad m \in u \in \mathbf{U}, n \in \mathbf{M}$$

Proof. We may write $\mathbb{B} = \xi' \mathbb{B}'$. By Proposition 8, the power allocation problem is feasible iff

$$\xi' \boldsymbol{\rho}(\mathbb{B}') < 1$$

Note that $\mathbb{B}'_{m,m} = \alpha$ then $\boldsymbol{\rho}(\mathbb{B}') \geq \sum_{n \in \mathbf{M}} \mathbb{B}'_{m,n} > \alpha$. Since $\xi' = \xi / (1 + \alpha \xi)$, we get the desired result. (The author [126] gives a similar result.) ■

Comparison of uplink and downlink. We shall now compare the uplink and the downlink. The matrices playing an important role in the power allocation problem are denoted by $\mathbb{A}$ and $\mathbb{B}$ and are given by (3.3) and (3.22) for downlink and uplink respectively. Note that $\mathbb{B}$ is generally different from $\mathbb{A}^T$. Nevertheless the following proposition proves that, in some cases, they have the same eigenvalues, and therefore the same spectral radius.

Proposition 9 *If the orthogonality factor α and the SINR targets $\{\xi_m\}$ are the same for uplink and downlink*[2]*, then the matrices $\mathbb{A}$ and $\mathbb{B}$ defined by (3.3) and (3.22) respectively have identical eigenvalues. In particular, they have the same spectral radius, that is $\boldsymbol{\rho}(\mathbb{A}) = \boldsymbol{\rho}(\mathbb{B})$.*

Proof. Consider the diagonal matrix defined by

$$\mathbf{D}_{m,n} = L_{u,m}\xi'_m\delta_{m,n}, \quad m \in u \in \mathbf{U}, n \in \mathbf{M}$$

where $\delta_{m,n}$ designates the KRONECKER SYMBOL defined by

$$\delta_{m,n} = \begin{cases} 1 & \text{if } m = n \\ \alpha & \text{if } m \neq n \end{cases}$$

Using the expression (3.22) of the matrix $\mathbb{B}$,

$$\mathbb{B}_{m,n} = \alpha_{uv} L_{u,m}/L_{u,n}\xi'_m, \quad m \in u \in \mathbf{U}, n \in v \in \mathbf{U}$$

we deduce that

$$\left(\mathbf{D}^{-1}\mathbb{B}\mathbf{D}\right)_{m,n} = \mathbf{D}^{-1}_{m,m}\mathbb{B}_{m,n}\mathbf{D}_{n,n} = \alpha_{uv}\xi'_n L_{v,n}/L_{u,n}, \quad m \in u \in \mathbf{U}, n \in v \in \mathbf{U}$$

Denote $\tilde{\mathbb{B}} = \mathbf{D}^{-1}\mathbb{B}\mathbf{D}$. The characteristic polynomials of $\tilde{\mathbb{B}}$ and $\mathbb{B}$ are identical.

On the other hand, from the expression (3.3) of the matrix $\mathbb{A}$, we get

$$\mathbb{A}_{n,m} = \alpha_{uv}\xi'_n L_{v,n}/L_{u,n}$$

Hence $\tilde{\mathbb{B}} = \mathbb{A}^T$. The determinant of the transpose of a matrix equals that of the matrix. This implies that the characteristic polynomials of $\mathbb{A}$ and $\tilde{\mathbb{B}}$ are identical. We deduce that the characteristic polynomials of $\mathbb{A}$ and $\mathbb{B}$ are identical. Then the two matrices have identical eigenvalues.

(A similar result for narrowband systems was given in [127].) ■

Corollary 3 *If the orthogonality factor α and the SINR targets $\{\xi_m\}$ are the same for uplink and downlink, then the uplink power allocation problem (3.21) is feasible iff the downlink power allocation problem (3.2) is feasible.*

Proof. Immediate from Propositions 1, 8, and 9.

(An analogous result is given in [1].) ■

[2] The assumtion that the SINR targets are the same for uplink and downlink is not realistic for non symmetric services where the bit-rate of the downlink is different (generally larger) than that of the uplink.

Remark 2 COMPARISON OF DOWNLINK AND UPLINK CONTINUED. *Assume that $\xi_m \ll 1$ for all user m (which is the case for voice service) for both downlink and uplink. Hence $\xi'_m \simeq \xi_m$ for all user m. In this case $\boldsymbol{\rho}(\mathbb{A})$ is non-decreasing with α, and since, practically, the orthogonality factor for the downlink is smaller than that for the uplink, we deduce that $\boldsymbol{\rho}(\mathbb{A}) \leq \boldsymbol{\rho}(\mathbb{B})$. Hence, in this context, the uplink is the limiting case.*

Reduced problem. We aim to formulate the problem in terms of the total intereference $\{I_u; u \in \mathbf{U}\}$ received at base stations. To this end we rearrange the inequality (3.20) as follows

$$\begin{aligned} P_m/L_{u,m} &\geq \xi_m \left(I_u - \alpha P_m/L_{u,m}\right) \\ (1+\alpha\xi_m)\, P_m/L_{u,m} &\geq \xi_m I_u \\ P_m &\geq \xi'_m L_{u,m} I_u \end{aligned}$$

We rewrite the above inequality for some user $n \in v \in \mathbf{U}$, that is

$$P_n \geq \xi'_n L_{v,n} I_v$$

We divide by $L_{u,n}$ an then add over $n \in v$ and over $v \in \mathbf{U}$, which gives

$$I_u - N \geq \sum_{v\in\mathbf{U}} \sum_{n\in v} L_{v,n}/L_{u,n}\xi'_n I_v$$

We say that the REDUCED POWER ALLOCATION PROBLEM is feasible if there exist antenna interferences $\{I_u \in \mathbb{R}^+; u \in \mathbf{U}\}$ such that

$$I_u \geq N + \sum_{v\in\mathbf{U}} \sum_{n\in v} L_{v,n}/L_{u,n}\xi'_n I_v, \quad u \in \mathbf{U} \tag{3.24}$$

Proposition 10 *(Matrix representation of the reduced problem) The power allocation problem (3.24) may be written*

$$\begin{cases} (\mathbf{1} - B)\, I \geq b, \\ I \geq 0, \end{cases} \tag{3.25}$$

where

$$B_{uv} = \sum_{n\in v} L_{v,n}/L_{u,n}\xi'_n, \quad u, v \in \mathbf{U} \tag{3.26}$$

$$b_u = N, \quad u \in \mathbf{U} \tag{3.27}$$

and $B = [B_{uv}]$, $b = (b_u)^T$ and $I = (I_u)^T$.

The above power allocation problem is feasible iff

$$\boldsymbol{\rho}(B) < 1 \tag{3.28}$$

in which case

$$I^* = (\mathbf{1} - B)^{-1}\, b \tag{3.29}$$

is the minimal solution.

Proof. The first part of the proposition is just a matrix representation. Corollary 12 gives the last part of the proposition.

(Reducing the uplink power allocation problem was already made in [59] where the original problem is called *microscopic* and the reduced one is called *macroscopic*. Unlike [59], we don't make the approximation of including the useful signal in the interference. Moreover we treat the problem with inequalities, whereas the author of [59] considers a system of equalities.) ■

Original versus reduced problem.

Proposition 11 *The power allocation problem (3.21) is feasible iff the reduced problem (3.25) is feasible. In this case their respective minimal solutions* $\mathbb{P}^*$ *and* I^* *are related by*

$$I_u^* = N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} P_n^* / L_{u,n}, \quad u \in \mathbf{U} \tag{3.30}$$

and

$$P_m^* = \xi'_m L_{u,m} I_u^*, \quad m \in u \in \mathbf{U}, \quad m \in u \in \mathbf{U} \tag{3.31}$$

Proof. (i) Suppose that Problem (3.21) is feasible and let $\mathbb{P}^* = (\mathbf{1} - \mathbb{B})^{-1}\mathbf{b}$ be the minimal solution. Arguments analogous to those leading to the reduced power allocation problem (3.24) show that $I^* = \{I_u^*; u \in \mathbf{U}\}$ defined by (3.30) is a solution (in fact the minimal one) of the reduced problem (3.25).

(ii) Suppose now that the reduced problem (3.25) is feasible and let $I^* = (\mathbf{1} - B)^{-1} b$ be the minimal solution. Consider $\mathbb{P}^* = \{P_m^*; m \in \mathbf{M}\}$ given by (3.31). From $(\mathbf{1} - B) I^* = b$ we get

$$I_u^* = N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} L_{v,n} / L_{u,n} \xi'_n I_v^* \tag{3.32}$$

By construction

$$P_n^* = \xi'_n L_{v,n} I_v^*, \quad n \in v \in \mathbf{U}$$

then

$$\sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} P_n^* / L_{u,n} = \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} L_{v,n} / L_{u,n} \xi'_n I_v^*$$

hence, using (3.32)

$$\sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} P_n^* / L_{u,n} = I_u^* - N$$

or equivalently

$$I_u^* = N + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} P_n / L_{u,n}$$

Using the above equality and Equation (3.31), we deduce that

$$P_m = N L_{m,u} \xi'_m + \sum_{v \in \mathbf{U}} \alpha_{uv} \sum_{n \in v} L_{u,m} / L_{u,n} \xi'_m P_n, \quad m \in u \in \mathbf{U}$$

which proves that P_m is the minimal solution of Problem (3.21). ■

The above proposition shows that the condition for the feasibility of the reduced problem should be the same as that for the original problem. Therefore

$$\boldsymbol{\rho}(\mathbb{B}) < 1 \Leftrightarrow \boldsymbol{\rho}(B) < 1$$

In fact we have the following stronger result.

Proposition 12 *We have*

$$\boldsymbol{\sigma}(\mathbb{B}) \setminus \{0\} = \boldsymbol{\sigma}(B) \setminus \{0\}$$

where $\boldsymbol{\sigma}(\mathbb{B})$ and $\boldsymbol{\sigma}(B)$ designate the sets of eigenvalues of $\mathbb{B}$ and B respectively; and in particular

$$\boldsymbol{\rho}(\mathbb{B}) = \boldsymbol{\rho}(B)$$

Proof. Consider the diagonal matrix defined by

$$\mathbf{D}_{m,n} = L_{u,m}\xi'_m \delta_{m,n}$$

Using Equation (3.22) we deduce that

$$\left(\mathbf{D}^{-1}\mathbb{B}\mathbf{D}\right)_{m,n} = \mathbf{D}^{-1}_{m,m}\mathbb{B}_{m,n}\mathbf{D}_{n,n} = L_{v,n}/L_{u,n}\xi'_n$$

Denote $\tilde{\mathbb{B}} = \mathbf{D}^{-1}\mathbb{B}\mathbf{D}$. The characteristic polynomials of $\tilde{\mathbb{B}}$ and $\mathbb{B}$ are identical. Then the two matrices have identical eigenvalues. Consider a eigenvector $\tilde{\mathbf{x}}$ of $\tilde{\mathbb{B}}$ corresponding to a non zero eigenvalue λ. Then

$$\sum_{n \in v} L_{v,n}/L_{u,n}\xi'_n \tilde{\mathbf{x}}_n = \lambda \tilde{\mathbf{x}}_m$$

So $\tilde{\mathbf{x}}_m$ depends only on the base station u serving the user m. Denote the common value x_u. We can decompose the sum in the previous equation

$$\sum_{v \in \mathbf{U}} \sum_{n \in v} L_{v,n}/L_{u,n}\xi'_n \tilde{\mathbf{x}}_n = \lambda \tilde{\mathbf{x}}_m$$

Then

$$\sum_{v \in \mathbf{U}} \sum_{n \in v} L_{v,n}/L_{u,n}\xi'_n x_v = \lambda x_u$$

Using (3.26), we get $\sum_{v \in \mathbf{U}} B_{uv} x_v = \lambda x_u$ hence $Bx = \lambda x$. This means that $x = (x_u)^T$ is an eigenvector of B corresponding to the eigenvalue λ.

Inversely from an eigenvector $x = (x_u)^T$ of B corresponding to a non zero eigenvalue λ we construct a vector $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_m)^T$ by $\tilde{\mathbf{x}}_m = x_u$. It is easy to see that $\tilde{\mathbf{x}}$ is an eigenvector of $\tilde{\mathbb{B}}$ corresponding to the eigenvalue λ.

Then $\tilde{\mathbb{B}}$ and B share the same non zero eigenvalues. We deduce that $\mathbb{B}$ and B share the same non zero eigenvalues. (Making the approximation that the useful signal is part of interference, the author [59] proves a similar result.) ■

Comparison of uplink and downlink revisited. Assume that the orthogonality factor α and the SINR targets $\{\xi_m\}$ are the same for uplink and downlink, then from (3.7) and (3.26), we get

$$A_{uv} = \sum_{m \in u} L_{u,m}/L_{v,m}\xi'_m = B_{vu}$$

Then the matrices for the uplink and downlink are transpose of each other

$$B = A^T$$

Hence $\boldsymbol{\rho}(B) = \boldsymbol{\rho}(A)$. Hence we get another proof of Corollary 3.

Power limitation. In the case where there is a power limitation constraint, the power allocation problem becomes

$$\begin{cases} (\mathbf{1} - \mathbb{B})\,\mathbb{P} \geq \mathbf{b} \\ 0 \leq \mathbb{P} \leq \tilde{\mathbb{P}} \end{cases} \tag{3.33}$$

where $\tilde{\mathbb{P}} = \left(\tilde{\mathbb{P}}_m\right)^T$ designates the vector of user power limits.

Proposition 13 *The power allocation problem (3.33) is equivalent to*

$$\begin{cases} (\mathbf{1} - B)\,I \geq b, \\ 0 \leq I \leq \tilde{I} \end{cases} \tag{3.34}$$

where $\tilde{I} = \left(\tilde{I}_u\right)^T$ *with*

$$\tilde{I}_u = \inf_{m \in u} \frac{\tilde{P}_m}{\xi'_m L_{u,m}} \tag{3.35}$$

The above power allocation problem is feasible iff

$$\boldsymbol{\rho}(B) < 1 \quad \textit{and} \;\; (\mathbf{1} - B)^{-1}\, b \leq \tilde{I} \tag{3.36}$$

In this case, the minimal solution is $I^* = (\mathbf{1} - B)^{-1}\, b$.

The condition (3.36) is a NECESSARY AND SUFFICIENT FEASIBILITY CONDITION (abbreviated by NSFC).

Proof. (i) Suppose that Problem (3.33) is feasible and let $\mathbb{P}^* = (\mathbf{1} - \mathbb{B})^{-1}\,\mathbf{b}$ be its minimal solution (see Corollary 13). Similarly to the proof of Proposition 11, we show that I^* given by Equation (3.30) satisfies $(\mathbf{1} - B)\,I^* = b$ and that Equation (3.31) is satisfied. Then from $\mathbb{P}^* \leq \tilde{\mathbb{P}}$ we get $I^* \leq \tilde{I}$. Then Problem (3.34) is feasible.

(ii) Suppose now that Problem (3.34) is feasible and let I^* be its minimal solution. Similarly to the proof of Proposition 11, we show that $\mathbb{P}^* = (P_m)^T$ defined by (3.31) satisfies $(\mathbf{1} - \mathbb{B})\,\mathbb{P}^* = \mathbf{b}$. From $I^* \leq \tilde{I}$ we get $\mathbb{P}^* \leq \tilde{\mathbb{P}}$. Then Problem (3.33) is feasible.

The last part proof is deduced from Corollary 13. ■

3.2.2 Feasibility conditions

Without power limit. We use the fact that the spectral radius of a matrix is lower than the maximum column sum [85, Exercice 8.2.7] to establish a sufficient condition for the feasibility of the power allocation problem (3.25).

Proposition 14 *If*

$$\sum_{m\in u}\xi'_m\sum_{v\in\mathbf{U}}L_{u,m}/L_{v,m}<1,\quad u\in\mathbf{U} \tag{3.37}$$

then the power allocation problem (3.25) is feasible.

The inequality in the above proposition is called UPLINK FEASIBILITY CONDITION, abbreviated by UFC (or simply FC when there is no need to precise “for the uplink without power limit”).

Proof. The column sums of B may be written as follows

$$\sum_{v\in\mathbf{U}}B_{vu}=\sum_{v\in\mathbf{U}}\sum_{m\in u}L_{u,m}/L_{v,m}\xi'_m,\quad u\in\mathbf{U}$$

The spectral radius $\boldsymbol{\rho}(B)$ is less than the largest column sum of B which is less than 1 under UFC. Then UFC implies $\boldsymbol{\rho}(B)<1$ which implies that power allocation problem is feasible by Proposition 10. (The idea of UFC is inspired from the work for downlink of [16]. We have adapted this idea to the uplink. UFC is different from [57, Equation (8.20)] and [46, Equation (11)] which include the users served by all the base stations. Moreover [46, Equation (11)] is obtained with the following assumptions: the noise is neglected; the received power at base station from some user in its cell is proportional to its SINR threshold.) ∎

Remark 3 *Since UFC is decentralized, Remark 1 also applies here. This may seem surprising, because the classical load control algorithms in the uplink depend on the positions of the users of all the base stations. Intuitively UFC controls the interference caused by the users in each cell on the other base stations. Consequently the interference caused by users in other cells on each base station will be limited enough to assure the existence of a solution for the power allocation problem.*

With power limit. We shall now establish a sufficient condition for the feasibility of the power allocation problem with power limit.

Proposition 15 *If*

$$(\mathbf{1}-B)\,\tilde{I}\geq b \tag{3.38}$$

where $\tilde{I}$ is given by (3.35), then the power allocation problem (3.34) is feasible and admits $I^*=(\mathbf{1}-B)^{-1}b$ as the minimal solution.

Proof. Since the vector $\tilde{I}$ is non-negative and satisfies (3.38) we deduce that $\boldsymbol{\rho}(B) < 1$ and hence $(\mathbf{1} - B)^{-1}$ is non-negative. Then $\tilde{I} = (\mathbf{1} - B)^{-1}(\mathbf{1} - B)\tilde{I} \geq (\mathbf{1} - B)^{-1} b$. Hence (3.36) is satisfied, which finishes the first claim of the proposition. ■

Condition (3.38) may be written as follows

$$\sum_{v\in\mathbf{U}}\sum_{n\in v} L_{v,n}/L_{u,n}\xi'_n \tilde{I}_v \leq \tilde{I}_u - N, \quad u \in \mathbf{U} \tag{3.39}$$

The condition in the above display is called SUFFICIENT FEASIBILITY CONDITION (abbreviated by SFC). It is not decentralized. The following proposition gives a decentralized condition.

Proposition 16 *If for some collection of real numbers $\gamma_{u,v} \geq 0$, such that $\gamma_u = \sum_{v\in\mathbf{U}} \gamma_{u,v} < \infty$ the following two conditions are satisfied*

$$\gamma_u = \tilde{I}_u - N, \quad u \in \mathbf{U} \tag{3.40}$$

where $\tilde{I}$ is given by (3.35), and

$$\frac{N + \gamma_u}{\gamma_{v,u}} \sum_{m\in u} L_{u,m}/L_{v,m}\xi'_m \leq 1, \quad u, v \in \mathbf{U} \tag{3.41}$$

then the power allocation problem (3.34) is feasible.

Proof. Denote $C = [C_{u,v}]$ where

$$C_{u,v} = \frac{\gamma_{u,v}}{N + \gamma_v}$$

Note that Condition (3.41) may be written

$$\sum_{m\in u} L_{u,m}/L_{v,m}\xi'_m \leq C_{v,u}$$

The left term is $B_{v,u}$. Then we have $B \leq C$ coordinate-wise. Note that

$$\left((\mathbf{1} - C)\tilde{I}\right)_u = \tilde{I}_u - \sum_{v\in\mathbf{U}} C_{u,v}\tilde{I}_v = N$$

Then $(\mathbf{1} - C)\tilde{I} = b$. Hence

$$\tilde{I} = b + C\tilde{I} \geq b + B\tilde{I}$$

which may be written $(\mathbf{1} - B)\tilde{I} \geq b$. Proposition 15 finishes the proof. ■

Consider the symmetric case $\gamma_{uv} = \gamma_{vu}$. If we replace the collection of inequalities (3.41) by one condition adding the inequalities up, we get

$$\sum_{m\in u}\sum_{v\in\mathbf{U}} L_{u,m}/L_{v,m}\xi'_m \leq 1 - N/\tilde{I}_u, \quad u \in \mathbf{U} \tag{3.42}$$

The condition in the above display is called EXTENDED UPLINK FEASIBILITY CONDITION abbreviated by EUFC (or simply FC when there is no need to precise "for the uplink with power limit").

Remark 4 *Note that Propositions 15 and 15 remain true if we replace $\tilde{I}$ by some $\hat{I} \leq \tilde{I}$. Each such $\hat{I}$ gives a different version of SFC and EUFC. Our numerical applications are made with the version of SFC and EUFC obtained by replacing $\tilde{I}$ with $\hat{I}$ given by*

$$\hat{I}_u = \frac{\tilde{P}}{L(R)\tilde{\xi}'}$$

where $\tilde{P} = \inf_m \tilde{P}_m$ and $\tilde{\xi}' = \sup_m \xi'_m$ where the infimum and the supremum are taken over all the possible values (not only over the set of users currently served by the base station). This permits to make the right-hand side of EUFC (3.42) independent of the particular set of users currently served by the base station.

Remark 5 LINK BUDGET. *It is easy to see that NSFC, SFC and EUFC imply*

$$\tilde{I}_u \geq N, \quad u \in \mathbf{U}$$

The versions of SFC and EUFC described in Remark 4 imply

$$\hat{I}_u \geq N, \quad u \in \mathbf{U}$$

hence, the cell radius R should be less than

$$R_{\max} = \frac{1}{K} \left(\frac{\tilde{P}}{\tilde{\xi}' N} \right)^{1/\eta}$$

3.3 Admission control schemes

Recall that we call NSFC the Necessary and Sufficient Feasibility Condition (Equation (3.15) for the downlink and Equation (3.36) for the uplink). Recall also that we call SFC the uplink Sufficient Feasibility Condition (Equation (3.39) which is not decentralized).

We built in the previous section *decentralized* feasibility conditions of the power allocation problem, denoted by FC which comprise:

- DFC: DOWNLINK Feasibility Condition without power limit (Equation (3.16));
- EDFC: EXTENDED (i.e. with power limit) DOWNLINK Feasibility Condition (Equation (3.18));
- UFC: UPLINK Feasibility Condition without power limit (Equation (3.37));
- EUFC: EXTENDED (i.e. with power limit) UPLINK Feasibility Condition (Equation (3.42)).

Recall that the above conditions, except EUFC, assure the feasibility of the power allocation problem. Note that the FC may be written in the general form (4.13). The FC may be extended to the case of an infinite network as in [116]. Hence our FC are *scalable*, i.e. may be applied to an infinite network in a decentralized way.

Assume the bit rates of all users, or equivalently all the (modified) SINR's $\{\xi'_m\}$, to be specified. The admission control problem can then be formulated as follows. Should a base station admit a new user requesting for admission?

The first idea is to admit the new user if and only if the power allocation problem is feasible, i.e. to use the NSFC as an admission criteria. Unfortunately this is difficult because we should collect information from all the users in the network before taking the admission decision (centralized). Therefore the NSFC is impractical in the field but also very time consuming in simulation tools. Moreover the analytic evaluation of its performance is yet an open problem.

We shall now build admission control schemes for streaming traffic based on the feasibility conditions of the power allocation problem. We call such scheme FEASIBILITY BASED admission control. (Feasibility based congestion control schemes for elastic traffic will be proposed in §10.2.)

FC may be used as an admission criteria. When a new user applies to some base station, the base station accepts it if the respective FC is satisfied with this additional user and rejects it otherwise. This admission scheme inherits the decentralized property of the FC. If each base station applies this admission scheme, then the global power allocation problem is feasible (with high probability for EUFC and certainly for the other schemes).

Chapter 4

First performance evaluation

The objective of the present chapter is to evaluate the performance of the decentralized feasibility conditions established in the previous chapter in a *semi-static* traffic model.

Hypothesis 1 *In this chapter we shall make the following assumptions:*

- *We consider an infinite network on the plan $\mathbb{R}^2$ where base station positions are either hexagonal or Poisson-Voronoi with intensity λ_{S}.*
- *Users are distributed as a Poisson point process with intensity λ_{M}.*
- *Each user is served by the nearest base station.*
- *The power limits $\tilde{P}_u$ and the common channel powers P'_u are the same for all base station $u \in \mathbf{U}$. The power limits $\tilde{P}_m$ are the same for all user $m \in \mathbf{M}$.*

The present chapter is organized as follows. In Section 4.1 we introduce a mean model which permits to define precisely the classical notions of pole capacity and load. In Section 4.2 we define the notion of infeasibility probability and build approximate explicit expression for it.

4.1 Mean model

In this section we introduce a mean model which permits to define precisely the classical notions of pole capacity and load. Besides Hypothesis 1, we assume in the present section that the random entries of A and a in Problems (3.6) and (3.14) (respectively B and b in Problems (3.25) and (3.34)) are replaced by their means $\mathbf{E}[A]$ and $\mathbf{E}[a]$ (respectively $\mathbf{E}[B]$ and $\mathbf{E}[b]$). This defines the MEAN MODEL.

4.1.1 Downlink

Each base station plays the same role as the others, then the row sums of the matrix $\mathbf{E}[A]$ are identical; that is $\sum_{v\in\mathbf{U}}\mathbf{E}[A_{uv}]$ is independent of $u\in\mathbf{U}$. Hence, the spectral radius of the matrix $\mathbf{E}[A]$ equals this row sum

$$\boldsymbol{\rho}(\mathbf{E}[A])=\sum_{v\in\mathbf{U}}\mathbf{E}[A_{uv}], \quad \forall u\in\mathbf{U}$$

If this latter quantity is less than 1, then the vector $(\mathbf{1}-\mathbf{E}[A])^{-1}\mathbf{E}[a]$ has indentical components which are given by

$$\dot{P}=\frac{\mathbf{E}[a_u]}{1-\sum_{v\in\mathbf{U}}\mathbf{E}[A_{uv}]} \tag{4.1}$$

Is is easy to see that

$$\sum_{v\in\mathbf{U}}\mathbf{E}[A_{uv}]=\bar{\xi}'\left(\alpha+\bar{f}\right)\bar{M}$$

$$\mathbf{E}[a_u]=\bar{\xi}'\bar{M}N\bar{l}L(R)+P'$$

where we denote

$$\bar{l}=\mathbf{E}[l(\mathbf{r})]=\mathbf{E}[L(\mathbf{r})]/L(R)$$

and

$$\bar{M}=\lambda_{\mathrm{M}}/\lambda_{\mathrm{S}}=\lambda_{\mathrm{M}}\pi R^2$$

From (4.1) we get

$$\dot{P}=\frac{\bar{\xi}'\bar{M}N\bar{l}L(R)+P'}{1-\bar{\xi}'\left(\alpha+\bar{f}\right)\bar{M}} \tag{4.2}$$

DFC. For the mean model, DFC (3.16) is not only sufficient but also necessary, and it has the form

$$\bar{M}<\Gamma=\frac{1}{\bar{\xi}'\left(\alpha+\bar{f}\right)} \tag{4.3}$$

where the right-hand side of the above display, denoted Γ, is called the POLE CAPACITY. If we apply a scale on the network the pole capacity Γ remains unchanged, hence Γ is independent of the cell radius R. We define the LOAD as

$$\bar{\theta}=\frac{\bar{M}}{\Gamma}=\bar{\xi}'\left(\alpha+\bar{f}\right)\bar{M}$$

which equals the spectral radius of the mean matrix $\mathbf{E}[A]$, that is $\bar{\theta}=\boldsymbol{\rho}(\mathbf{E}[A])$.

Note that the average number of users per cell equals $\bar{M}=\pi R^2\lambda_{\mathrm{M}}$. Then Inequality (4.3) illustrates the phenomenon called CELL BREATHING: If the traffic intensity λ_{M} increases the cell sizes shrink in order to keep the average number of users per cell constant.

Remark 6 *The only difference in the pole capacity between the hexagonal and the Poisson-Voronoi model is the values of* $\bar{f}$*. For Poisson-Voronoi model, we have (see [16])* $\bar{f} = 2/(\eta - 2)$ *whereas for hexagonal model we have approximately,* $\bar{f} \simeq 0.94/(\eta - 2)$*. (This approximation is established in Annex 13.B.) The pole capacity for hexagonal model is larger than that for Poisson-Voronoi by a factor equal to*

$$\frac{\alpha + 2/(\eta - 2)}{\alpha + 0.94/(\eta - 2)}$$

which is about 1.7.

EDFC. For the mean model, EDFC (3.18) takes the form

$$(\mathbf{1} - \mathbf{E}[A])\,\tilde{P} \geq \mathbf{E}[a]$$

The necessary and sufficient feasibility condition (3.15) is

$$\bar{M} < \frac{1}{\bar{\xi}'\left(\alpha + \bar{f}\right)}, \quad \text{and } \dot{P} \leq \tilde{P}$$

(where $\dot{P}$ is given by (4.1)) which is equivalent to

$$\bar{M} \leq \Gamma = \frac{1 - \epsilon}{\left(\alpha + \bar{f} + \bar{l}L(R)N/\tilde{P}\right)\bar{\xi}'} \tag{4.4}$$

which is equivalent to EDFC. Hence EDFC is not only sufficient but also necessary.

Remark 7 *The only difference between the hexagonal and the Poisson-Voronoi model is in* $\bar{f}$ *and* $\bar{l}$*. These parameters characterize the influence of the network geometry on the relation between* $\bar{M}$ *and* R*, as shown by Inequality (4.4). For hexagonal model, we have (by an easy direct calculation)* $\bar{l} = (1 + \eta/2)^{-1}$*, whereas for Poisson-Voronoi model, we have (see [16])* $\bar{l} = \Gamma(1 + \eta/2)$ *(where* $\Gamma(\cdot)$ *is the gamma function not to be confused with the ressource* Γ *defined in Equation (4.4).)*

We now evaluate the effect of the power limitation by comparing the two conditions (4.3) and (4.4). In the condition (4.4), for relatively small R, it is the fraction $\epsilon = P'/\tilde{P}$ that reduces the pole capacity of the downlink

$$\bar{M} \leq \Gamma \simeq \frac{1 - \epsilon}{\bar{\xi}'\left(\alpha + \bar{f}\right)}$$

We call this case INTERFERENCE-LIMITED case. In this case, the above display shows that the power limitation doesn't play an important role. On the other hand, for large R, the dominant restriction is

$$\bar{M} \leq \Gamma \simeq \frac{(1 - \epsilon)\,\tilde{P}}{\bar{\xi}' N \bar{l} L(R)}$$

We call this case NOISE-LIMITED. In this case, the above display shows that the power limit $\tilde{P}$ plays an important role, since the pole capacity is proportional to $\tilde{P}$.

4.1.2 Uplink

Each base station plays the same role as the others, then $\sum_{v\in\mathbf{U}} \mathbf{E}(B_{uv})$ is independent of u. Then $\boldsymbol{\rho}(B) = \sum_{v\in\mathbf{U}} \mathbf{E}(B_{uv})$. If this latter is less than 1, then the vector $(\mathbf{1} - \mathbf{E}[B])^{-1} \mathbf{E}[b]$ has constant components given by

$$\dot{I} = \frac{\mathbf{E}[b_u]}{1 - \sum_{v\in\mathbf{U}} \mathbf{E}[B_{uv}]} \tag{4.5}$$

Is is easy to see that

$$\sum_{v\in\mathbf{U}} \mathbf{E}[B_{uv}] = \bar{\xi}'\left(1 + \bar{f}\right)\bar{M}$$

$$\mathbf{E}[b_u] = N$$

From (4.5) we get

$$\dot{I} = \frac{N}{1 - \bar{\xi}'\left(1 + \bar{f}\right)\bar{M}} \tag{4.6}$$

Hence

$$\dot{P}_m = \xi'_m L_{u,m} \dot{I} = \xi'_m L_{u,m} \frac{N}{1 - \bar{\xi}'\left(1 + \bar{f}\right)\bar{M}}, \quad m \in u \tag{4.7}$$

UFC. For the mean model, UFC (3.37) is in fact necessary and it takes the form

$$\bar{M} < \Gamma = \frac{1}{\bar{\xi}'\left(1 + \bar{f}\right)}. \tag{4.8}$$

where Γ is called the *pole capacity*. If we apply a scale on the network the pole capacity Γ remains unchanged. Then Γ is independent of the radius R of a cell. We define the *load* as

$$\bar{\theta} = \frac{\bar{M}}{\Gamma} = \bar{\xi}'\left(1 + \bar{f}\right)\bar{M}$$

which equals the spectral radius of the mean matrix $\mathbf{E}[B]$, that is $\bar{\theta} = \boldsymbol{\rho}(\mathbf{E}[B])$.

As for the downlink, Inequality (4.8) illustrates the phenomenon called cell breathing.

Remark 8 *As for the downlink, the only difference in the pole capacity between the hexagonal and the Poisson-Voronoi model is the values of $\bar{f}$. The capacity for hexagonal model will be larger than that for Poisson-Voronoi by a factor equal to*

$$\frac{1 + 2/(\eta - 2)}{1 + 0.94/(\eta - 2)}$$

which is about 1.4.

EUFC. For the mean model, SFC (3.38) takes the form

$$\left(\mathbf{1}-\mathbf{E}\left[B\right]\right)\tilde{I}\geq\mathbf{E}\left[b\right]$$

The necessary and sufficient feasibility condition (3.36) is

$$\bar{M}<\frac{1}{\bar{\xi}'\left(1+\bar{f}\right)},\quad \text{and } \dot{I}\leq\tilde{I} \tag{4.9}$$

which is equivalent to

$$\bar{M}\leq\Gamma=\frac{1-N/\tilde{I}_{u}}{\bar{\xi}'\left(1+\bar{f}\right)} \tag{4.10}$$

which is equivalent to SFC. Hence SFC is also necessary.

Note that if we take the expectation of the left-hand side of EUFC (3.42), then we get a condition equivalent to the above display.

We summarize the results for the mean model in the proposition below.

Proposition 17 *The feasibility condition for the mean model may be written in the form*

$$\bar{M}\leq\Gamma=\frac{C}{\bar{\varphi}}$$

where $\bar{M}=\lambda_{\mathrm{M}}\pi R^{2}$*;*

$$\bar{\varphi}=\begin{cases}\left(\alpha+\bar{f}\right)\bar{\xi}' & \textit{for DFC, UFC, EUFC}\\ \left[\alpha+\bar{f}+\bar{l}L\left(R\right)N/\tilde{P}\right]\bar{\xi}' & \textit{for EDFC}\end{cases} \tag{4.11}$$

and

$$C=\begin{cases}1 & \textit{for DFC et UFC}\\ 1-\epsilon & \textit{for EDFC}\\ 1-N/\tilde{I}_{u} & \textit{for EUFC}\end{cases} \tag{4.12}$$

4.2 Infeasibility probability

In this section we define the notion of infeasibility probability and build approximate explicit expression for it. To this end, we make Hypothesis 1, and consider only the hexagonal model for base station positions.

Proposition 18 *The feasibility conditions DFC, EDFC, UFC and EUFC may be written in the generic form*

$$S=\sum_{m\in u}\varphi\left(m\right)<C \tag{4.13}$$

where

$$\varphi\left(m\right)=\begin{cases}\left[\alpha+f\left(m\right)\right]\xi'_{m} & \textit{for DFC, UFC, EUFC}\\ \left[\alpha+f\left(m\right)+NL_{u,m}/\tilde{P}\right]\xi'_{m} & \textit{for EDFC}\end{cases},\quad m\in u \tag{4.14}$$

and C *is given by Equation (4.12).*

Proof. The feasibility conditions DFC, EDFC, UFC and EUFC are given respectively by Inequalities (3.16), (3.18), (3.37) and (3.42) which are clearly in the form of Inequality (4.13) where the function $\varphi(\cdot)$ and the parameter C are those given in the proposition.

In fact, the equality in (4.13) is not strict for EDFC and EUFC. Replacing it by a strict inequality gives still a sufficient feasibility condition which looks more severe but has in fact the same performance when the random variable S has a continuous density. ■

The INFEASIBILITY PROBABILITY is defined as follows

$$\mathbf{P}(S \geq C) \tag{4.15}$$

where S is given by (4.13). Due to Hypothesis 1, the infeasibility probability is the same for all base station $u \in \mathbf{U}$.

Since $S = \sum_{m \in u} \varphi(m)$ is a shot noise, its mean and standard deviation are respectively given by

$$\bar{S} = \bar{M}\bar{\varphi}, \quad \sigma_S = \left(\bar{M}\bar{\varphi^2}\right)^{1/2} \tag{4.16}$$

Proposition 19 *The mean of $\varphi(m)$ is given by Equation (4.11). Its second moment is given by*

$$\bar{\varphi^2} = \begin{cases} \left(\bar{f^2} + 2\alpha\bar{f} + \alpha^2\right)\bar{\xi'^2} & \text{for DFC, UFC, EUFC} \\ \left[L(R)^2\,\bar{l^2}N^2/\tilde{P}^2 + \bar{f^2} + \alpha^2 + 2\alpha\bar{f} + 2\left\{\alpha\bar{l} + \bar{fl}\right\}L(R)\,N/\tilde{P}\right]\bar{\xi'^2} & \text{for EDFC} \end{cases}$$

Proof. Consider first DFC, UFC and EUFC. From the expression $\varphi(m) = [\alpha + f(m)]\,\xi'_m$, we obtain the expression of $\bar{\varphi}$ and $\bar{\varphi^2}$ in the proposition.

Consider now EDFC. From the expression $\varphi(m) = \left[\alpha + f(m) + NL_{u,m}/\tilde{P}\right]\xi'_m$ we get the expression of $\bar{\varphi}$ in the proposition. Moreover

$$\begin{aligned} \bar{\varphi^2} &= \mathbf{E}\left[\left(\alpha + f(m) + l(m)\,L(R)\,N/\tilde{P}\right)^2 \xi'^2_m\right] \\ &= \left[\bar{l^2}L(R)^2\,N^2/\tilde{P}^2 + \bar{f^2} + \alpha^2 + 2\alpha\bar{f} + 2\alpha\bar{l}L(R)\,N/\tilde{P} + 2\bar{fl}L(R)\,N/\tilde{P}\right]\bar{\xi'^2} \\ &= \left[L(R)^2\,\bar{l^2}N^2/\tilde{P}^2 + \bar{f^2} + \alpha^2 + 2\alpha\bar{f} + 2\left\{\alpha\bar{l} + \bar{fl}\right\}L(R)\,N/\tilde{P}\right]\bar{\xi'^2} \end{aligned}$$

■

4.2.1 Calculation methods

Note that the parameter C is constant for DFC, UFC and EDFC. Considering the version of EUFC described in Remark 4 assures that the parameter C is also constant for EUFC. Here are some methods to calculate the infeasibility probability:

Complete simulation

The principle of the complete simulation is the following. We choose a discrete set of test intensities (of users) $\lambda_0 < \lambda_1 < \ldots < \lambda_k$ and simulate k independent patterns of Poisson point processes $\mathcal{N}_i$ $(i = 0, \ldots, k)$ with respective intensities λ_0 and $\Lambda_i = \lambda_i - \lambda_{i-1}$ in the considered cell. Let $F_i(C) = 1_{S \geq C}$ be the indicator that the event $\{S \geq C\}$ holds for the population of users $\mathcal{N}_M = \sum_{j=0}^{i} \mathcal{N}_j$. Obviously $\mathbf{E}[F_i] = \mathbf{P}(S \geq C)$ at $\lambda_\mathrm{M} = \lambda_i$ and F_i is increasing in i. The same holds for $F_i^{(n)} = 1/n \sum_{u=1}^{n} F_{i,u}$, where $(F_{i,u}, i = 0, \ldots, k)$, $u = 1, \ldots, n$ are independent copies of $(F_i, i = 0, \ldots, k)$. In addition, $F_i^{(n)}$ converges a.s. to $\mathbf{P}(S \geq C)$ at $\lambda_\mathrm{M} = \lambda_i$ as $n \to \infty$.

Gaussian approximation

We approximate the infeasibity probability by assuming that S has a Gaussian distribution

$$\mathbf{P}(S \geq C) \simeq Q\left(\frac{C - \bar{S}}{\sigma_S}\right) \tag{4.17}$$

where Q is the Gaussian tail distribution function $Q(x) = 1/\sqrt{2\pi} \int_x^\infty e^{-t^2/2} dt$. The above approximation is called the GAUSSIAN APPROXIMATION.

[86] shows that the Gaussian approximation is good if the SKEWNESS

$$g_3 = \frac{\bar{\varphi^3}}{\bar{\varphi^2}^{3/2}} \frac{1}{\sqrt{\bar{M}}}$$

is small compared to the infeasibity probability. We calculate numerically for DFC

$$\frac{\bar{\varphi^3}}{\bar{\varphi^2}^{3/2}} \simeq 1.5$$

We expect the Gaussian approximation to be good when $\bar{M}$ is large enough which is the case for voice service.

Poisson approximation

The POISSON APPROXIMATION consists of replacing $\varphi(m)$ by its mean $\bar{\varphi}$ in the expression $S = \sum_{m \in u} \varphi(m)$ which gives

$$\mathbf{P}(S \geq C) \simeq \mathbf{P}(M \geq C/\bar{\varphi})$$

where M is the number of user in the cell u, which has a Poisson distribution with mean $\bar{M}$. Hence the infeasibility probability may be approximated by the complementary of the cumulative distribution function of a Poisson random variable.

The following lemma justifies the Poisson approximation from the linear regression theory. It shows that $\bar{\varphi}M$ is the best predictor of $\sum_{m \in u} \varphi(m)$ in terms of an affine function of M.

Lemma 2 *Let $S = \sum_{m \in u} \varphi(m)$. Among all variables $Z = aM + b$, where a and b are real numbers, the one that minimizes the error $\mathbf{E}[(Z - S)^2]$ is*

$$\hat{S} = \bar{\varphi} M$$

and the error is then

$$\mathbf{E}[(\hat{S} - S)^2] = \bar{M}\bar{\varphi^2}\left(1 - \rho_{MS}^2\right)$$

where the CORRELATION COEFFICIENT *of M and S, denoted ρ_{MS}, is given by*

$$\rho_{MS} = \frac{\bar{\varphi}}{\sqrt{\bar{\varphi^2}}}$$

Proof. From the linear regression theory we know that among all variables $Z = aM + b$, where a and b are real numbers, the one that minimizes the error $\mathbf{E}[(Z - S)^2]$ is

$$\hat{S} = \bar{S} + \frac{\sigma_{MS}}{\sigma_M^2}(M - \bar{M})$$

where $\sigma_{MS} = \mathbf{E}[\left(M - \bar{M}\right)\left(S - \bar{S}\right)]$ is the covariance of M and S and σ_M^2 is the variance of M.

Since M is a Poisson random variable, we have $\sigma_M^2 = \bar{M}$. The random variables S and M may be viewed as shot noises. From general results for shot noises, we get

$$\sigma_S^2 = \bar{M}\bar{\varphi^2}, \quad \bar{S} = \bar{\varphi}\bar{M}, \quad \sigma_{MS} = \bar{\varphi}\bar{M}$$

Hence

$$\hat{S} = \bar{\varphi}\bar{M} + \frac{\bar{\varphi}\bar{M}}{\bar{M}}(M - \bar{M}) = \bar{\varphi}M$$

The regression error equals

$$\mathbf{E}[(\hat{S} - S)^2] = \sigma_S^2(1 - \rho_{MS}^2)$$

where ρ_{MS} is the correlation coefficient of M and S defined by

$$\rho_{MS} := \frac{\sigma_{MS}}{\sigma_M \sigma_S} = \frac{\bar{\varphi}\bar{M}}{\sqrt{\bar{M}}\sqrt{\bar{M}\bar{\varphi^2}}} = \frac{\bar{\varphi}}{\sqrt{\bar{\varphi^2}}}$$

■

4.3 Numerical results

Unless otherwise specified, all the numerical applications are made using the default values specified in Section 2.2.5. The versions of SFC and EUFC used in the calculations are those described in Remark 4.

Table 4.1 gives the pole capacity for the hexagonal model without power limit (DFC for the downlink and UFC for the uplink).

FC\service	voice	data 64kbps([1])	data 144kbps	data 384kbps
DFC	38.0	12.3	7.9	3.4
UFC	38.7	15.8	10.2	4.4

Table 4.1: Pole capacity for the hexagonal model without power limit (DFC for the downlink and UFC for the uplink).

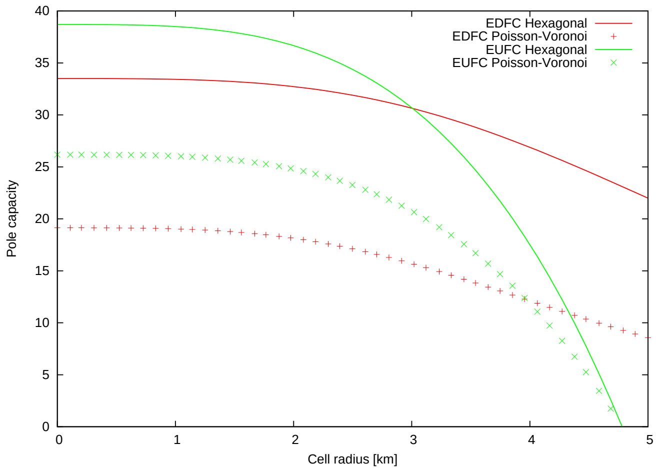

Figure 4.1: Pole capacity for the hexagonal and Poisson-Voronoi models with power limit (EDFC for the downlink and EUFC for the uplink).

Figure 4.1 represents the pole capacity for the hexagonal and Poisson-Voronoi models with power limit (EDFC for the downlink and EUFC for the uplink). This figure shows that the pole capacity of the Poisson-Voronoi model is about 2/3 that of the hexagonal model. For a each model, the pole capacity of the downlink and the uplink depend on the cell radius. For small cell radii the uplink pole capacity is larger than the downlink pole capacity, whereas for large cell radii the situation is reversed. Hence the answer to the question "which link is limiting: the downlink or the uplink?" depends on the cell radius (and the other radio parameters).

Figure 4.2 represents the infeasibility probability obtained with complete simulation, Gaussian approximation and Poisson approximation for EDFC. Visual inspection of this figure shows that the Gaussian and Poisson approximations are both good and that the Gaussian one is better.

We have also made the comparison between complete simulation, Gaussian approximation and Poisson approximation for data 64, 144 and 384kbps respectively. The results are similar to the voice case. In particular the Gaussian and Poisson approximations are both good. It is surprising that the Gaussian approximation is good even for small $\bar{M}$ as for data service 384kbps.

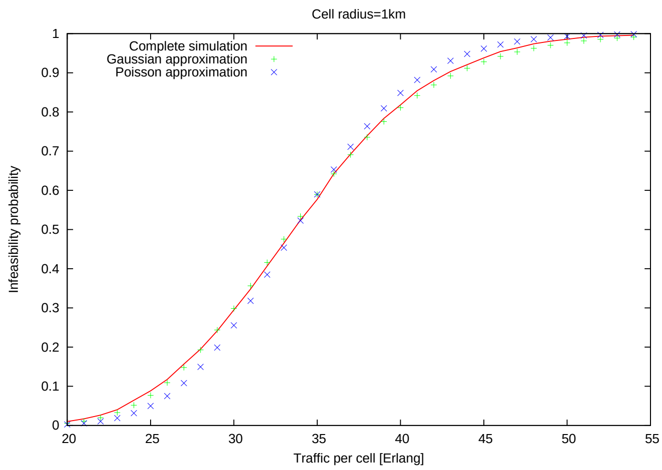

Figure 4.2: EDFC. (Similar results are obtained for other cell radii and for DFC,UFC and EUFC.)

Part II

Spatial Markov Queueing

Chapter 5

Introduction

The present part focuses on the second objective described in Section 0.2, i.e. to develop a general enough model for wireless cellular networks to account for their geometry and dynamics and permitting an analytic evaluation of their performance. In this perspective, mush as in [106, 67], we introduce a *Spatial Markov Queueing* (SMQ) process which is a *pure jump Markov process* that takes its values in the space of *finite point measures* on some general, complete, separable, metric space. We think of these point measures as describing locations of some *individuals* in the space. The SMQ process evolves by individuals being born, moving or dying individually, with only one such event being possible at a time. The process is defined by its generator, which describes the behavior of each individual by a common, fixed *Markov routing kernel* and *departure-arrival rates* that are supposed to be the intensities at which this individual is "repulsed" from its present location and is "attracted" by a new one. These rates possibly depend on the entire, actual configuration of calls. Some special cases of the SMQ processes are *Spatial Birth-and-Death* (SBD) processes, where individuals do not move in the space, and *Markov Poisson Location* (MPL) processes where individuals are being born, move and die independently of each other. Spatial queueing Jackson and Whittle networks are special cases too.

5.1 Related works

The SMQ process has already been studied. In particular Preston [97] establishes sufficient conditions for the regularity (i.e. non-explosion) and ergodicity of the SBD processes. Iglehart [68] establishes such conditions for discrete SMQ processes. Serfozo introduces spatial Whittle processes in [106, 67], SMQ processes in [107, 66], and some extensions of SMQ in [108, 109].

The regularity of the SMQ generator is not adressed by Serfozo. We study this question by extending Preston's arguments.

Our approach to ergodicity is inspired by Preston's paper and is different from Serfozo's work. In this latter, sufficient conditions for the null measure to be positive recurrent are found by comparing the process to a M/GI/∞ queueing system, via a MPL process with some modified system of traffic equations. Our method is more general, and our sufficient conditions for ergodicity seem to be less constraining in cases when both approaches can be applied. In particular we do not need uniformly bounded arrival rates.

Preston gave the explicit form of the invariant measure of SBD processes. The invariant measure of general SMQ processes is due to the pioneering work of Serfozo. For the sake of completeness, we recall the results in our report and give proofs based on results of point process theory (Proposition 20).

Stability of some general (not necessarily Markov) spatial queueing systems, where the individuals are motionless, is studied in [27].

5.2 Our contribution

We give sufficient conditions for the SMQ generator to be *regular* and *ergodic* which may be seen as extensions of Iglehart's conditions [68] to the spatial case and of Preston's conditions [97] to the case with mobility. We prove both regularity and ergodicity by comparing our SMQ process to a discrete birth-and-death process, for which the respective conditions are known (given in [102] and in [72]). More specifically, in order to prove ergodicity, we use a dominating birth-and-death process to give sufficient conditions for the null measure (representing the empty-system) to be positive recurrent. Then the limiting and invariant measure is given by the classical cycle formula.

The spatial component of wireless networks is subject to changes due to the mobility of users. In such a perspective, we believe that SMQ processes are suitable for modeling modern wireless communication systems.

To make this claim evident, we use the SMQ process to model wireless cellular networks serving elastic traffic. We build two mobility models and we give explicit formulae for the delay and average throughput.

We also use our SMQ process to define and analyze two loss models (which in some case cannot be seen as Whittle networks). Our goal in this part is to give possibly explicit formulae for blocking and cuts probabilities for streaming traffic. In particular, in one setting of our models we prove an expression for blocking rates that might be seen as a spatial version of the classical Erlang loss formula.

5.3 Organization

In chapter 6, we introduce and study the SMQ process. In Chapters 7 and 8 we use the SMQ process to model wireless cellular networks serving elastic and streaming traffic respectively.

Chapter 6

Spatial Markov queueing process (SMQ)

The present chapter presents the SMQ process and its basic properties. It is organized as follows. In Section 6.1 we introduce the notation and recall some basic facts concerning point processes and measure valued pure-jump Markov processes. In Section 6.2 we introduce the SMQ generator. In Sections 6.2.2 and 6.3 we give sufficient conditions for it to be regular and ergodic respectively. In section 6.4 we give conditions for its invariant measure to be Gibbsian.

6.1 Preliminaries

6.1.1 Point process

Very much as in [106, 67], we will consider a system in which users are located in a complete, separable metric space $\mathbb{D}$ with its Borel σ-field $\mathcal{D}$. Typically $\mathbb{D}$ would be a bounded subset of some Euclidean space. If $\mathbb{D}$ is a finite set of points, the system is DISCRETE. In the general case, we will represent the state of the system by a finite counting measure ν on $\mathbb{D}$. Suppose that $x_1, \ldots, x_k \in \mathbb{D}$ are the locations of users. These locations can be described by a COUNTING MEASURE ν on $\mathbb{D}$ defined by

$$\nu(A) = \sum_{i=1}^{k} \delta_{x_i}(A), \quad \text{for all } A \in \mathcal{D}$$

where k is some given integer and δ_x is a DIRAC MEASURE with unit mass at x, i.e. $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. As a simple consequence of this notation we have for a given real valued measurable function f on $\mathbb{D}$: $\int f(x)\nu(dx) = \sum_{i=1}^{k} f(x_i)$.

A random configuration N of points at a given time, will be modeled by a POINT PROCESS that is a measurable mapping from some given probability space to the state space $\mathbb{M}$ of all finite counting measures on $\mathbb{D}$ (with the smallest σ-algebra $\mathcal{M}$ making the mappings $\mathbb{M} \ni \nu \mapsto \nu(B)$ measurable for all $B \in \mathcal{D}$).

The MEAN MEASURE $\lambda(\cdot)$ of the point process N is defined as $\lambda(B) = \mathbf{E}[N(B)]$, $B \in \mathcal{D}$; it represents the expected numbers of points present in subsets of $\mathbb{D}$.

Here are two examples of point processes.

Example 1 *The most prominent point processes are* POISSON PROCESSES *defined as follows: Let λ be a finite measure on $(\mathbb{D}, \mathcal{D})$, a point process N is Poisson with mean measure λ if for each $A \in \mathcal{D}$ the random variable $N(A)$ is Poisson with mean $\lambda(A)$ and for all mutually disjoint $A_1, \ldots, A_k \in \mathcal{D}$ the random variables $N(A_1), \ldots, N(A_k)$ are independent.*

Example 2 *Another important class of point processes is that of* GIBBS PROCESSES. *For a given non-negative measurable function $\Psi : \mathbb{M} \to \mathbb{R}_+$ ($\mathbb{R}_+$ denotes the set of non-negative real numbers) and a measure ρ on $\mathbb{D}$, the* GIBBS DISTRIBUTION *on $\mathbb{M}$, with* DENSITY *or* ENERGY FUNCTION *Ψ with respect to the Poisson* WEIGHT PROCESS *N of mean measure ρ, is the distribution Π_Ψ on $\mathbb{M}$ defined by*

$$\Pi_\Psi(\Gamma) = Z^{-1}\mathbf{E}\left[1\left\{N \in \Gamma\right\}\Psi\left(N\right)\right], \quad \textit{for all } \Gamma \in \mathcal{M}$$

where $Z = \mathbf{E}\left[\Psi\left(N\right)\right]$ is the NORMALIZING CONSTANT *assumed to be positive and finite (called also* partition function *or* statistical sum*). The energy function can often be expressed as follows*

$$-\log(\Psi(\nu)) = \sum_{k=1}^{\nu(\mathbb{D})} \mathcal{E}(x_k, \sum_{i=1}^{k-1} \delta_{x_i})$$

where $\nu = \sum_{i=1}^{\nu(\mathbb{D})} \delta_{x_i}$, and where $\mathcal{E} : \mathbb{D} \times \mathbb{M} \to \mathbb{R}$ is called the LOCAL ENERGY FUNCTION.

Proposition 20 *Let N be a Gibbs process on $\mathbb{D}$, with density Ψ with respect to a Poisson weight process with finite measure ρ. For any measurable function $g : \mathbb{D} \times \mathbb{M} \to \mathbb{R}_+$ we have*

$$\mathbf{E}\left[\int_{\mathbb{D}} g\left(x, N - \delta_x\right) N\left(dx\right)\right] = \mathbf{E}\left[\int_{\mathbb{D}} g(x, N)\frac{\Psi(N + \delta_x)}{\Psi(N)}\rho(dx)\right] \tag{6.1}$$

(The term $p(x, N) = \frac{\Psi(N+\delta_x)}{\Psi(N)}$ is called the PAPANGELOU'S EXVISIBLE INTENSITY.*)*

Proof. (The proposition may be deduced from [113, Theorem 5.1, p. 179];

but no proof is provided there.) The left-hand side of (6.1) equals

$$A = \mathbf{E}\left[\int_{\mathbb{D}} g\left(x, N-\delta_x\right) N\left(dx\right)\right] = \mathbf{E}_N\left[\int_{\mathbb{D}} g\left(x, \nu-\delta_x\right) \nu\left(dx\right)\right]$$
$$= \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n+1}} g\left(x, \sum_{j=1}^{n} \delta_{x_j} - \delta_x\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j}\right) \rho\left(dx_1\right)\ldots\rho\left(dx_n\right) \sum_{j=1}^{n} \delta_{x_j}\left(dx\right)$$
$$= \sum_{n=1}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n}} \sum_{i=1}^{n} g\left(x_i, \sum_{j=1, j\neq i}^{n} \delta_{x_j}\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j}\right) \rho\left(dx_1\right)\ldots\rho\left(dx_n\right)$$

The right-hand side of (6.1) equals

$$B = \mathbf{E}\left[\int_{\mathbb{D}} g(x,N)p(x,N)\rho(dx)\right] = \mathbf{E}_N\left[\int_{\mathbb{D}} g\left(x,\nu\right) p\left(x,\nu\right) \rho(dx)\right]$$
$$= \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n+1}} g\left(x, \sum_{j=1}^{n} \delta_{x_j}\right) p\left(x, \sum_{j=1}^{n} \delta_{x_j}\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j}\right) \rho\left(dx_1\right)\ldots\rho\left(dx_n\right)\rho(dx)$$
$$= \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n+1}} g\left(x, \sum_{j=1}^{n} \delta_{x_j}\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j} + \delta_x\right) \rho\left(dx_1\right)\ldots\rho\left(dx_n\right)\rho(dx)$$

With the change of notation $x \to x_i, x_i \to x_{i+1}, \ldots, x_n \to x_{n+1}$, we obtain

$$\int_{\mathbb{D}^{n+1}} g\left(\sum_{j=1}^{n} \delta_{x_j}, x\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j} + \delta_x\right) \rho\left(dx_1\right)\ldots\rho\left(dx_n\right)\rho(dx)$$
$$= \int_{\mathbb{D}^{n+1}} g\left(\sum_{j=1, j\neq i}^{n+1} \delta_{x_j}, x_i\right) \Psi\left(\sum_{j=1}^{n+1} \delta_{x_j}\right) \rho(dx_1)\ldots\rho(dx_{n+1})$$

Therefore

$$B = \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n+1}} \frac{1}{n+1} \sum_{i=1}^{n+1} g\left(x_i, \sum_{j=1, j\neq i}^{n+1} \delta_{x_j}\right) \Psi\left(\sum_{j=1}^{n+1} \delta_{x_j}\right) \rho(dx_1)\ldots\rho(dx_{n+1})$$
$$= \sum_{n=1}^{\infty} \frac{e^{-\rho(\mathbb{D})}}{n!} \int_{\mathbb{D}^{n}} \sum_{i=1}^{n} g\left(x_i, \sum_{j=1, j\neq i}^{n} \delta_{x_j}\right) \Psi\left(\sum_{j=1}^{n} \delta_{x_j}\right) \rho(dx_1)\ldots\rho(dx_n)$$

and therefore $A = B$. ■

6.1.2 Measure-valued Markov process

We will model the temporal evolution of the system of points by a time-homogeneous Markov jump process $N_. = \{N_t; t \geq 0\}$ taking values in the state space $\mathbb{M}$ of all finite counting measures on $\mathbb{D}$. A survey of general results concerning such processes is made in [41].

Let $q(\cdot) : \mathbb{M} \to \mathbb{R}^+$ be $\mathcal{M}$-measurable, and let $q(\cdot,\cdot) : \mathbb{M} \times \mathcal{M} \to \mathbb{R}^+$ be a KERNEL (that is: for each $\Gamma \in \mathcal{M}$, $q(\cdot,\Gamma)$ is a non-negative measurable function on $\mathbb{M}$; and for each $\nu \in \mathbb{M}$, $q(\nu,\cdot)$ is a measure on $\mathcal{M}$).

A stochastic process $N. = \{N_t; t \geq 0\}$ with state space $(\mathbb{M},\mathcal{M})$ is called a MARKOV JUMP PROCESS with (infinitesimal) GENERATOR $(q(\cdot), q(\cdot,\cdot))$ (called also q-PAIR) if, given that the process is in state ν, then the waiting time to the next jump has an exponential distribution with expectation $1/q(\nu)$ (if $q(\nu) = 0$ then the process remains indefinitely in state ν) and is independent of the past history, and the probability that the following jump leads to a value in $\Gamma \in \mathcal{M}$ is $q(\nu,\Gamma)/q(\nu)$ (cf. [48]). The TRANSITION KERNEL $P_t(\nu,\Gamma) = \mathbf{P}\{N_t \in \Gamma | N_0 = \nu\}$ then satisfies the following CHAPMAN-KOLMOGOROV EQUATIONS

$$P_{t+s}(\nu,\Gamma) = \int_{\mathbb{M}} P_s(\mu,\Gamma)P_t(\nu,d\mu) \qquad \nu \in \mathbb{M}, \Gamma \in \mathcal{M}, t,s \geq 0$$
$$P_0(\nu,\Gamma) = 1\{\nu \in \Gamma\}, \quad \nu \in \mathbb{M}, \Gamma \in \mathcal{M}$$

where $1\{A\} = 1$ if A true and 0 otherwise. The transition kernel $\{P_t\}$ also satisfies the following set of BACKWARD KOLMOGOROV EQUATIONS:

$$\frac{\partial P_t(\nu,\Gamma)}{\partial t} = -P_t(\nu,\Gamma)q(\nu) + \int_{\mathbb{M}} P_t(\mu,\Gamma)q(\nu,\mu)d\mu, \quad \nu \in \mathbb{M}, \Gamma \in \mathcal{M} \qquad (6.2)$$
$$P_0(\nu,\Gamma) = 1\{\nu \in \Gamma\}, \quad \nu \in \mathbb{M}, \Gamma \in \mathcal{M}$$

If $q(\nu,\{\nu\}) = 0, \forall \nu \in \mathbb{M}$ (i.e. there is no pseudo-transitions) then we have also the following relations between the transition kernel and the generator

$$q(\nu,\Gamma) = \lim_{t \searrow 0} t^{-1} P_t(\nu, \Gamma \setminus \{\nu\}), \quad \nu \in \mathbb{M}, \Gamma \in \mathcal{M} \qquad (6.3)$$
$$q(\nu) = \lim_{t \searrow 0} t^{-1} \left(1 - P_t(\nu,\{\nu\})\right), \quad \nu \in \mathbb{M}$$

Suppose that the q-pair is STABLE; i.e., that $q(\nu) < \infty$ for all $\nu \in \mathbb{M}$ and CONSERVATIVE; i.e., that $q(\nu) = q(\nu, \mathbb{M} \setminus \{\nu\})$. Given the intial distribution, the q-pair defines the evolution of a Markov jump process $\{N_t^{(\infty)}(\omega) : t \in [0, t_\infty(\omega))\}$ where the EXPLOSION TIME $t_\infty(\omega)$ can be either the first time of accumulation of jumps if such time is finite or otherwise is equal to ∞. We call this process the MINIMAL MARKOV JUMP PROCESS associated to the given q-pair. Given, its initial distribution, the minimal process is unique in distribution. The Markov transition kernel $P_t^\infty(\cdot,\cdot)$ describing the evolution of this process is the so called MINIMAL SOLUTION of the Backward Kolmogorov equations (6.2). (It is minimal in the sense that any other solution $P_t(\cdot,\cdot)$ of (6.2) satisfies $P_t(\nu,\Gamma) \geq P_t^\infty(\nu,\Gamma)$, for all $\nu \in \mathbb{M}, \Gamma \in \mathbb{M}$.) The minimal solution P_t^∞ is SUBSTOCHASTIC, i.e., $P_t^{(\infty)}(\nu,\mathbb{M}) \leq 1$ for all $\nu \in \mathbb{M}$ and $t \geq 0$. The minimal solution P_t^∞ is stochastic iff $t_\infty = \infty$ a.s.. In this case $P_t^{(\infty)}$ is the unique solution of (6.2) (unique among the substochastic kernels which are solutions of (6.2)) and we say that q is REGULAR.

Suppose that q is regular so we can speak about the unique Markov kernel $P_t(\cdot,\cdot)$. We say that a non-null measure Π on $\mathbb{M}$ is INVARIANT for $P_t(\cdot,\cdot)$ if

$$\Pi(\Gamma) = \int_{\mathbb{M}} P_t(\nu,\Gamma)\Pi(d\nu), \quad \Gamma \in \mathcal{M}, t \geq 0 \tag{6.4}$$

For a given $t \geq 0$, the right hand side of (6.4) defines a measure denoted ΠP_t, that is

$$\left(\Pi P_t\right)(\Gamma) = \int_{\mathbb{M}} P_t(\nu,\Gamma)\Pi(d\nu), \quad \Gamma \in \mathcal{M}$$

It is known for a probability measure $\Pi(\cdot)$ that (6.4) is equivalent to the following GLOBAL BALANCE EQUATIONS

$$\int_{\Gamma} q(\nu,\mathbb{M})\Pi(d\nu) = \int_{\mathbb{M}} q(\mu,\Gamma)\Pi(d\mu), \quad \Gamma \in \mathcal{M} \tag{6.5}$$

(see e.g. [34, Theorem 4.17, p.129]).

We call $\nu \in \mathbb{M}$ a POSITIVE RECURRENT state of a Markov jump process $\{N_t; t \geq 0\}$ if $E[T^\nu | N_0 = \nu] < \infty$ where T^ν is the RETURN TIME of $\{N_t\}$ to ν (strictly after the first jump of the process).

We will say that a Markov kernel $P_t(\cdot,\cdot)$ (or the associated process $\{N_t; t \geq 0\}$) is ERGODIC if there exists a probability measure Π satisfying

$$\lim_{t\to\infty} \sup_{\Gamma\in\mathcal{M}} |P_t(\nu,\Gamma) - \Pi(\Gamma)| = 0$$

for all $\nu \in \mathbb{M}$. We say in this case that $P_t(\nu,\cdot)$ converges in TOTAL VARIATION to $\Pi(\cdot)$.

We say that the Markov kernel $P_t(\cdot,\cdot)$ is REVERSIBLE with respect to a non-null measure Π on $\mathbb{M}$ if

$$\int_{\Gamma_1} P_t(\nu,\Gamma)\Pi(d\nu) = \int_{\Gamma} P_t(\nu,\Gamma_1)\Pi(d\nu), \quad \Gamma,\Gamma_1 \in \mathcal{M}, t \geq 0 \tag{6.6}$$

It is known for a probability measure $\Pi(\cdot)$ that (6.6) is equivalent to a set of DETAILED BALANCE EQUATIONS which has the form of (6.5) with $\mathbb{M}$ replaced by Γ_1, for all $\Gamma_1 \in \mathcal{M}$. (see e.g. [34, Theorem 6.7 p.230]).

In what follows, for some $\nu \in \mathbb{M}$, we will write $\mathbf{E}_\nu[\cdot] = \mathbf{E}[\cdot|N_0 = \nu]$ and similarly, for some measure Π on $(\mathbb{M},\mathcal{M})$, we will denote by $\mathbf{E}_\Pi[\cdot]$ the expectation conditionally to N_0 being distributed according to Π. Recall also that for some measurable $\Psi : \mathbb{M} \to (0,\infty)$ and some measure Π on $(\mathbb{M},\mathcal{M})$ it is usual to denote $\mathbf{E}_\Pi[\Psi] = \int_{\mathbb{M}} \Psi(\nu)\,\Pi(d\nu)$. Note that there is a slight conflict of notation which is not harmful since which $\mathbf{E}_\Pi[\cdot]$ we are considering would be clear from the context.

6.2 Spatial Markov queueing process

As in [106, 67], the system consists of a set $\mathbb{D}$ of LOCATIONS which is a complete separable metric space, typically $\mathbb{D}$ is a subset of $\mathbb{R}^d$ induced with the Euclidean distance. We will consider the topology generated by the metric and take $\mathcal{D}$ as the Borel σ-algebra. USERS (or CALLS) move along the locations where they are served. The evolution of the system is represented by a continuous-time stochastic process $N_\cdot = \{N_t\}_{t\geq 0}$ whose *states* are in $\mathbb{M}$, the space of all finite counting measures on $\mathbb{D}$. Let $\mathcal{M}$ be the σ-algebras on $\mathbb{M}$ generated by the mappings $\nu \mapsto \nu(B), B \in \mathcal{D}$. For $\nu \in \mathbb{M}$, we denote $\nu_x = \nu\{x\}$.

Lemma 3 *If $\mathbb{D}$ is countable, then $\mathbb{M}$ is countable.*

Proof. We may assume without loss of generality that $\mathbb{D} = \{1, 2, \ldots\}$. For $y \in \mathbb{D}$, let $\mathbb{M}^{(y)} = \left\{\nu \in \mathbb{N}^{\mathbb{D}} : \nu_x = 0, \forall x \geq y\right\}$. Clearly $\mathbb{M} = \bigcup_{y\in\mathbb{D}} \mathbb{M}^{(y)}$. Using the fact that the countable union of finite sets (and even of countable sets) is countable, proves that $\mathbb{M}$ is countable. ■

The users move within the set of locations

$$\bar{\mathbb{D}} = \mathbb{D} \cup \{o\} \tag{6.7}$$

where the location o designates the outside of the system. We will associated to $\bar{\mathbb{D}}$ the σ-algebra $\bar{\mathcal{D}} = \mathcal{D} \cup \{\Gamma \cup \{o\} : \Gamma \in \mathcal{D}\}$. The system state ν doesn't record any population size for location o. In order to write some equations in a more compact form, we will view each element $\nu \in \mathbb{M}$ as a counting measures on $\bar{\mathbb{D}}$ such that $\nu(\{o\}) = 1$.

6.2.1 Infinitesimal generator

We aim to model the system by a Markov process. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be the underlying probability space. A typical transition of the process will be triggered by the movement of a user from some location x to some location $y \neq x$ in $\bar{\mathbb{D}}$. Then we have a transition from state ν to state $T_{xy}\nu = \nu - \delta_x + \delta_y$ where, by convention, $\delta_o = 0$ and for $x \in \mathbb{D}$, δ_x is the measure with unit mass concentrated at x. The transition is called a MIGRATION or MOTION if $x, y \in \mathbb{D}$, a BIRTH if $x = o$ and a DEATH if $y = o$.

We consider an infinitesimal generator $(q(\cdot), q(\cdot,\cdot))$ given by

$$\begin{cases} q(\nu, \Gamma) = \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} r_{xy}(\nu) 1_\Gamma(T_{xy}\nu) \lambda(x, dy) \nu(dx), & \Gamma \in \mathcal{M}, \nu \in \mathbb{M} \\ q(\nu) = \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} r_{xy}(\nu) \lambda(x, dy) \nu(dx), & \nu \in \mathbb{M} \end{cases} \tag{6.8}$$

where $\lambda : \bar{\mathbb{D}} \times \bar{\mathcal{D}} \to \mathbb{R}^+$ constitute a kernel (i.e. for each $A \in \bar{\mathcal{D}}$, $\lambda(\cdot, A)$ is a non-negative measurable function on $\bar{\mathbb{D}}$; and for each $x \in \bar{\mathbb{D}}$, $\lambda(x, \cdot)$ is a measure on $\bar{\mathcal{D}}$) called the ROUTING RATES and $r : \bar{\mathbb{D}} \times \bar{\mathbb{D}} \times \mathbb{M} \to \mathbb{R}^+, (x, y, \nu) \mapsto r_{xy}(\nu)$ are measurable functions called the SERVICE RATES. We sometimes denote $r_{xy}(\nu, T_{xy}\nu) = r_{xy}(\nu)$. **From now on we assume that the routing rates satisfy**

$$\lambda(x, \{x\}) = 0, \quad x \in \bar{\mathbb{D}} \tag{6.9}$$

(This implies that $q(\nu, \{\nu\}) = 0$.) **and that the service rates satisfy**

$$(x \in \mathbb{D} \text{ and } \nu(\{x\}) = 0) \Rightarrow (\forall y \in \bar{\mathbb{D}}, r_{xy}(\nu) = 0)$$

The first equation in (6.8) is written in a compact form with some abuse of notation. It may be written more explicitly as follows

$$\begin{aligned} q(\nu,\Gamma) &= \int_{\mathbb{D}} r_{oy}(\nu)\, 1_\Gamma(T_{oy}\nu)\, \lambda(o,dy) + \int_{\mathbb{D}} \left[\int_{\bar{\mathbb{D}}} r_{xy}(\nu)\, 1_\Gamma(T_{xy}\nu)\, \lambda(x,dy) \right] \nu(dx) \\ &= \int_{\mathbb{D}} r_{oy}(\nu)\, 1_\Gamma(T_{oy}\nu)\, \lambda(o,dy) + \int_{\mathbb{D}\times\mathbb{D}} r_{xy}(\nu)\, 1_\Gamma(T_{xy}\nu)\, \lambda(x,dy)\, \nu(dx) \\ &\quad + \int_{\mathbb{D}} r_{xo}(\nu)\, 1_\Gamma(T_{xo}\nu)\, \lambda(x,\{o\})\, \nu(dx) \end{aligned} \tag{6.10}$$

The sum in the right-hand side of the above equation comprises three terms which correspond respectively to births, motions and deaths.

Definition 1 *We call the generator (6.8)* SPATIAL MARKOV QUEUEING *(SMQ) or equivalently* SPATIAL BIRTH, MOTION AND DEATHS *generator. It is called* WHITTLE *if the service rates have the form $r_{xy}(\nu) = \psi_x(\nu)$ and* JACKSON *if $\psi_x(\nu)$ is a function of only ν_x denoted, with a slight abuse of notation, $\psi_x(\nu_x)$.*

Remark 9 *The idea of the spatial Whittle generator is due to [106, 67] and that of the general SMQ appears is due to [107, 66].*

Remark 10 *Serfozo [106] uses the term Whittle network only when the service rates $r_{xy}(\nu) = \psi_x(\nu)$ have a certain balance property which will be defined later (Definition 3).*

In order to simplify some equations, we denote

$$T_{AB}\nu = \{T_{xy}\nu : y \neq x, x \in A, y \in B\}, \quad A, B \in \bar{\mathcal{D}}, \nu \in \mathbb{M} \tag{6.11}$$

If A is a singleton $\{x\}$ we write $T_{\{x\}B}\nu$ simply $T_{xB}\nu$. The same notation simplification applies when B is a singleton. Hence the quantity $q(\nu,\Gamma)$ in (6.10) may be written as the sum

$$q(\nu,\Gamma) = q(\nu,\Gamma\cap T_{o\mathbb{D}}\nu) + q(\nu,\Gamma\cap T_{\mathbb{D}\mathbb{D}}\nu) + q(\nu,\Gamma\cap T_{\mathbb{D}o}\nu)$$

where the right-hand side comprises three terms which correspond respectively to the intensity of births, motions and deaths:

$$\begin{cases} q(\nu,\Gamma\cap T_{o\mathbb{D}}\nu) = \int_{\mathbb{D}} r_{oy}(\nu)\, 1_\Gamma(T_{oy}\nu)\, \lambda(o,dy) \\ q(\nu,\Gamma\cap T_{\mathbb{D}\mathbb{D}}\nu) = \int_{\mathbb{D}\times\mathbb{D}} r_{xy}(\nu)\, 1_\Gamma(T_{xy}\nu)\, \lambda(x,dy)\, \nu(dx) \\ q(\nu,\Gamma\cap T_{\mathbb{D}o}\nu) = \int_{\mathbb{D}} r_{xo}(\nu)\, 1_\Gamma(T_{xo}\nu)\, \lambda(x,\{o\})\, \nu(dx) \end{cases}$$

Example 3 *The* MARKOV-POISSON LOCATION *(MPL) process can be seen as a SMQ process where users arrive, move and depart from $\mathbb{D}$ completely independently of each other. Thus, $r_{xy}(\nu) \equiv 1$.*

Example 4 *A* SPATIAL BIRTH-AND-DEATH *(SBD) process is a SMQ process without mobility. Thus $\lambda(x,\mathbb{D}) = 0$, for all $x \in \mathbb{D}$.*

6.2.2 Regularity

We aim now to find sufficient conditions assuring that there exists a unique q-process (or transition functions) associated to the generator q given by (6.8). In this case, given the intial distribution, there will be a unique (in distribution) homogeneous Markov process $N_{\cdot} = \{N_t : t \in \mathbb{R}_+\}$ associated to the generator q. In such a case we could talk about the process defined by the generator (6.8) since this process is uniquely defined (in distribution).

Conservative By definition of the generator (6.8), $q(\nu) = q(\nu, \mathbb{M}\setminus\{\nu\})$ which assures that the generator q is conservative.

Stable Recall that the stability condition for a generator q reads $q(\nu, \mathbb{M}\setminus\{\nu\}) < \infty$. For the SMQ generator (6.8) we may write

$$q(\nu, \mathbb{M}\setminus\{\nu\}) = q(\nu, T_{o\mathbb{D}}\nu) + q(\nu, T_{\mathbb{D}\mathbb{D}}\nu) + q(\nu, T_{\mathbb{D}o}\nu)$$

where the three terms correspond, respectively, to the intensity of births, motions and deaths. The death intensity is always finite, hence the SMQ generator is stable iff

$$\int_{\mathbb{D}} r_{xy}(\nu)\lambda(x, dy) < \infty, \quad x \in \nu \cup \{o\}, \nu \in \mathbb{M} \tag{6.12}$$

what will be assumed from now on.

Uniqueness The generator q defines uniquely the evolution of a jump process until the first (random) time of accumulation of jumps, say τ_∞ (called *explosion* time). Let $N_{\cdot}^{\infty} = \{N_t^{\infty}(\omega) : t \in [0, \tau_\infty(\omega))\}$ be the process describing this evolution, and call it the *minimal q-process*.

Recall that a stable and conservative generator q is called *regular* if there exists a unique Markov jump process defined for all $t \in \mathbb{R}_+$ associated to q, which is equivalent to $\tau_\infty = \infty$, $\mathbf{P}$-a.s. or equivalently {the number of jumps is finite within any finite time interval $(0, t]$} $\mathbf{P}$-a.s.

Sufficient conditions for regularity

We aim to establish sufficient conditions for q to be regular. We begin by a particular case which will be useful later.

Lemma 4 *Let q be a generator given by (6.8). If*

$$\sup_{\nu \in \mathbb{M}} q(\nu, T_{o\mathbb{D}}\nu) + q(\nu, T_{\mathbb{D}\mathbb{D}}\nu) < \infty \tag{6.13}$$

then q is regular.

Proof. In order to prove that q is regular, we will construct the minimal q-process in a manner assuring that $\tau_\infty = \infty$, $\mathbf{P}$-a.s. or equivalently {the number of jumps of $N_{\cdot}^{\infty}$ is finite within any finite time interval $(0, t]$} $\mathbf{P}$-a.s.

First, we construct a sequence of independent and identically random variable $Y_1, Y_2, \ldots$ which are exponential with parameter $\sup_{\nu\in\mathbb{M}} q\left(\nu, T_{o\mathbb{D}}\nu\right)+q\left(\nu, T_{\mathbb{D}\mathbb{D}}\nu\right)$ (we may extend the original probability space if necessary to include these random variables). Let $S_n = \sum_{k=1}^{n} Y_k$ ($\{S_n\}$ is called a renewal process). From (6.13) we deduce that

$$\lim_{n\to\infty} S_n = \infty, \quad \mathbf{P}-a.s. \tag{6.14}$$

We will construct the minimal q-process in a way assuring that the (increasing) sequence of birth and motion instants, denoted $0 \le T_1, T_2, \ldots$, satisfy

$$T_n \ge S_n, \quad \text{almost surely} \tag{6.15}$$

To this end we construct the minimal q-process recursively using competing exponential random variables (or equivalently homogeneous Poisson processes) as in [32]. Let ν be the SMQ state at the time origin $t = 0$. The competing events for the determination of the next jump after $t = 0$ are birth, motion and death. We generate the times to these (eventual) events as exponential random variables, say B, M and D, with parameters $q\left(\nu, T_{o\mathbb{D}}\nu\right)$, $q\left(\nu, T_{\mathbb{D}\mathbb{D}}\nu\right)$ and $q\left(\nu, T_{\mathbb{D}o}\nu\right)$ respectively. We may generate B and M such that $B \ge S_1$ and $M \ge S_1$ (Strassen Theorem). The lowest variable among B, M and D will determine the jump. At the time of this jump, the same procedure recommences with the following precaution. Unless the first birth or motion occurs, B and M are generated such that $B \ge S_1$ and $M \ge S_1$. Once a birth or a motion occurs, we consider the following S_n.

As the state of the SMQ is finite, the number of deaths, before the following birth or motion, is finite. Moreover as the above construction satisfies (6.15), the number of births and motions in each finite time interval is finite due to (6.14). Then the number of jumps of the minimal q-process in each finite time interval is finite. ■

Define

$$b_n = \sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q\left(\nu, T_{o\mathbb{D}}\nu\right), \quad d_n = \inf_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q\left(\nu, T_{\mathbb{D}o}\nu\right) \tag{6.16}$$

From now on we will assume that for each $n \ge 0$

$$b_n < \infty \quad \text{and} \quad \sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q(\nu, T_{\mathbb{D}\mathbb{D}}\nu) < \infty \tag{6.17}$$

The following result gives sufficient conditions for the q to be regular.

Proposition 21 *Let q be a generator given by (6.8). If the discrete birth-death generator q' on $\mathbb{N}$ with birth rates b_n and death rates d_n defined by (6.16) is regular then so is the generator q.*

We begin by giving simple proofs in some particular cases, then we give the proof in the general case.

Proof. Proposition 21 in the case of finite networks.

(Cf. [8, p.309] and the references [101, 68] therein.) ■

Proof. Proposition 21 in the case of birth-death networks with general state space.

Our proof consists of verifying the conditions of [97, Proposition 5.1]. The generator is given by

$$q(\nu,\Gamma)=\int_{\mathbb{D}} r_{oy}(\nu)\,1_{\Gamma}(T_{oy}\nu)\,\lambda(o,dy)+\int_{\mathbb{D}} r_{xo}(\nu)\,1_{\Gamma}(T_{xo}\nu)\,\lambda(x,o)\,\nu(dx)\quad \Gamma\in\mathcal{M},\nu\in\mathbb{M}\backslash\Gamma$$

Using the same notations as [97], denote

$$B(\nu,\Gamma)=\int_{\Gamma} r_{oy}(\nu)\,\lambda(o,dy)\,,\quad D(\nu,x)=r_{xo}(T_{ox}\nu)\,\lambda(x,o)$$

$$\alpha(\nu)=q(\nu)$$

$$K(\nu,\Gamma)=q(\nu,\Gamma)/q(\nu)\,,\quad \text{if } q(\nu)>0$$

(if $q(\nu)=0$ then it doesn't matter how to define $K(\nu,\cdot)$) then

$$q(\nu,\Gamma)=B(\nu,T_{oL}\nu)+\sum_{x\in\nu} D(T_{xo}\nu,x)\,1_{\Gamma}(T_{xo}\nu)$$

which is equivalent to [97, Formula (4.4a)]. Note that assumption (6.9) is implicit in [97]. Moreover the stability condition (6.12) reduces to $q(\nu,T_{o\mathbb{D}}\nu)<\infty,\forall\nu\in\mathbb{M}$ which is equivalent to Preston's condition that, for each $\nu\in\mathbb{M}$, $B(\nu,\cdot)$ is a finite measure on $(\mathbb{D},\mathcal{D})$. Finally, Condition (6.17) reduces to $\sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q(\nu,T_{o\mathbb{D}}\nu)<\infty,\forall n\in\mathbb{N}$ which is equivalent to Preston's condition $b_n<\infty,\forall n\in\mathbb{N}$. Hence our proposition for birth-death networks is just a reformulation of [97, Proposition 5.1]. ■

In order to prove Proposition 21 in the general case we will construct a coupling between the SMQ generator q given by (6.8) and the birth-death generator q' analogous to that given by [97]. Consider an infinitesimal generator $\tilde{q}$ on $\tilde{\mathbb{M}}=\mathbb{M}\times\mathbb{N}$ defined as follows.

- If $\nu(\mathbb{D})\neq n$, then

$$\begin{aligned}\tilde{q}((\nu,n);\Gamma\times\{n\})&=q(\nu,\Gamma)\\ \tilde{q}((\nu,n);\{\nu\}\times\{n+1\})&=b_n\\ \tilde{q}((\nu,n);\{\nu\}\times\{n-1\})&=d_n\\ \tilde{q}((\nu,n))&=q(\nu)+b_n+d_n\end{aligned}$$

- If $\nu(\mathbb{D}) = n$, then

$$\tilde{q}((\nu, n); \Gamma \times \{n+1\}) = q(\nu, \Gamma), \quad \Gamma \subset T_{o\mathbb{D}}\nu$$
$$\tilde{q}((\nu, n); \{\nu\} \times \{n+1\}) = b_n - q(\nu, T_{o\mathbb{D}}\nu) \tag{6.18}$$
$$\tilde{q}((\nu, n); \Gamma \times \{n-1\}) = q(\nu, \Gamma)\frac{d_n}{q(\nu, T_{\mathbb{D}o}\nu)}, \quad \Gamma \subset T_{\mathbb{D}o}\nu \tag{6.19}$$
$$\tilde{q}((\nu, n); \Gamma \times \{n\}) = q(\nu, \Gamma)\left(1 - \frac{d_n}{q(\nu, T_{\mathbb{D}o}\nu)}\right), \quad \Gamma \subset T_{\mathbb{D}o}\nu \tag{6.20}$$
$$\tilde{q}((\nu, n); \Gamma \times \{n\}) = q(\nu, \Gamma), \quad \Gamma \subset T_{\mathbb{D}\mathbb{D}}\nu \tag{6.21}$$
$$\tilde{q}((\nu, n)) = b_n + q(\nu, T_{\mathbb{D}o}\nu) + q(\nu, T_{\mathbb{D}\mathbb{D}}\nu) \tag{6.22}$$

  In order to adapt the coupling defined by [97] for our generator q, we added the term (6.21) to take into account motions and modify the term (6.22) in order to assure conservation of the coupling generator.

Note that, for every $(\nu, n) \in \tilde{\mathbb{M}}$, and $\Gamma \in \mathcal{M}$ such that $\nu \notin \Gamma$,

$$\tilde{q}((\nu, n); \Gamma, \mathbb{N}) = q(\nu, \Gamma), \quad \text{which is independent of } n$$
$$\tilde{q}((\nu, n); \mathbb{M}, n+1) = b_n, \quad \text{which is independent of } \nu$$
$$\tilde{q}((\nu, n); \mathbb{M}, n-1) = d_n, \quad \text{which is independent of } \nu$$

which assures that $\tilde{q}$ is a *coupling* of two generators. (The above conditions which are taken from [33] mean that the transition rate of each component is independent of the initial state of the other component. Note that $\tilde{q}((\nu, n); \{\nu\}, \mathbb{N})$ may be different from $0 = q(\nu, \{\nu\})$.) We see that the marginal generators (cf. [33]) of the first and second components are equal to q and q' respectively (where q' is a birth-death generator on $\mathbb{N}$ with birth rates b_n and death rates d_n) except for the diagonal terms which may be non-null which means that the marginal processes may have pseudo-transitions. Recall that two generators differing only by the diagonal terms lead to Markov jump processes which are equal in distribution.

The coupling construction above has the following interpretation. Consider the minimal $\tilde{q}$- process $\tilde{N}^{\infty}_{\cdot} = (N_{\cdot}, N'_{\cdot})$ (recall that $\tilde{N}^{\infty}_t(\omega)$ is defined until the first explosion time, say $\tilde{\tau}_{\infty}(\omega)$). Denote its components on $\mathbb{M}$ and $\mathbb{N}$ by $N_{\cdot}$ and $N'_{\cdot}$ respectively. Assume that at the time origin $N_0(\mathbb{D}) \leq N'_0$ and consider times $t \in [0, \tilde{\tau}_{\infty})$. As long as $\nu(\mathbb{D}) < n$ (i.e. $N_t(\mathbb{D}) < N'_t$) the components $N_{\cdot}$ and $N'_{\cdot}$ evolve independently of each other. If $\nu(\mathbb{D}) = n$ (i.e. $N_t(\mathbb{D}) = N'_t$), then we consider first $N'_{\cdot}$-births at rate b_n and $N_{\cdot}$-deaths at rate $q(\nu, \Gamma)$. Each $N'_{\cdot}$-birth is either accompanied by an $N_{\cdot}$-birth at rate $q(\nu, T_{o\mathbb{D}}\nu)$ or not at rate $b_n - q(\nu, T_{o\mathbb{D}}\nu)$. We can generate a Bernoulli random variable with parameter $q(\nu, T_{o\mathbb{D}}\nu)/b_n$ saying whether an $N'_{\cdot}$-birth is accompanied by an $N_{\cdot}$-birth or not. Analogously, each $N_{\cdot}$-death is either accompanied by an $N'_{\cdot}$-death at rate (6.19) or not at rate (6.20). The rates here are a little bit more complicated to write as they depend of the set Γ where the $N_{\cdot}$-death occurs. We can generate a Bernoulli random variable with parameter $d_n/q(\nu, T_{\mathbb{D}o}\nu)$ saying whether an $N_{\cdot}$-death is accompanied by an $N'_{\cdot}$-death or not. Finally when there is an $N_{\cdot}$-motion, $N'_{\cdot}$ remains unchanged. This coupling assures that, if $N_0(\mathbb{D}) \leq N'_0$ then

$$N_t(\mathbb{D}) \leq N'_t, \quad 0 \leq t < \tilde{\tau}_\infty \tag{6.23}$$

Note that the first explosion time of the q-process τ_∞ and of the q'-process τ'_∞ are both larger than $\tilde{\tau}_\infty$, i.e.

$$\tau_\infty \geq \tilde{\tau}_\infty, \quad \tau'_\infty \geq \tilde{\tau}_\infty \tag{6.24}$$

Then the paths $N_t(\omega)$ and $N'_t(\omega)$ where $t \in [0, \tilde{\tau}_\infty(\omega))$ are portions of the paths of the minimal q-process $N^\infty_{\cdot}$ and the minimal q'-process $N'^\infty_{\cdot}$ respectively.

We will need the following lemma to prove Proposition 21. Denote $\mathbb{M}_m = \{\nu \in \mathbb{M} : \nu(\mathbb{D}) \leq m\}$, $\mathbb{M}'_m = \{1, \ldots, m\}$ and $\tilde{\mathbb{M}}_{m,k} = \mathbb{M}_m \times \mathbb{M}'_k$.

Lemma 5 *Consider the minimal $\tilde{q}$-process $\left\{\tilde{N}^\infty_t\right\} = \{(N_t, N'_t), t \in [0, \tilde{\tau}_\infty)\}$. Then for all $m, n \in \mathbb{N}$*

$$\mathbf{P}\{\tilde{N}^\infty_s \in \tilde{\mathbb{M}}_{m,n} \textit{ for all } s \in (0, \tau_\infty);\ \textit{and } \tilde{\tau}_\infty < \infty\} = 0 \tag{6.25}$$

Proof. It is enough to prove that if all visited states within a finite interval are in $\tilde{\mathbb{M}}_{m,n}$, then the number of jumps is finite $\mathbf{P}$-a.s.

Fix some $m \in \mathbb{N}, t \geq 0$. Let's first prove that, within $(0, t]$, if all visited states are in $\tilde{\mathbb{M}}_{m,n}$ then the number of jumps is finite. Let $A(t)$ denote the number of jumps within $(0, t]$ and $A'(t) = A(t) \times 1\{$all visited states within $(0, t]$ are in $\tilde{\mathbb{M}}_{m,n}\}$. Let T denote the time of the first transition to a state outside $\tilde{\mathbb{M}}_{m,n}$ and $\tau(t) = \min(t, T)$. Note that

$$A'(t) \leq A(\tau(t))$$

Note that $A(\tau(t))$ may be viewed as the number of jumps within $(0, t]$ of a modified Markov generator $\tilde{q}_{m,m}$ where each state outside $\tilde{\mathbb{M}}_{m,n}$ is made absorbing, i.e.

$$\begin{aligned}
\tilde{q}_{m,m}((\nu, n); \Gamma) &= q((\nu, n), \Gamma) 1\left\{(\nu, n) \in \tilde{\mathbb{M}}_{m,m}\right\}, \qquad && \nu \in \tilde{\mathbb{M}}, \Gamma \subset \tilde{\mathbb{M}} \\
\tilde{q}_{m,m}((\nu, n)) &= q((\nu, n)) 1\left\{(\nu, n) \in \tilde{\mathbb{M}}_{m,m}\right\}, && \nu \in \tilde{\mathbb{M}}
\end{aligned}$$

By (6.17), the generator $\tilde{q}_{m,m}$ satisfies (6.13), then it is regular by Lemma 4. Hence $A(\tau(t))$ is finite which implies that $A'(t)$ is finite. This completes the proof. ■

Proof. Proposition 21 in the case of general networks (i.e. including motions in addition to births and deaths).

Let q, q' be as in Proposition 21. Let $\tilde{q}$ be the coupling defined above. Let $Q^m_t(\nu, \Gamma)$ be the probability of a transition of $N^\infty_{\cdot}$ from state ν to Γ; in time t; with a finite number of jumps; and with visited states in $\mathbb{M}_m$. Similarly $Q'^m_t(n, A)$ designates the probability of a transition of $N'^\infty_{\cdot}$ from state n to A;

in time t; with a finite number of jumps; and with visited states in $\mathbb{M}'_m$; and $\tilde{Q}_t^{m,k}(\nu, n; F)$ designates the probability of a transition of $\tilde{N}^{\infty}_{\cdot}$ from state (ν, n) to F; in time t; with a finite number of jumps; and with visited states in $\tilde{\mathbb{M}}_{m,k}$.

From Inequality (6.23) we deduce that for $\nu(\mathbb{D}) \leq n$, we have $\tilde{Q}_t^{m,k}(\nu, n; F) = \tilde{Q}_t^{k,k}(\nu, n; F), \forall m \geq k$. We aim now to prove that

$$\tilde{Q}_t^{m,m}(\nu, n; \mathbb{M} \times A) = Q_t'^{m}(n, A) \tag{6.26}$$

We will prove that each side of (6.26) is not larger than the other one. (i) Consider a transition of $\tilde{N}^{\infty}_{\cdot}$ from state (ν, n) to $\mathbb{M} \times A$; in time t; with a finite number of jumps; and with visited states in $\tilde{\mathbb{M}}_{m,m}$. In such case we have a transition of $N'_{\cdot}$ from state n to A; in time t; with a finite number of jumps; and with visited states in $\mathbb{M}'_m$. By (6.24), such a transition may be viewed as a transition of $N'^{\infty}_{\cdot}$, hence $\tilde{Q}_t^{m,m}(\nu, n; \mathbb{M} \times A) \leq Q_t'^{m}(n, A)$. (ii) Consider now a transition of $N'^{\infty}_{\cdot}$ from state n to A; in time t; with a finite number of jumps; and with visited states in $\mathbb{M}_m$. If $t < \tilde{\tau}_{\infty}$, then we have a transition of $\tilde{N}^{\infty}_{\cdot}$ from state (ν, n) (for some $\nu \in \mathbb{M}$) to $\mathbb{M} \times A$; in time t; and with visited states in $\tilde{\mathbb{M}}_{m,m}$. It remains to show that $\mathbf{P}\{N'^{\infty}_s \in \mathbb{M}_m$ for all $s \in (0, t]$; and $t \geq \tilde{\tau}_{\infty}\} = 0$; to obtain $Q_t'^{m}(n, A) \leq \tilde{Q}_t^{m,m}(\nu, n; \mathbb{M} \times A)$. Note that

$$\begin{aligned}
&\{N'^{\infty}_s \in \mathbb{M}_m \text{ for all } s \in (0, t];\text{ and } t \geq \tilde{\tau}_{\infty}\} \\
&\quad \subset \{\tilde{N}^{\infty}_s \in \tilde{\mathbb{M}}_{m,m} \text{ for all } s \in (0, \tilde{\tau}_{\infty});\text{ and } t \geq \tilde{\tau}_{\infty}\} \\
&\quad \subset \{\tilde{N}^{\infty}_s \in \tilde{\mathbb{M}}_{m,m} \text{ for all } s \in (0, \tilde{\tau}_{\infty});\text{ and } \tilde{\tau}_{\infty} < \infty\}
\end{aligned}$$

The event in the right-hand side has probability 0 by Lemma 5. This finishes the proof of (6.26).

Now by the continuity property of the probability

$$\lim_{m\to\infty} \tilde{Q}_t^{m,m}(\nu, n; \mathbb{M} \times A) = \tilde{P}_t^{(\infty)}(\nu, n; \mathbb{M}\times A), \quad \lim_{m\to\infty} Q_t'^{(m)}(n, A) = P_t'^{(\infty)}(n, A)$$

where $\tilde{P}_t^{(\infty)}$, $P_t'^{(\infty)}$ designate the minimal solutions of the backward Kolmogorov equations with the generators, respectively, $\tilde{q}$ and q'. Then, by (6.26)

$$\tilde{P}_t^{(\infty)}(\nu, n; \mathbb{M} \times A) = P_t'^{(\infty)}(n, A)$$

and in particular $\tilde{P}_t^{(\infty)}(\nu, n; \mathbb{M} \times \mathbb{N}) = P_t'^{(\infty)}(n, \mathbb{N}) = 1$.

Observe that $\tilde{P}_t^{(\infty)}(\nu, n; \mathbb{M} \times \mathbb{N}) \leq P_t^{(\infty)}(\nu, \mathbb{M})$ (since if $\tilde{N}^{\infty}_{\cdot}$ makes a finite number of jumps, then $N_{\cdot}$ makes a finite number of jumps, hence $N^{\infty}_{\cdot}$ makes a finite number of jumps as it is equal in distribution to $N_{\cdot}$). We deduce that $P_t^{(\infty)}(\nu, \mathbb{M}) = 1$ which finishes the proof. ■

Discrete case

Remark 11 *In the discrete case, one may apply [34, Theorem 3.19] to obtain sufficient conditions for regularity of the SMQ generator. But in order to apply* *this theorem we should have* $\sup_{\nu\in\mathbb{M}_m} q(\nu) < \infty, \forall m \in \mathbb{N}$*, and in particular the death rates are uniformly bounded over* $\mathbb{M}_m$*, which is more restrictive than Condition (6.17).*

Conditions for a birth-death generator q' on $\mathbb{N}$ with birth rates $q'_{n,n+1} = b_n$ and death rates $q'_{n,n-1} = d_n$ (where b_n, d_n are given non-negative constants) to be regular are given in [102]. The following are sufficient conditions

$$b_n = 0, \quad \forall n \geq n_0 \tag{6.27}$$

or

$$b_n > 0, \quad \forall n \geq n_0 \text{ and } \sum_{n=n_0}^{\infty} w_n = \infty \tag{6.28}$$

where

$$w_n = \frac{1}{b_n} + \frac{d_n}{b_n b_{n-1}} + \ldots + \frac{d_n \ldots d_{n_0+1}}{b_n \ldots b_{n_0}} + \frac{d_n \ldots d_{n_0}}{b_n \ldots b_{n_0}}$$

Note that (6.28) is satisfied if either

$$b_n > 0, \quad \forall n \geq n_0 \text{ and } \sum_{n=n_0}^{\infty} \frac{1}{b_n} = \infty$$

or

$$b_n > 0, \quad \forall n \geq n_0 \text{ and } \sum_{n=n_0}^{\infty} \frac{d_n \ldots d_{n_0}}{b_n \ldots b_{n_0}} = \infty$$

Corollary 4 *Consider a birth-death generator on* $\mathbb{N}$ *with birth rates* b_n *and death rates* d_n*. If* $\sup_{n\in\mathbb{N}} b_n < \infty$ *then the generator is regular.*

Proof. Let $b = \sup_{n\in\mathbb{N}} b_n$. The birth-death generator on $\mathbb{N}$ with birth rates b and death rates d_n is regular by (6.27) if $b = 0$ and by (6.28) otherwise. Then by Proposition 21 the birth-death generator on $\mathbb{N}$ with birth rates b_n and death rates d_n is regular. ■

Corollary 5 *Consider the context of Proposition 21, if* $\sup_{\nu\in\mathbb{M}} q(\nu, T_{o\mathbb{D}}\nu) < \infty$ *then* q *is regular.*

Proof. We use the same notation as in Proposition 21. As $\sup_{\nu\in\mathbb{M}} q(\nu, T_{o\mathbb{D}}\nu) = \sup_{n\in\mathbb{N}} b_n < \infty$, the birth-death generator on $\mathbb{N}$ with birth rates b_n and death rates d_n is regular by Corollary 4. Then q is regular by Proposition 21. ■

6.2.3 Interpretation of Whittle SMQ

In this section we consider a Whittle SMQ with a regular generator q which defines a (unique in distribution) Markov process, say $\{N_t\}_{t\geq 0}$. We aim to describe the dynamics of such a process. Assume that

$$\psi_o(\nu) = 1 \tag{6.29}$$

We have

$$q\left(\nu, T_{x\bar{\mathbb{D}}}\nu\right) = \lambda\left(x, \bar{\mathbb{D}}\right)\nu_x\psi_x\left(\nu\right)$$

In particular, the exogenous arrival rate equals

$$q\left(\nu, T_{o\bar{\mathbb{D}}}\nu\right) = \lambda\left(o, \bar{\mathbb{D}}\right)\psi_o\left(\nu\right) = \lambda\left(o, \bar{\mathbb{D}}\right)$$

(Since by assumption (6.29, we have $\psi_o(\nu) = r_{oy}(\nu) = 1$.). On the other hand

$$\frac{q\left(\nu, T_{xA}\nu\right)}{q\left(\nu, T_{x\bar{\mathbb{D}}}\nu\right)} = \frac{\lambda\left(x, A\right)}{\lambda\left(x, \bar{\mathbb{D}}\right)}, \quad A \in \bar{\mathcal{D}}$$

Hence, we have the following description of the Whittle network dynamics [106, 67]. Given the routing rates $\{\lambda(x, A)\}_{x\in\bar{\mathbb{D}}, A\in\bar{\mathcal{D}}}$ and the service rates $\{\psi_x(\nu)\}_{x\in\bar{\mathbb{D}}, \nu\in\mathbb{M}}$, the system dynamics are the following:

(a1) Exogenous arrivals come to dx as a Poisson process with intensity $\lambda(o, dx)$.

(b1) Each departure from $x \in \mathbb{D}$ is routed to dy according to the probability Kernel $\lambda(x, dy)/\lambda\left(x, \bar{\mathbb{D}}\right)$, independently of everything else.

(c1) Whenever the system is in state ν, the time to the next departure from location x is exponentially distributed with rate $\nu_x\psi_x(\nu)\lambda\left(x, \bar{\mathbb{D}}\right)$.

6.3 Limiting behavior

In this section we consider q a generator given by (6.8) and we suppose that the hypotheses of Proposition 21 are satisfied. We will give sufficient conditions for q to be ergodic. We will use again the coupling generator $\tilde{q}$ introduced in the previous section and will show that if 0 is the positive recurrent state for q' then the null measure $\emptyset$ ($\emptyset(A) \equiv 0$ for any $A \in \mathcal{D}$) is such a state for q. Then the ergodicity of q will follow from the standard arguments for regenerative processes.

We denote P_t, P'_t and $\tilde{P}_t$ the unique transition functions associated to q, q' and $\tilde{q}$. We denote $N_., N'_.$ and $\tilde{N}_. = (N_., N'_.)$ the respective corresponding Markov processes (which are unique in distribution). We take the initial state such that $N_0(\mathbb{D}) \leq N'_0$ which assures (6.23).

6.3.1 Limiting distribution

We aim first to give a sufficient condition under which, if the process starts from state $\emptyset$, then the limiting distribution of the process exists.

Lemma 6 *Let q' be the discrete birth-death generator with birth rates b_n and death rates d_n defined by (6.16); T'^0 be the return time of the q'-process to state 0; and $T^{\emptyset}$ be the return time of the q-process to state $\emptyset$.*

(i) If the state 0 *is* positive recurrent *for q' then $\emptyset$ is a positive recurrent state for q.*

(ii) *If for some* $n \in \mathbb{N}$, $\mathbf{P}\left\{T'^{0} < \infty | N'_0 = n\right\} = 1$, *then* $\mathbf{P}\left\{T^{\emptyset} < \infty | N_0 = \nu\right\} = 1$ *for all* $\nu \in \mathbb{M}_n$.

Proof.

(i) Let's denote by T'^{0} is the return time of $N'_{\cdot}$ to state 0. Let's denote the times of the first jumps of $N_{\cdot}$ and $N'_{\cdot}$ by τ_1 and τ'_1 respectively. Starting from $N_0 = \emptyset$ and $N'_0 = 0$, the following first event for both processes $N_{\cdot}$ and $N'_{\cdot}$ is necessarily a birth. Denote . By the coupling inequality (6.23) the first birth in $N_{\cdot}$ occurs after that in $N'_{\cdot}$, then $\tau'_1 \leq \tau_1$. We will show that in fact $\tau'_1 = \tau_1$ (conditionally to the fact that $N_0 = \emptyset, N'_0 = 0$). Note that the birth rate at state 0 of $N'_{\cdot}$ is, by definition, given by

$$b_0 = \sup_{\nu \in \mathbb{M}: \nu(\mathbb{D}) = 0} q\left(\nu, T_{o\mathbb{D}}\nu\right) = q\left(\emptyset, T_{o\mathbb{D}}\emptyset\right)$$

which equals the birth rate at state $\emptyset$ of $N_{\cdot}$. From (6.18) we deduce that $\tilde{q}\left(\emptyset, 0; \{\emptyset\} \times \{1\}\right) = 0$. Hence the first $N'_{\cdot}$-birth is accompanied by an $N_{\cdot}$-birth. Hence $\tau'_1 = \tau_1$. Observe now that

$$\begin{aligned}
\mathbf{E}\left[T'^{0} | N'_0 = 0\right] &= \mathbf{E}\left[\inf\left\{t : t \geq \tau'_1, N'_t = 0\right\} | N'_0 = 0\right] \\
&= \mathbf{E}\left[\inf\left\{t : t \geq \tau'_1, N_t \in \mathbb{M}, N'_t = 0\right\} | N_0 = \emptyset, N'_0 = 0\right] \\
&= \mathbf{E}\left[\inf\left\{t : t \geq \tau_1, N_t = \emptyset, N'_t = 0\right\} | N_0 = \emptyset, N'_0 = 0\right] \\
&\geq \mathbf{E}\left[\inf\left\{t : t \geq \tau_1, N_t = \emptyset, N'_t \in \mathbb{N}\right\} | N_0 = \emptyset, N'_0 = 0\right] \\
&= \mathbf{E}\left[\inf\left\{t : t \geq \tau_1, N_t = \emptyset\right\} | N_0 = \emptyset\right] \\
&= \mathbf{E}\left[T^{\emptyset} | N_0 = \emptyset\right]
\end{aligned}$$

Then $E\left[T^{\emptyset} | N_0 = \emptyset\right] < \infty$, which means that $\emptyset$ is positive recurrent for the SMQ generator q.

(ii) Fix some $\nu \in \mathbb{M}_n$. Let's denote the times of the first jumps of $N_{\cdot}$ and $N'_{\cdot}$ by τ_1 and τ'_1 respectively. We have

$$\begin{aligned}
\mathbf{P}\left\{T^{\emptyset} = \infty | N_0 = \nu\right\} &= \mathbf{P}\left\{N_t \neq \emptyset, \forall t \geq \tau_1 | N_0 = \nu\right\} \\
&= \mathbf{P}\left\{N_t \neq \emptyset, N'_t \in \mathbb{N}, \forall t \geq \tau_1 | N_0 = \nu, N'_0 = n\right\} \\
&= \mathbf{P}\left\{N_t \neq \emptyset, N'_t \neq 0, \forall t \geq \tau_1 | N_0 = \nu, N'_0 = n\right\} \\
&\leq \mathbf{P}\left\{N_t \in \mathbb{M}, N'_t \neq 0, \forall t \geq \tau_1 | N_0 = \nu, N'_0 = n\right\} \\
&= \mathbf{P}\left\{N'_t \neq 0, \forall t \geq \tau_1 | N'_0 = n\right\} \\
&= \mathbf{P}\left\{N'_t \neq 0, \forall t \geq \tau'_1 | N'_0 = n\right\} \\
&= \mathbf{P}\left\{T'^{0} = \infty | N'_0 = n\right\} = 0
\end{aligned}$$

where for the third equality we use the coupling inequality (6.23), and for the sixth line we use the fact that $\tau'_1 = \tau_1$ proved in (i).

∎

Proposition 22 *If $\emptyset$ is a positive recurrent state for the generator q then*

$$\Pi(\Gamma) = \lim_{t\to\infty} P_t(\emptyset, \Gamma)$$

exists for all $\Gamma \in \mathcal{M}$ and it is given by

$$\Pi(\Gamma) = \frac{1}{\mathbf{E}_\emptyset\left[T^\emptyset\right]} \mathbf{E}_\emptyset\left[\int_0^{T^\emptyset} 1\{N_t \in \Gamma\}\, dt\right] \tag{6.30}$$

where $T^\emptyset$ is the return time of $N.$ to state $\emptyset$ and $E_\emptyset[\cdot]$ is the conditional expectation $E_\emptyset[\cdot] = E[\cdot|N_0 = \emptyset]$.

Proof. Clearly, conditionally to $N_0 = \emptyset$, the process $\{N_t\}_{t\geq 0}$ is a zero-delayed regenerative process with generation points $\{t \geq 0 : N_t = \emptyset\}$. We aim to apply the limit theorem [10, Theorem 1.2 p.170] for such processes. The conditions of this theorem, besides the fact that $\{N_t\}_{t\geq 0}$ is regenerative, are: (i) the state space $\mathbb{M}$ is a metric space; (ii) $\{N_t\}_{t\geq 0}$ has right-continuous paths; (iii) the cycle length distribution, say F, defined by $F(A) = \mathbf{P}\left\{T^\emptyset \in A | N_0 = \emptyset\right\}, A \in \mathcal{B}(\mathbb{R})$ has a finite mean; and (iv) F is non-lattice (called also non-arithmetic). Let's prove that these conditions hold.

(i) The state space $\mathbb{M}$ may be metrizable (cf. [40, p.629]).

(ii) We assume that, by construction, the Markov jump process $\{N_t\}_{t\geq 0}$ has right-continuous paths.

(iii) $\emptyset$ is a positive recurrent state for the generator q means that the cycle length distribution F has finite mean.

(iv) It is sufficient to show that F is absolutely continuous with respect to the Lebesgue measure. We denote by $N_1, N_2, \ldots, N_K$ the sequence of visited states (other than $\emptyset$) in the first cycle (the number K of these states is also random). For $A \in B(\mathbb{R})$,

$$F(A) = \mathbf{E}\left[\mathbf{P}\left\{T^\emptyset \in A | N_0 = \emptyset, N_1, \ldots, N_K\right\} 1\{N_1 \neq \emptyset, \ldots, N_K \neq \emptyset\}\right]$$

Note that, conditionally to $N_0 = \emptyset, N_1, \ldots, N_K$, the random variable $T^\emptyset$ is a sum of exponential random variables. If A has a zero Lebesgue measure, then $\mathbf{P}\left\{T^\emptyset \in A | N_0 = \emptyset, N_1, \ldots, N_K\right\} = 0$, hence $F(A) = 0$.

Conditions of [10, Theorem 1.2 p.170] are all satisfied, then $\mathbf{P}_\emptyset\{N_t \in \cdot\}$ converges weakly of to $\Pi(\cdot)$. This means that for all Π-continuity set $\Gamma \in \mathcal{M}$ (i.e. $\Pi(\partial\Gamma) = 0$), $\mathbf{P}_0\{N_t \in \Gamma\} \to \Pi(\Gamma)$.

(v) In fact we may show that the previous convergences holds true for all $\Gamma \in \mathcal{M}$. To this end, we proceed exactly as in the proof of [10, Theorem 1.2 p.170], but here we consider a measurable function $f : \mathbb{M} \to \mathbb{R}$

which is bounded but not necessarily continuous (for example $f(x) = 1\{x \in \Gamma\}$). It is then not immediate that the function z defined by $z(t) = E_{\emptyset}\left[f(N_t); t < T^{\emptyset}\right] = \mathbf{P}_{\emptyset}\left\{N_t \in \Gamma; t < T^{\emptyset}\right\}$ is right-continuous. In fact, this is true because our process $\{N_t\}$ is a regular jump Markov process. (Fix some $s \geq 0$. Note that for each $\omega \in \Omega$, there exists some $\epsilon > 0$ such that $N_t(\omega) = N_s(\omega)$ for each $t \in [s, s+\epsilon)$, hence $\lim_{t \to s+} f(N_t) = f(N_s)$. The dominated convergence theorem gives

$$\lim_{t \to s+} \mathbf{E}_{\emptyset}\left[f(N_t); t < T^{\emptyset}\right] = \mathbf{E}_{\emptyset}\left[f(N_s); s < T^{\emptyset}\right]$$

Hence $z(t) = E_{\emptyset}\left[f(N_t); t < T^{\emptyset}\right]$ is right-continuous.)

■

The following corollary gives a sufficient condition under which, if the process starts from some state ν, then the limiting distribution of the process exists.

Corollary 6 *If $\emptyset$ is a positive recurrent state for the generator q and if for some $\nu \in \mathbb{M}$, $\mathbf{P}\left\{T^{\emptyset} < \infty | N_0 = \nu\right\} = 1$, then, for all $\Gamma \in \mathcal{M}$,*

$$\lim_{t \to \infty} P_t(\nu, \Gamma) = \Pi(\Gamma) \tag{6.31}$$

where Π is given by (6.30).

Proof. Conditionally to $N_0 = \nu$, $\{N_t\}$ is a delayed regenerative process with a delay equal to the first time the process attains the regeneration point $\emptyset$, which is precisely $T^{\emptyset}$ the return time to $\emptyset$. [10, Theorem 1.2 p.170] assures the weak convergence which may be strengthened in the same manner as (v) of the proof of Proposition 22. ■

The following result strengthens the convergence assured in Proposition 22 and Corollary 6 by assuring that it holds in the sense of total variation.

Proposition 23 *Under the conditions of Corollary 6, the convergence*

$$\lim_{t \to \infty} P_t(\nu, \cdot) = \Pi(\cdot)$$

holds in the sense of total variation. (Recall that if this is true for all $\nu \in \mathbb{M}$, then the process is said to be ergodic*.)*

Proof. We apply [10, Corollary 1.4 p.188] to prove convergence of $\mathbf{P}_{\nu}\{N_t \in \cdot\}$ in total variation to Π. The conditions of this corollary, besides the fact that $\{N_t\}_{t \geq 0}$ is regenerative, are: (i) the cycle length distribution, say F, is spread out; (ii) F has a finite mean; (iii) $N_t(\omega)$ is measurable jointly in (t, ω). Let's prove that these conditions hold.

(i) In (iv) of the proof of Proposition 22 we showed that F is absolutely continuous with respect to the Lebesgue measure which implies that F is spread out.

(ii) Proved in (iii) of the proof of Proposition 22.

(iii) We have to show that for each $A \in \mathcal{M}$,

$$B = \{(\omega, t) \in \Omega \times \mathbb{R}_+ : N_t(\omega) \in A\} \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$$

To this end, let $\{t_n\}_{n\in\mathbb{N}}$ be some countable dense set in $\mathbb{R}_+$; let $\{\epsilon_k\}_{k\in\mathbb{N}}$ be a sequence of positive reals such that $\lim_{k\to\infty} \epsilon_k = 0$; let $B_{nk} = N_{t_n}^{-1}(A) \times [t_n - \epsilon_k, t_n]$; and let $B' = \bigcap_k \bigcup_n B_{nk}$. We aim to show that $B' = B$ which will prove the measurability of B.

Consider first some $(\omega, t) \in B'$, for each ϵ_k there exists some t_n such that $N_{t_n}(\omega) \in A$ and $t \in [t_n - \epsilon_k, t_n]$. Due to the fact that our process is a regular jump Markov process, there exists some $\epsilon > 0$ such that $N_s(\omega) = N_t(\omega)$ for each $s \in [t, t + \epsilon]$. Then for ϵ_k sufficiently small $N_{t_n}(\omega) = N_t(\omega)$, hence $N_t(\omega) \in A$. We get that $(\omega, t) \in B$. We deduce that $B' \subset B$.

Consider now some $(\omega, t) \in B$, i.e. $N_t(\omega) \in A$. There exists some $\epsilon > 0$ such that $N_s(\omega) = N_t(\omega)$ for each $s \in [t, t + \epsilon]$. For each ϵ_k consider some $t_n \in [t, t + \epsilon_k \wedge \epsilon]$. Clearly $(\omega, t) \in B_{nk}$. We get that $(\omega, t) \in B'$. We deduce that $B \subset B'$.

Hence $B = B'$ which finishes the proof of the measurability of B for each $A \in \mathcal{M}$; hence the measurability of $N_t(\omega)$ jointly in (t, ω).

■

Remark 12 *Recall that the weak convergence assures that* $\lim_{t\to\infty} E_\nu [f(N_t)] = \int_{\mathbb{M}} f d\Pi$ *only for bounded continuous functions. The convergence assured in Proposition 22 and Corollary 6; i.e.* $\lim_{t\to\infty} P_t(\nu, \Gamma) = \Pi(\Gamma), \forall \Gamma \in \mathcal{M}$*; is equivalent to,* $\lim_{t\to\infty} E_\nu [f(N_t)] = \int_{\mathbb{M}} f d\Pi$ *for every bounded measurable function* $f : \mathbb{M} \to \mathbb{R}$*. (One way is trivial and the other follows from dominated convergence Theorem.) This convergence is stronger than weak convergence showed in [10, Theorem 1.2 p.170] which is due the fact that we consider a regular jump process whereas [10, Theorem 1.2 p.170] considers general regenerative processes.*

Recall that the total variation convergence assures that

$$\lim_{t\to\infty} \sup_f \left| \mathbf{E}_\nu [f(N_t)] - \int_{\mathbb{M}} f d\Pi \right| = 0$$

where the $\sup_f$ *is taken over all bounded measurable functions* f*. Hence the convergence assured in 22 and Corollary 6 is intermediate between the weak convergence and the total variation convergence.*

[72] show that the state 0 is positive recurrent for the birth-death process N' on $\mathbb{N}$ iff either $b_0 = 0$ or $b_0 > 0$ and one of the following holds:

$$b_m = 0 \text{ for some } m \geq 1 \text{ and } d_n > 0 \ \forall 0 < n \leq m \tag{6.32}$$

or

$$b_n > 0 \text{ for all } n \geq 0 \text{ and } d_n > 0 \; \forall n \geq 1 \text{ and} \\ \sum_{n=1}^{\infty} \frac{b_0 \dots b_{n-1}}{d_1 \dots d_n} < \infty \tag{6.33}$$

If $b_0 > 0$, then Condition (6.33) implies that the hypotheses of Lemma 6(ii) holds for all $n \in \mathbb{N}$.

6.3.2 Novelty of our ergodicity conditions

Recall that a Markov Poisson location (MPL) process is defined in Example 3 as a SMQ process for which

$$r_{xy}(\nu) \equiv 1$$

Proposition 24 *Let $\{X_t\}$ be a MPL process. The total number of users $\{X_t(\mathbb{D})\}$ may be viewed as a* M/GI/∞ *queue with arrival rate $\lambda = \lambda(o, \mathbb{D})$. If the traffic equations (these are the balance equations for the routing rates which will be written later (6.47)) admit a solution $\rho(\cdot)$ with finite mass, $\rho(\mathbb{D}) < \infty$, then the mean service duration of the equivalent* M/GI/∞ *queue, say μ^{-1}, is given by*

$$\mu^{-1} = \frac{\rho(\mathbb{D})}{\lambda(o, \mathbb{D})} \tag{6.34}$$

The average time between successive visits of $\{X_t\}$ to $\emptyset$ is given by

$$\mathbf{E}_{\emptyset}\left[T^{\emptyset}\right] = \frac{e^{\rho(\mathbb{D})}}{\lambda(o, \mathbb{D})} < \infty \tag{6.35}$$

and the return time of $\{X_t\}$ to $\emptyset$ from any initial state is almost surely finite.

Proof. In the case of a MPL process, users arrive, move and depart from $\mathbb{D}$ completely independently of each other. The motion dynamics of each user may be described by a Markov process (called MOTION PROCESS) on the state space $\bar{\mathbb{D}}$ with generator $\lambda(\cdot, \cdot)$. Let τ^o be the return time of the motion process to o.

Consider the whole system as a single queue. Let $T_1, T_2, \dots$ be the sequence of exogenous arrival times. The arrival point process $\Phi = \sum_{n \in \mathbb{Z}} \delta_{T_n}$ is a Poisson process with intensity $\lambda(o, \mathbb{D})$. The user arriving at time T_n will be called user n for simplicity. The service duration of a user n equals the return time of the motion process of user n to o, say τ_n^o. The random variables $\{\tau_n^o\}$ are i.i.d. (independent and identically distributed) and independent from the arrival process. Hence we get a M/GI/∞ queue with arrival rate $\lambda = \lambda(o, \mathbb{D})$ and mean service duration, say $\mu^{-1} = \mathbf{E}[\tau_n^o]$, which we will determine later. This finishes the proof of the first part of the proposition.

Assume now that the traffic equations admit a solution $\rho(\cdot)$ with finite mass, $\rho(\mathbb{D}) < \infty$. Then the motion process of each user is positive recurrent and its invariant distribution $\pi(\cdot)$ is obtained by normalizing $\rho(\cdot)$,

$$\pi(A) = \frac{\rho(A)}{\rho(\bar{\mathbb{D}})} = \frac{\rho(A)}{1+\rho(\mathbb{D})}, \quad A \in \bar{\mathcal{D}}$$

in particular

$$\pi(\{o\}) = \frac{1}{1+\rho(\mathbb{D})}$$

By properties of Markov processes, we have

$$\pi(\{o\}) = \frac{1}{\lambda(o,\mathbb{D})\,\mathbf{E}_o[\tau^o]}$$

where $\mathbf{E}_o[\cdot]$ designates the expectation conditionally to that the motion process starts at state o. Hence

$$\mathbf{E}_o[\tau^o] = \frac{1+\rho(\mathbb{D})}{\lambda(o,\mathbb{D})}$$

In order to get the mean service duration of the equivalent M/GI/∞ queue, we have to remove from $\mathbf{E}_o[\tau^o]$ the mean sojourn duration of the motion process at state o which equals $1/\lambda(o,\mathbb{D})$, hence we get

$$\mu^{-1} = \frac{\rho(\mathbb{D})}{\lambda(o,\mathbb{D})}$$

which proves (6.34).

The average time between successive visits of a MPL process $\{X_t\}$ to $\emptyset$ equals the expectation of the busy period in a M/GI/∞ queue which equals [37, p.37]

$$\frac{e^{\lambda/\mu}}{\lambda} = \frac{e^{\rho(\mathbb{D})}}{\lambda(o,\mathbb{D})}$$

hence we get (6.35).

The last part of the proposition is due to the ergodicity of the M/GI/∞ queue when $\lambda/\mu = \rho(\mathbb{D}) < \infty$. ■

Remark 13 *The problem adressed in Proposition 24 was already adressed in [109, Theorem 4] where the ergodicity condition is $\rho(\mathbb{D}) < 1/2$ which is much more constraining than the condition $\rho(\mathbb{D}) < \infty$ in Proposition 24. (Indeed, with the convention $\rho(\{0\}) = 1-\rho(\mathbb{D})$ made in [109, Theorem 4], $\rho(\cdot)$ isn't reversible with respect to the routing rates $\lambda(\cdot,\cdot)$. Such reversibility holds with the convention $\rho(\{0\}) = 1$.)*

Remark 14 *Our ergodicity conditions presented in Section 6.3.1 can be seen as en extension of the results on ergodicity of spatial birth death processes in [97] and we will now comment on the ergodicity result for spatial queueing systems given in [106, Chapter 10]. The open spatial queueing system considered there is* *a special case of our SMQ process, where* $r_{xy}(\nu) = \psi_x(\nu)$ *for some function* $\psi.(\cdot)$. *In this case in [106, Theorem 10.5] one considers, instead of our dominating discrete birth-and-death process, a MPL process with the routing kernel*

$$\hat{\lambda}(x, B) = \bar{b}_x \lambda(x, B), \quad x \in \bar{\mathbb{D}}, B \in \bar{\mathcal{D}}$$

where

$$\bar{b}_x = \begin{cases} \inf_{\nu \neq \emptyset} \psi_x(\nu) & \text{for } x \in \mathbb{D} \\ \sup_\nu \psi_o(\nu) & \text{for } x = o \end{cases}$$

provided $\bar{b}_0 < \infty$ *and* $\bar{b}_x > 0$ *for* $x \in \mathbb{D}$. *As observed in the prove of the Theorem 10.5 there, it is possible to couple the original Whittle process* $\{N_t\}$ *with this MPL process, call it* $\{\hat{X}_t\}$, *in such a way that* $N_t(\mathbb{D}) \leq \hat{X}_t(\mathbb{D})$ *for all* $t \geq 0$. *Then, as observed in this proof too, for* $\emptyset$ *to be the positive recurrent state of* $\{N_t\}$ *it suffices to assume that the same holds true for* $\{\hat{X}_t\}$. *By Proposition 24, the necessary and sufficient condition for this latter is*

$$\int_{\mathbb{D}} 1/\bar{b}_x \, \rho(dx) < \infty \tag{6.36}$$

where $\rho(\cdot)$ *is the solution of the traffic equations (6.47), which is much less constraining than* $\bar{b}_o \int_{\mathbb{D}} 1/\bar{b}_x \, \rho(dx) < 1$ *used in [106] (c.f. condition (10.11) there). Note also that our dominating birth-and-death process in the case of a Whittle network has the following rates*

$$b_n = \sup_{\nu(\mathbb{D})=n} \lambda(o, \mathbb{D})\psi_o(\nu) \leq \bar{b}_o \lambda(o, \mathbb{D}) \tag{6.37}$$

$$d_n = \inf_{\nu(\mathbb{D})=n} \int_{\mathbb{D}} \psi_x(\nu)\lambda(x, \{o\}) \, \nu(dx) \geq \inf_{\nu(\mathbb{D})=n} \int_{\mathbb{D}} \bar{b}_x \lambda(x, \{o\}) \, \nu(dx) \tag{6.38}$$

In contrast to [106, Theorem 10.5] we do not require $\bar{b}_o < \infty$, *which would imply* $\sup_n b_n < \infty$. *Our Lemma 6 combined with the results on the ergodicity of the discrete birth-and-death process gives the result also when* b_n *are unbounded. Indeed suppose that*

$$\inf_{x \in \mathbb{D}} \bar{b}_x \lambda(x, \{o\}) = \epsilon > 0, \tag{6.39}$$

then we have by (6.38) that $d_n \geq n\epsilon > 0$ *and thus by Lemma 6 and (6.33)* $\emptyset$ *is positive recurrent for* N_t *if*

$$\sum_{n=1}^{\infty} \frac{b_0 \ldots b_{n-1}}{\epsilon^n n!} < \infty \tag{6.40}$$

For this latter condition $\sup_n b_n < \infty$ *is sufficient but not necessary. As a final comment, note also that in our approach to ergodicity, we do not require a'priori explicit form of the invariant measure and even existence of the solution of the traffic equations.*

6.3.3 Invariance of the limiting distribution

We aim now to prove that under suitable sufficient conditions, the probability measure Π given by (6.30) is the unique invariant probability measure. (Recall that a non-null measure is said invariant if it satisfies (6.4).)

Lemma 7 *If $\emptyset$ is a positive recurrent state for the generator q, then*

$$\mathbf{P}\left\{T^{\emptyset}<\infty|N_0=\nu\right\}=1,\quad \text{for } \Pi\text{-almost all } \nu\in\mathbb{M}$$

Proof. Consider some $\Gamma\in\mathcal{M}$ such that $\forall\nu\in\Gamma$,

$$\mathbf{P}\left\{T^{\emptyset}<\infty|N_0=\nu\right\}<1$$

We will show that $\Pi(\Gamma)=0$ which will proves the lemma. Let $\mathcal{F}_t=\sigma\{N_s:s\leq t\}$; T be the return time to Γ which is a stopping time (for a Markov jump process, the return time to any measurable set is a stopping time). Let τ be the return time to $\emptyset$ after T, i.e.

$$\tau=\inf\{t\geq T:N_t=\emptyset\}$$

Observe that $\tau=h(N_{T+t}:t\geq 0)$ where h is defined as follows: for every $x:\mathbb{R}_+\to\mathbb{M}$, $h(x)=\inf\{t\geq 0:x_t=\emptyset\}$. We have a.s. on $\{T<\infty\}$,

$$\begin{aligned}
\mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\tau\right] &= \mathbf{E}_{\emptyset}\left[\mathbf{E}\left[1\left\{T\leq T^{\emptyset}\right\}\tau|\mathcal{F}_T\right]\right]\\
&= \mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\mathbf{E}\left[\tau|\mathcal{F}_T\right]\right]\\
&= \mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\mathbf{E}\left[h\left(N_{T+t}:t\geq 0\right)|\mathcal{F}_T\right]\right]\\
&= \mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\mathbf{E}_{N_T}\left[h\left(N_t:t\geq 0\right)|\mathcal{F}_T\right]\right]\\
&= \mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\mathbf{E}_{N_T}\left[T^{\emptyset}\right]\right]
\end{aligned}$$

where the first and second equalities are due to the properties of conditional expectation and the fourth equality is due to the strong Markov property [10, p.34]. (Our process is a Markov jump process, then it has the strong Markov property by [71, Theorem 12.14 p.237]). On the other hand $1\left\{T\leq T^{\emptyset}\right\}\tau\leq T^{\emptyset}$, then

$$\mathbf{E}_{\emptyset}\left[1\left\{T\leq T^{\emptyset}\right\}\mathbf{E}_{N_T}\left[T^{\emptyset}\right]\right]\leq\mathbf{E}_{\emptyset}\left[T^{\emptyset}\right]<\infty$$

But for each $\nu\in\Gamma,\mathbf{E}_{\nu}\left[T^{\emptyset}\right]=\infty$ (because $\mathbf{P}\left\{T^{\emptyset}=\infty|N_0=\nu\right\}>0$) then for each $\omega\in\Omega,\mathbf{E}_{N_T(\omega)}\left[T^{\emptyset}\right]=\infty$. We deduce that $\mathbf{P}_{\emptyset}\left\{T\leq T^{\emptyset}\right\}=0$. From (6.30) we deduce that

$$\Pi(\Gamma)=\frac{1}{\mathbf{E}_{\emptyset}\left[T^{\emptyset}\right]}\mathbf{E}_{\emptyset}\left[\int_0^{T^{\emptyset}}1\{N_t\in\Gamma\}\,dt\right]=0$$

which finishes the proof. ∎

Proposition 25 *If $\emptyset$ is a positive recurrent state for the generator q, then*

(i) $\lim_{t\to\infty} \mathbf{P}_\Pi \{N_t \in \Gamma\} = \Pi(\Gamma)$ *for all* $\Gamma \in \mathcal{M}$*; and*

$$\lim_{t\to\infty} \mathbf{E}_{\boldsymbol{\Pi}} [f(N_t)] = \int_{\mathbb{M}} f d\Pi$$

for every bounded measurable function $f : \mathbb{M} \to \mathbb{R}$*;*

(ii) Π *is an invariant probability measure;*

(iii) If Q is an invariant probability measure, such that

$$\mathbf{P}\left\{T^{\emptyset} < \infty | N_0 = \nu\right\} = 1, \quad \textit{for } Q\textit{-almost all } \nu \in \mathbb{M}$$

then $Q = P$.

Proof.

(i) For $\Gamma \in \mathcal{M}$,

$$\begin{aligned}
\lim_{t\to\infty} \mathbf{P}_\Pi \{N_t \in \Gamma\} &= \lim_{t\to\infty} \int_{\mathbb{M}} \mathbf{P}\{N_t \in \Gamma | N_0 = \nu\} \Pi(d\nu) \\
&= \lim_{t\to\infty} \int_{\mathbb{M}} \mathbf{P}_\nu \{N_t \in \Gamma\} \Pi(d\nu) \\
&= \int_{\mathbb{M}} \lim_{t\to\infty} \mathbf{P}_\nu \{N_t \in \Gamma\} \Pi(d\nu)
\end{aligned}$$

where the last equality is due to the dominated convergence theorem. Using Lemma 7 and Corollary 6, finishes the proof of $\lim_{t\to\infty} \mathbf{P}_\Pi \{N_t \in \Gamma\} = \Pi(\Gamma)$ for all $\Gamma \in \mathcal{M}$. The second assertion in (i) follows from dominated convergence Theorem.

(ii) Fix some $t \geq 0$. By (i) $\lim_{s\to\infty} \Pi P_{s+t}(\Gamma) = \Pi(\Gamma)$. On the other hand

$$\begin{aligned}
\Pi P_{s+t}(\Gamma) &= (\Pi P_s) P_t(\Gamma) \\
&= \int_{\mathbb{M}} P_t(\nu, \Gamma) (\Pi P_s)(d\nu) \\
&= \mathbf{E}_\Pi [P_t(N_s, \Gamma)]
\end{aligned}$$

which converges when $s \to \infty$ to $\int_{\mathbb{M}} P_t(\nu, \Gamma) \Pi(d\nu) = (\Pi P_t)(\Gamma)$ which follows from (i) applied to the function $f(\nu) = P_t(\nu, \Gamma)$. Then $\Pi(\Gamma) = (\Pi P_t)(\Gamma)$ for all $t \in \mathbb{R}_+, \Gamma \in \mathcal{M}$. (Another proof of the fact that Π is invariant may be obtained by proceeding as in the proof of [10, Theorem 3.2 p.200] using only the weak convergence.)

(iii) Lets take an invariant measure Q; i.e. $(QP_t)(\Gamma) = Q(\Gamma)$. For all $t \geq 0$, the distribution of N_t will be $\mathbf{P}_Q\{N_t \in \Gamma\} = (QP_t)(\Gamma) = Q(\Gamma)$. On the other hand

$$\begin{aligned}\lim_{t\to\infty} \mathbf{P}_Q\{N_t \in \Gamma\} &= \lim_{t\to\infty} QP_t(\Gamma) \\ &= \lim_{t\to\infty} \int_{\mathbb{M}} \mathbf{P}_\nu\{N_t \in \Gamma\} Q(d\nu) \\ &= \int_{\mathbb{M}} \lim_{t\to\infty} \mathbf{P}_\nu\{N_t \in \Gamma\} Q(d\nu) \\ &= \int_{\mathbb{M}} \Pi(\Gamma) Q(d\nu) = \Pi(\Gamma)\end{aligned}$$

where the third equality is due to the dominated convergence theorem; and the fourth equality is due to the fact that $\mathbf{P}\left\{T^{\emptyset} < \infty | N_0 = \nu\right\} = 1$ for Q-almost $\nu \in \mathbb{M}$ and Corollary 6. Hence $Q = P$.

■

6.3.4 Time averages

We will give one result concerning some time average limits (such limits are used as the definition of the blocking probability in wireless networks).

Assume that there exists an invariant probability measure Π. First observe that there exists a stationary version of the q-process $\{N_t\}_{t\in\mathbb{R}}$ with doubly infinite time which may be viewed as a random object taking its values in the set of functions $\mathbb{R} \to \mathbb{D}$ which are right-continuous, piecewise constant and with a finite number of discontinuities in each finite interval. Equivalently $\{N_t\}$ may be viewed a marked point process $\sum_{n\in\mathbb{Z}} \delta_{T_n,N_n}$ where $\{T_n\}_{n\in\mathbb{Z}}$ are the jump times (with the usual convention $T_0 \leq 0 < T_1$) and $N_n = N_{T_n}$. Considering the canonical space as in [17, Example 1.1.6 p.9], we see that there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P}_\Pi)$ and a flow $\{\theta_t\}$ such that $\mathbf{P}_\Pi$ is θ_t-invariant and $\{N_t\}$ is θ_t-compatible. (This probability space is eventually enlarged to include some supplementary randomness.) From now on we consider this (enlarged) probability space and we denote it by $(\Omega, \mathcal{F}, \mathbf{P}_\Pi)$ and we denote the flow by $\{\theta_t\}$.

Let Φ be the point process on $\mathbb{R}$ counting the jump times $\{T_n\}_{n\in\mathbb{Z}}$ and $\Phi_\emptyset$ be the point process on $\mathbb{R}$ counting times $\{T_n^\emptyset\}_{n\in\mathbb{Z}}$ of visits to state $\emptyset$ ($T_0^\emptyset \leq 0 < T_1^\emptyset$). Note that $T_1^\emptyset$ equals the return time of the q-process to state $\emptyset$ which was previously denoted $T^\emptyset$. Note that Φ and $\Phi_\emptyset$ are stationary with respective intensities $\lambda = \mathbf{E}_\Pi[\Phi(0,1]]$ and $\lambda_\emptyset = \mathbf{E}_\Pi[\Phi_\emptyset(0,1]]$.

Lemma 8 *If $\emptyset$ is a positive recurrent state for the generator q, then $0 < \lambda_\emptyset < \infty$. In this case we denote $\mathbf{P}_\emptyset$ the Palm probability associated to $\Phi_\emptyset$.*

Proof. Note that $\lambda_\emptyset = 1/E_\emptyset\left[T^\emptyset\right] > 0$ because $E_\emptyset\left[T^\emptyset\right] < \infty$ by assumption. On the other hand $E_\emptyset\left[T^\emptyset\right] \geq \mathbf{E}_\emptyset[T_1] = 1/q(\emptyset, \mathbb{M}) > 0$, then $\lambda_\emptyset < \infty$.

By the Slivnyak inverse construction [17, Section 1.3.5 p.25], it is easy to prove that $\mathbf{P}_\emptyset$ equals the Palm probability $\mathbf{P}^0_\emptyset$ associated to $\Phi_\emptyset$. This justifies the notation suggested in the Lemma. ■

Let $\{\tau_n\}_{n\geq 1}$ be a renewal process on $(0,+\infty)$ and denote $\tau_0 = 0$. We say that a process $\{X_t\}_{t\geq 0}$ is *cumulative* with respect to $\{\tau_n\}_{n\geq 1}$, if $X_0 = 0$; for any $n \geq 0$, $\{X_{\tau_n+t} - X_{\tau_n}\}_{t\geq 0}$ is independent from $\tau_1, \ldots, \tau_n$ and $\{X_t\}_{t<\tau_n}$; and the distribution of $\{X_{\tau_n+t} - X_{\tau_n}\}_{t\geq 0}$ is independent of n [10, p.178]. We will say that a point process N on $\mathbb{R}$ is cumulative with respect to $\{\tau_n\}_{n\geq 1}$ if the associated process $\{N(0,t]\}_{t\geq 0}$ is so.

Let H be a measurable subset of $\mathbb{D}\times\mathbb{D} - \mathrm{diag}(\mathbb{D})$ and let Φ_H be the point process counting the H-transitions of $\{N_t\}$ defined by

$$\Phi_H(B) = \sum_{n\in\mathbb{Z}} 1\{(N_{n-1}, N_n) \in H\}\, 1_B(T_n), \quad B \in \mathcal{B}(\mathbb{R}) \tag{6.41}$$

Note that we may write $\Phi_H(0,t] = \sum_{k\geq 1} 1\{(N_{k-1},N_k)\in H\}\, 1\{0 < T_k \leq t\}$ then $\Phi_H(0,T^\emptyset_n+t]-\Phi_H(0,T^\emptyset_n] = \sum_{k\geq 1} 1\{(N_{k-1},N_k)\in H\}\, 1\left\{T^\emptyset_n < T_k \leq T^\emptyset_n + t\right\}$ which shows that Φ_H is cumulative with respect to $\left\{T^\emptyset_n\right\}$.

Proposition 26 *Suppose that $\emptyset$ is a positive recurrent state for the generator q and let Φ' be a point process on $\mathbb{R}$. We have*

$$\mathbf{E}_\Pi[\Phi'(0,1]] = \lambda_\emptyset \mathbf{E}_\emptyset\left[\Phi'(0,T^\emptyset]\right] \tag{6.42}$$

Moreover

(i) *If Φ' is cumulative with respect to $\Phi_\emptyset$ and $\mathbf{E}_\Pi[\Phi'(0,1]] < \infty$, then*

$$\lim_{t\to\infty} t^{-1}\Phi'(0,t] = \mathbf{E}_\Pi[\Phi'(0,1]], \quad a.s. \tag{6.43}$$

for all initial state ν such that $\mathbf{P}\left\{T^\emptyset < \infty | N_0 = \nu\right\} = 1$.

(ii) *If Φ' is a point process counting the H-transitions, then*

$$\mathbf{E}_\Pi[\Phi'(0,1]] = \mathbf{E}_\Pi[q(N, H_N)] \tag{6.44}$$

where $H_N = \{\nu \in \mathbb{M} : (N,\nu) \in H\}$.

Proof. We aim now to apply the Swiss army formula [17, Formula (1.3.28) p.29]. With the notations of the previous reference, we define $A = \Phi_\emptyset$, D such that $\tau_n = T^\emptyset_{n+1}$, $X(t) = 1$, $Z(t) = 1$, $B(t) = \Phi'(0,t]$, then we have $W_0 = \tau_0 - T^\emptyset_0 = T^\emptyset_1 - T^\emptyset_0 = T^\emptyset_1$, $\mathbf{P}_\emptyset$-a.s. Then the Swiss army formula gives (6.42).

(i) We aim now to apply [10, Theorem 3.1 p.178] to the process $\{\Phi'(0,t]\}$. Let $U = \Phi'(0,T^\emptyset]$ and $V = \max_{0\leq t<T^\emptyset} |\Phi'(0,t]| = |U|$. By (6.42), we have $E_\emptyset|U| < \infty$ then $E_\emptyset[V] < \infty$. Then [10, Theorem 3.1 p.178] gives $\lim_{t\to\infty} t^{-1}\Phi'(0,t] = \lambda_\emptyset \mathbf{E}_\emptyset\left[\Phi'(0,T^\emptyset]\right]$, a.s. which together with (6.42) implies (6.43).

(ii) We have

$$\begin{aligned}
\mathbf{E}_\Pi\left[\Phi'(0,1]\right] &= \mathbf{E}_\Pi\left[\sum_{n\in\mathbb{Z}} 1\left\{(N_{n-1},N_n)\in H\right\}1\left\{0<T_n\le 1\right\}\right]\\
&= \mathbf{E}_\Pi\left[\int_0^1\int_{\mathbb{M}} 1\left\{(N_s,\nu)\in H\right\}q\left(N_s,d\nu\right)ds\right]\\
&= \mathbf{E}_\Pi\left[\int_0^1 q\left(N_s,H_{N_s}\right)ds\right]\\
&= \mathbf{E}_\Pi\left[q\left(N,H_N\right)\right] \quad \text{(by stationarity)}
\end{aligned}$$

where the second equality is due to Lévy's formula.

■

Corollary 7 *If $\emptyset$ is a positive recurrent state for the generator q, then the intensity λ of Φ is given by $\lambda = \mathbf{E}_\Pi\left[q\left(N\right)\right] > 0$.*

Proof. From Proposition 26 (ii) we have $\lambda = \mathbf{E}_\Pi\left[q\left(N\right)\right]$. We aim now to show that λ is non-null. Observe that $\lambda = \mathbf{E}_\Pi\left[q\left(N\right)\right] \ge q\left(\emptyset\right)\Pi\left(\emptyset\right) = q\left(\emptyset,\mathbb{M}\right)\Pi\left(\emptyset\right)$. Note first that $0 < 1/q\left(\emptyset,\mathbb{M}\right) = E_\emptyset\left[T_1\right] \le E_\emptyset\left[T^\emptyset\right] < \infty$ by assumption; then $q\left(\emptyset,\mathbb{M}\right) > 0$. On the other hand by (6.30) $\Pi\left(\emptyset\right) = E_\emptyset\left[T_1\right]/E_\emptyset\left[T^\emptyset\right] > 0$, then $\lambda > 0$. ■

Remark 15 *Here is another proof of (6.44). Let $Z_k = 1\left\{(N_{k-1},N_k)\in H\right\}$. Note that $\mathbf{E}_\emptyset\left[\Phi'(0,T^\emptyset]\right]$ equals*

$$\begin{aligned}
&\mathbf{E}_\emptyset\left[\int_{(0,T^\emptyset]} Z_0\circ\theta_t\Phi\left(dt\right)\right]\\
&= \mathbf{E}_\emptyset\left[\sum_{k\ge1} Z_k 1\left\{T_k\le T^\emptyset\right\}\right]\\
&= \mathbf{E}_\emptyset\left[\sum_{k\ge1} Z_k 1\left\{T_k\le T^\emptyset\right\}q\left(N_{k-1}\right)\mathbf{E}_\emptyset\left[T_k-T_{k-1}|\left\{N_n\right\}\right]\right]\\
&= \mathbf{E}_\emptyset\left[\sum_{k\ge1} \mathbf{E}_\emptyset\left[Z_k 1\left\{T_k\le T^\emptyset\right\}q\left(N_{k-1}\right)\left(T_k-T_{k-1}\right)|\left\{N_n\right\}\right]\right]\\
&= \mathbf{E}_\emptyset\left[\sum_{k\ge1} Z_k 1\left\{T_k\le T^\emptyset\right\}q\left(N_{k-1}\right)\left(T_k-T_{k-1}\right)\right]\\
&= \mathbf{E}_\emptyset\left[\int_0^{T^\emptyset} Z_0\circ\theta_t q\left(N_{-1}\circ\theta_t\right)dt\right]\\
&= \lambda_\emptyset^{-1}\mathbf{E}_\Pi\left[Z_0 q\left(N_{-1}\right)\right] \quad \textit{(Inversion formula [17, Formula (1.2.25) p.20])}\\
&= \lambda_\emptyset^{-1}\mathbf{E}_\Pi\left[q\left(N_{-1}\right)1\left\{(N_{-1},N_0)\in H\right\}\right] \qquad (6.45)
\end{aligned}$$

where for the third equality we use $\left\{T_k \leq T^{\emptyset}\right\} = \{N_1 \neq \emptyset, \ldots, N_{k-1} \neq \emptyset\}$*. The sequence* $\{N_k\}_{k\in\mathbb{Z}}$ *is a discrete time Markov jump process with transition Kernel* $P(\nu, d\mu) = q(\nu, d\mu)/q(\nu)$*, then, for any* $A, A_0 \in \mathcal{M}$*,*

$$\begin{aligned}\mathbf{P}_\Pi \{N_{-1} \in A, N_0 \in A_0\} &= \int_A \Pi(d\nu) \frac{q(\nu, A_0)}{q(\nu)} \\ &= \int_{\mathbb{M}} \Pi(d\nu) 1\{\nu \in A\} \int_{\mathbb{M}} \frac{q(\nu, d\nu_0)}{q(\nu)} 1\{\nu_0 \in A_0\}\end{aligned}$$

then, for any measurable function $f : \mathbb{M} \times \mathbb{M} \to \mathbb{R}_+$*,*

$$\mathbf{E}_\Pi [f(N_{-1}, N_0)] = \int_{\mathbb{M}} \Pi(d\nu) \int_{\mathbb{M}} \frac{q(\nu, d\nu_0)}{q(\nu)} f(\nu, \nu_0)$$

In particular

$$\begin{aligned}\mathbf{E}_\Pi [q(N_{-1}) 1\{(N_{-1}, N_0) \in H\}] &= \int_{\mathbb{M}} \Pi(d\nu) \int_{\mathbb{M}} \frac{q(\nu, d\nu_0)}{q(\nu)} q(\nu) 1\{(\nu, \nu_0) \in H\} \\ &= \mathbf{E}_\Pi \left[\int_{\mathbb{M}} 1\{(N_{-1}, \nu) \in H\} q(N_{-1}, d\nu)\right] \\ &= \mathbf{E}_\Pi \left[q\left(N_{-1}, H_{N_{-1}}\right)\right] \qquad (6.46)\end{aligned}$$

Equations (6.45) and (6.46) implies $\lambda_\emptyset \mathbf{E}_\emptyset \left[\Phi'(0, T^{\emptyset}]\right] = \mathbf{E}_\Pi [q(N, H_N)]$ *which together with (6.42) imply (6.44).*

6.3.5 Countable case

Proposition 27 *Consider an at most countable SMQ. Suppose that the SMQ generator* q *given by (6.8) is regular and irreducible. Suppose that a probability measure* Π *is invariant with respect to* q*. Then* q *is ergodic.*

Proof. By [34, Theorem 4.17 p.129], Π is invariant with respect to the transition functions associated to q, then [8, Proposition 5.1.6 p.160] finishes the proof. ■

6.4 Invariant probability measure

In this section we gather results concerning invariant measures for the SMQ generator q given by (6.8).

The formula (6.30) does not give Π in an explicit form. Studying the global balance equation, we can sometimes express Π in a more tractable way. This the case when the routing kernel $\lambda(\cdot, \cdot)$ satisfies certain traffic equations and $r(\cdot, \cdot)$ are "balanced" in some way which will be defined soon.

Definition 2 TRAFFIC EQUATIONS. *We call a locally finite measure* $\rho(\cdot)$ *on* $\bar{\mathbb{D}}$ *(defined in (6.7)) a* SOLUTION OF THE TRAFFIC EQUATIONS *if*

$$\rho\{o\} = 1, \quad \int_B \lambda\left(x, \bar{\mathbb{D}}\right) \rho(dx) = \int_{\bar{\mathbb{D}}} \lambda(y, B) \rho(dy), \quad B \in \bar{\mathcal{D}} \qquad (6.47)$$

Moreover, we will say that $\lambda(\cdot,\cdot)$ *is* reversible with respect to ρ *if the equations (6.47) hold with* $\bar{\mathbb{D}}$ *replaced by any* $A \in \bar{\mathcal{D}}$.

Note that (6.47) may be decomposed in two equations

$$\lambda\left(o,\bar{\mathbb{D}}\right)=\int_{\bar{\mathbb{D}}}\lambda\left(y,\{o\}\right)\rho\left(dy\right) \tag{6.48}$$

$$\int_{B}\lambda\left(x,\bar{\mathbb{D}}\right)\rho\left(dx\right)=\int_{\bar{\mathbb{D}}}\lambda\left(y,B\right)\rho\left(dy\right),\quad B\in\mathcal{D} \tag{6.49}$$

By standard arguments from indicator functions, to simple functions and then to non-negative functions, we may show that

$$\int_{\bar{\mathbb{D}}}\lambda\left(x,\bar{\mathbb{D}}\right)h\left(x\right)\rho\left(dx\right)=\int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}}h\left(x\right)\lambda\left(y,dx\right)\rho\left(dy\right) \tag{6.50}$$

for every $\bar{\mathcal{D}}$-measurable function $h:\bar{\mathbb{D}}\to\mathbb{R}_{+}$.

Definition 3 Service rate balance. *We say that the service rates* $r:\bar{\mathbb{D}}\times\bar{\mathbb{D}}\times\mathbb{M}\to\mathbb{R}^{+},(x,y,\nu)\mapsto r_{xy}\left(\nu\right)$ *are* balanced *if there exists some measurable function* $\Psi:\mathbb{M}\to(0,\infty)$ *such that*

$$\Psi\left(\nu\right)r\left(\nu,T_{xy}\nu\right)=\Psi\left(T_{xy}\nu\right)r\left(T_{xy}\nu,\nu\right),\quad x\neq y\in\bar{\mathbb{D}},\nu\in\mathbb{M},\nu_{x}>0$$

In particular Whittle service rates $\psi_{x}\left(\nu\right)$ are Ψ-balanced iff

$$\Psi\left(\nu\right)\psi_{x}\left(\nu\right)=\Psi\left(T_{xy}\nu\right)\psi_{y}\left(T_{xy}\nu\right),\quad x\neq y\in\bar{\mathbb{D}},\nu\in\mathbb{M},\nu_{x}>0$$

Remark 16 *Serfozo [106] calls the quantity*

$$\phi_{x}\left(\nu\right)=\nu_{x}\psi_{x}\left(\nu\right)$$

service rates. In fact $\phi_{x}\left(\nu\right)$ *may be viewed as the service rate per location and* $\psi_{x}\left(\nu\right)$ *may be viewed as the service rate per user. It is easy to see that if* ψ_{x} *is* Ψ*-balanced, then* ϕ_{x} *is* Φ*-balanced where*

$$\Phi\left(\nu\right)=\Psi\left(\nu\right)\frac{1}{\prod_{x\in\operatorname{supp}(\nu)}\nu_{x}!}$$

6.4.1 Balance for Whittle SMQ

Proposition 28 *Consider a Whittle generator* q *such that the traffic equations (6.47) have a solution some measure* ρ *on* $\bar{\mathbb{D}}$ *and such that the service rates are balanced with respect to some measurable function* $\Psi:\mathbb{M}\to(0,\infty)$. *Suppose that*

$$\rho\left(\bar{\mathbb{D}}\right)<\infty \tag{6.51}$$

and

$$\mathbf{E}_{\Pi_{\rho}}\left[\Psi\right]=\int_{\mathbb{M}}\Psi\left(\nu\right)\Pi_{\rho}\left(d\nu\right)<\infty \tag{6.52}$$

where Π_ρ is the distribution of the Poisson process on $\mathbb{D}$ with intensity measure ρ. If we normalize Ψ such that $E_{\Pi_\rho}[\Psi] = 1$, then the Gibbs distribution Π_Ψ having density Ψ with respect to the Poisson process Π_ρ is an invariant measure for q. (The Gibbs distribution is defined in Example 2.)

Proof. By the form of the SMQ generator q, the left-hand side of (6.5) may be written

$$\begin{aligned} I &= \int_\Gamma q(\nu, \mathbb{M}) \Pi_\Psi(d\nu) \\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) 1_\Gamma(\nu) \left[\psi_o(\nu) \lambda\left(o, \bar{\mathbb{D}}\right) + \int_{\mathbb{D}} \nu(dx) 1_\Gamma(\nu) \psi_x(\nu) \lambda\left(x, \bar{\mathbb{D}}\right)\right] \end{aligned} \tag{6.53}$$

Note that the second term in the right-hand side of Equation (6.53) can be written as

$$I_2 = \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \nu(dx) g(x, \nu - \delta_x)$$

where

$$g(x, \nu) = 1_\Gamma(\nu + \delta_x) \psi_x(\nu + \delta_x) \lambda\left(x, \bar{\mathbb{D}}\right)$$

Since Π_Ψ is the Gibbs distribution, by Proposition 20,

$$\begin{aligned} I_2 &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \rho(dx) g(x, \nu) \frac{\Psi(\nu + \delta_x)}{\Psi(\nu)} \\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \rho(dx) 1_\Gamma(\nu + \delta_x) \psi_x(\nu + \delta_x) \lambda\left(x, \bar{\mathbb{D}}\right) \frac{\Psi(\nu + \delta_x)}{\Psi(\nu)} \end{aligned}$$

Note also that $\rho(\{o\}) = 1$ and interpreting $\nu + \delta_o \equiv \nu$ we can write the first term in the right-hand side of Equation (6.53) as

$$I_1 = \int_{\mathbb{M}} \Pi_\Psi(d\nu) \rho(\{o\}) 1_\Gamma(\nu + \delta_o) \psi_o(\nu + \delta_o) \lambda\left(o, \bar{\mathbb{D}}\right) \frac{\Psi(\nu + \delta_o)}{\Psi(\nu)}$$

Consequently

$$\begin{aligned} I &= I_1 + I_2 \\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\bar{\mathbb{D}}} \rho(dx) 1_\Gamma(\nu + \delta_x) \psi_x(\nu + \delta_x) \lambda\left(x, \bar{\mathbb{D}}\right) \frac{\Psi(\nu + \delta_x)}{\Psi(\nu)} \end{aligned}$$

Consider the function $h : \bar{\mathbb{D}} \to \mathbb{R}_+$ defined by

$$h(x) = 1_\Gamma(\nu + \delta_x) \psi_x(\nu + \delta_x) \Psi(\nu + \delta_x)$$

We have

$$\begin{aligned}\int_{\bar{\mathbb{D}}} \rho(dx)\,\lambda\left(x,\bar{\mathbb{D}}\right)h(x) &= \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} \rho(dy)\,h(x)\,\lambda(y,dx)\\ &= \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} \rho(dy)\,1_\Gamma(\nu+\delta_x)\,\psi_x(\nu+\delta_x)\,\Psi(\nu+\delta_x)\,\lambda(y,dx)\\ &= \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} \rho(dy)\,1_\Gamma(\nu+\delta_x)\,\psi_y(\nu+\delta_y)\,\Psi(\nu+\delta_y)\,\lambda(y,dx)\\ &= \int_{\bar{\mathbb{D}}} \rho(dy)\,\psi_y(\nu+\delta_y)\,\Psi(\nu+\delta_y)\,\lambda(y,T_{o\Gamma}\nu)\end{aligned}$$

where we use (6.50) for the first equality and Ψ-balance property for the third one. (The notation $T_{o\Gamma}\nu$ is defined by (6.11).) Hence

$$I = \int_{\mathbb{M}} \Pi_\Psi(d\nu)\int_{\bar{\mathbb{D}}} \rho(dy)\,\psi_y(\nu+\delta_y)\,\frac{\Psi(\nu+\delta_y)}{\Psi(\nu)}\lambda(y,T_{o\Gamma}\nu)$$

The right-hand side of (6.5) may be written

$$\begin{aligned}J &= \int_{\mathbb{M}} q(\nu,\Gamma)\,\Pi_\Psi(d\nu)\\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu)\left[\psi_o(\nu)\,\lambda(o,T_{o\Gamma}\nu) + \int_{\mathbb{D}} \nu(dx)\,\psi_x(\nu)\,\lambda(x,T_{x\Gamma}\nu)\right] \qquad (6.54)\end{aligned}$$

Note that the second term in the right-hand side of Equation (6.54) can be written as

$$J_2 = \int_{\mathbb{M}} \Pi_\Psi(d\nu)\int_{\mathbb{D}} g(x,\nu-\delta_x)\,\nu(dx)$$

where

$$g(x,\nu) = \psi_x(\nu+\delta_x)\,\lambda(x,T_{o\Gamma}(\nu))$$

Since Π_Ψ is the Gibbs distribution, by Proposition 20,

$$\begin{aligned}J_2 &= \int_{\mathbb{M}} \Pi_\Psi(d\nu)\int_{\mathbb{D}} \rho(dx)\,g(x,\nu)\,\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)}\\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu)\int_{\mathbb{D}} \rho(dx)\,\psi_x(\nu+\delta_x)\,\lambda(x,T_{o\Gamma}(\nu))\,\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)}\end{aligned}$$

Note also that $\rho(\{o\}) = 1$ and interpreting $\nu+\delta_o \equiv \nu$ we can write the first term in the right-hand side of Equation (6.54) as

$$J_1 = \int_{\mathbb{M}} \Pi_\Psi(d\nu)\,\rho(\{o\})\,\psi_o(\nu+\delta_o)\,\lambda(o,T_{o\Gamma}(\nu))\,\frac{\Psi(\nu+\delta_o)}{\Psi(\nu)}$$

Consequently

$$\begin{aligned}J &= J_1 + J_2\\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu)\int_{\bar{\mathbb{D}}} \rho(dx)\,\psi_x(\nu+\delta_x)\,\lambda(x,T_{o\Gamma}(\nu))\,\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)}\end{aligned}$$

Note that $J = I$ which finishes the proof. ■

The Gibbs distribution in Proposition 28 is given by

$$\Pi_\Psi(\Gamma) = \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}\rho(\mathbb{D})^n}{n!} \Pi_\Psi(\Gamma) = \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}\rho(\mathbb{D})^n}{n!} \int_\Gamma \Psi(\nu)\,\Pi_\rho^{(n)}(d\nu), \quad \Gamma \in \mathcal{M}$$

where $\Pi_\rho^{(n)}$ is given by

$$\Pi_\rho^{(n)}(\Gamma) = \int_{\mathbb{D}^n} 1\left(\sum_{k=1}^{n} \delta_{x_k} \in \Gamma\right) \frac{\rho(dx_1)}{\rho(\mathbb{D})} \cdots \frac{\rho(dx_n)}{\rho(\mathbb{D})}, \quad n \in \mathbb{N}, \Gamma \in \mathcal{M}$$

Remark 17 *An invariant measure for the spatial Whittle generator is proposed in Serfozo's pioneer work [106, 67]. We shall now compare this invariant measure with the invariant distribution given in Proposition 28.*

Note that the term π_n in [106, Equation (10.5) p.269] and [67, Equation (5)] should be replaced by $\frac{1}{n!}\pi_n$ as done in [107]. Indeed the proof in [106, 67] accounts well for the motions but not for the births and deaths.

With this replacement, the invariant distribution given in Proposition 28 is just a renomalized version of Serfozo's invariant measure.

Example 5 *In the particular case where $\mathbb{D}$ is finite $\mathbb{D} = \{1, 2, \ldots, D\}$, we have, for $\nu \in \mathbb{M}$,*

$$\Pi_\rho^{(n)}(\nu) = \binom{n}{\nu_1 \ldots \nu_D} \frac{1}{\rho(\mathbb{D})^n} \prod_{x\in\nu} \rho_x^{\nu_x} = \frac{n!}{\rho(\mathbb{D})^n} \prod_{x\in\nu} \frac{\rho_x^{\nu_x}}{\nu_x!}$$

where we use for $\nu = (\nu_1, \ldots, \nu_D)$ the notation $\nu = \sum_{x\in\mathbb{D}} \nu_x \delta_x$ and denote $n = \nu(\mathbb{D})$. Then

$$\Pi_\Psi(\nu) = e^{-\rho(\mathbb{D})} \Psi(\nu) \prod_{x\in\nu} \frac{\rho_x^{\nu_x}}{\nu_x!}$$

In this case

$$\begin{aligned}
\mathbf{E}_{\Pi_\rho}[\Psi] &= \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}\rho(\mathbb{D})^n}{n!} \mathbf{E}_{\Pi_{\rho n}}[\Psi] \\
&= \sum_{n=0}^{\infty} \frac{e^{-\rho(\mathbb{D})}\rho(\mathbb{D})^n}{n!} \sum_{\nu:\nu(\mathbb{D})=n} \Psi(\nu) \frac{n!}{\rho(\mathbb{D})^n} \prod_{x\in\nu} \frac{\rho_x^{\nu_x}}{\nu_x!} \\
&= e^{-\rho(\mathbb{D})} \sum_{n=0}^{\infty} \sum_{\nu:\nu(\mathbb{D})=n} \Psi(\nu) \prod_{x\in\nu} \frac{\rho_x^{\nu_x}}{\nu_x!} \qquad (6.55)
\end{aligned}$$

6.4.2 Balance for SMQ

Suppose that λ is reversible with respect to some measure ρ on $\bar{\mathbb{D}}$, i.e.

$$\int_B \lambda(x, C)\, \rho(dx) = \int_C \lambda(y, B)\, \rho(dy), \quad B, C \in \bar{\mathcal{D}} \tag{6.56}$$

Note that (6.56) may be decomposed in two equations

$$\lambda(o, C) = \int_C \lambda(y, \{o\})\, \rho(dy), \quad C \in \bar{\mathcal{D}} \tag{6.57}$$

$$\int_B \lambda(x, C)\, \rho(dx) = \int_C \lambda(y, B)\, \rho(dy), \quad B \in \mathcal{D}, C \in \bar{\mathcal{D}} \tag{6.58}$$

By standard arguments from indicator function, to simple functions and then to non-negative functions, it is easy to show that

$$\int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} h(x, y)\, \lambda(x, dy)\, \rho(dx) = \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} h(x, y)\, \lambda(y, dx)\, \rho(dy) \tag{6.59}$$

for every $\bar{\mathcal{D}} \times \bar{\mathcal{D}}$-measurable function $h : \bar{\mathbb{D}} \times \bar{\mathbb{D}} \to \mathbb{R}_+$.

Proposition 29 *If the conditions of Proposition 28 are satisfied for a SMQ generator q given by (6.8) and if λ is reversible with respect to ρ, then q is reversible with respect to Π_Ψ.*

Proof. It suffices to prove that

$$I(\Gamma, K) = \int_\Gamma q(\nu, K)\, \Pi_\Psi(d\nu)$$

is symmetric with respect to Γ, K, i.e.; that $I(\Gamma, K) = I(K, \Gamma)$ for all $K, \Gamma \in \mathcal{M}$. By the form of the SMQ generator q

$$\begin{aligned} I(\Gamma, K) &= \int_\Gamma q(\nu, K)\, \Pi_\Psi(d\nu) \\ &= \left[\int_{\mathbb{M}} \Pi_\Psi(d\nu)\, 1_\Gamma(\nu) \int_{\bar{\mathbb{D}}} \lambda(o, dy)\, r(\nu, T_{oy}\nu)\, 1_K(T_{oy}\nu)\right] \\ &+ \left[\int_{\mathbb{D}} \nu(dx) \int_{\bar{\mathbb{D}}} \lambda(x, dy)\, r(\nu, T_{xy}\nu)\, 1_K(T_{xy}\nu)\right] \end{aligned} \tag{6.60}$$

Note that the second term in the right-hand side of Equation (6.60) can be written as

$$I_2 = \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \nu(dx)\, g(x, \nu - \delta_x)$$

where

$$g(x, \nu) = 1_\Gamma(\nu + \delta_x) \int_{\bar{\mathbb{D}}} \lambda(x, dy)\, r(\nu + \delta_x, T_{oy}\nu)\, 1_K(T_{oy}\nu)$$

interpreting $T_{oo}\nu \equiv \nu$. Since Π_Ψ is the Gibbs distribution, by Proposition 20,

$$\begin{aligned} I_2 &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \rho(dx)\, g(x,\nu) \frac{\Psi(\nu+\delta_x)}{\Psi(\nu)} \\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\mathbb{D}} \rho(dx) \frac{\Psi(\nu+\delta_x)}{\Psi(\nu)} 1_\Gamma(\nu+\delta_x) \int_{\bar{\mathbb{D}}} \lambda(x,dy)\, r(\nu+\delta_x, T_{oy}\nu)\, 1_K(T_{oy}\nu) \end{aligned}$$

Note also that $\rho(\{o\}) = 1$ and interpreting $\nu + \delta_o \equiv \nu$ we can write the first term in the right-hand side of Equation (6.60) as

$$I_1 = \int_{\mathbb{M}} \Pi_\Psi(d\nu)\, \rho(\{o\}) \frac{\Psi(\nu+\delta_o)}{\Psi(\nu)} 1_\Gamma(\nu+\delta_o) \int_{\bar{\mathbb{D}}} \lambda(o,dy)\, r(\nu+\delta_o, T_{oy}\nu)\, 1_K(T_{oy}\nu)$$

Consequently

$$\begin{aligned} I(\Gamma,K) &= I_1 + I_2 \\ &= \int_{\mathbb{M}} \Pi_\Psi(d\nu) \int_{\bar{\mathbb{D}}\times\bar{\mathbb{D}}} \rho(dx)\, \lambda(x,dy)\, 1(T_{ox}\nu \in \Gamma, T_{oy}\nu \in K)\, r(T_{ox}\nu, T_{oy}\nu) \frac{\Psi(T_{ox}\nu)}{\Psi(\nu)} \end{aligned}$$

The symmetry of $I(\cdot,\cdot)$ follows from (6.59) and the symmetry of $\Psi(\nu)\, r(\nu,\zeta)$.
■

Remark 18 *Proposition 29 is due to Serfozo [107, 66]. A more general result is given in [108, 109].*

6.4.3 Invariant probability measure

Proposition 30 *Assume that the generator q given by (6.8) is regular and let Π be a probability measure on $(\mathbb{M}, \mathcal{M})$. The following statements hold true.*

(i) If Π satisfies the balance equations (6.4), then Π is invariant with respect to the transition functions associated to q.

(ii) If Π is reversible with respect to q, then Π is reversible with respect to the transition functions associated to q.

Proof.

(i) (Cf. [34, Theorem 4.17 p.129]).

(ii) (Cf. [34, Theorem 6.7 p.230]).

■

6.5 Mobility process for wireless networks

We give in the present section a Markov jump mobility process which may be used for wireless networks. The reader may doubt about the relevance of a Markov jump process to model mobility in wireless networks. In fact we don't assume that the user remains at a given position for some time and then jumps to another place, but rather we account for the effect of the mobility on the state of our system only if the position of the user changes appreciably.

Denote by $\mathbb{D} \subset \mathbb{R}^2$ the cell served by the given base station $u = 0$. We may model the internal mobility of users is $\mathbb{D}$ by a Markov jump process.

Specifically, assume that the users move independently of each other in $\mathbb{D}$. The sojourn duration of a given user at location $x \in \mathbb{D}$ is exponentially distributed with parameter $\lambda'(x)$. Any user finishing its sojourn at location x, is routed to a new location dy according to some probability kernel $p'(x, dy)$, where $p'(x, \mathbb{D}) = 1$.

The above description corresponds to the Markov process on $\mathbb{D}$ with the following generator (of the individual user mobility): $\lambda(x, dy) = \lambda'(x)p'(x, dy)$.

A probability measure $\varrho(\cdot)$ on $\mathbb{D}$ is invariant for this mobility Markov process iff it satisfies the following equations

$$\varrho(\mathbb{D}) = 1, \quad \int_A \lambda(x, \mathbb{D})\varrho(dx) = \int_{\mathbb{D}} \lambda(x, A)\varrho(dx), \quad A \in \mathcal{D} \tag{6.61}$$

6.5.1 Completely aimless mobility

We show now an example where we can calculate concretely the above mobility parameters $\lambda'(x)$ and $p'(x, dy)$.

We consider the model described in [115]. It is based on the following assumptions:

- The velocity vectors of the users are independent.
- The velocity direction is uniformly distributed in $[0, 2\pi)$.

The authors of [70] call this model *completely aimless motion*.

Sojourn duration

We are interested in the user's sojourn duration in a given geographic zone of area A and perimeter L. Much as in [115], we derive a relation between the average sojourn duration and the average velocity.

We are interested in the users crossing an infinitesimal element dl of the border (for example from outside to inside) within an infinitesimal duration dt. Such users are located in a rectangle of sides dl and $V \cos \alpha dt$, as illustrated in Figure 6.1, where:

- V is the user's velocity magnitude;

- and α is the angle formed by the user's velocity vector and the perpendicular to dl.

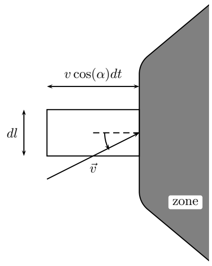

Figure 6.1: Rectangle containing customers crossing an element dl of the border during dt

Integrating over V and α, we obtain the average number of users crossing an element dl of the border of the zone, from outside to inside, during dt

$$\int_{-\pi/2}^{\pi/2}\int_{0}^{+\infty} VF\left(dV\right)\cos\alpha\frac{d\alpha}{2\pi}\rho dldt = \frac{\upsilon\rho}{\pi}dldt$$

where ρ is the intensity of users per surface unit, F is the cumulative distribution function of the user's velocity and $\upsilon = \mathbf{E}\left[V\right]$. Then the average number of users crossing the zone border per time-unit denoted λ (which is the average arrival rate of users to the zone) is given by

$$\lambda = \upsilon\rho L/\pi \tag{6.62}$$

where L is the perimeter of the zone.

Denote τ the average sojourn time of a user in a zone and $\bar{M}$ the average number of users in the zone. By Little's formula, we have

$$\bar{M} = \lambda\tau$$

which gives

$$\tau = \bar{M}\lambda^{-1} = \rho A\frac{\pi}{\upsilon\rho L} = \frac{\pi}{\upsilon}\frac{A}{L} \tag{6.63}$$

where A is the surface of the zone. For a disc of radius R, we have $A/L = R/2$.

The authors of [70] consider an exponential distribution for the sojourn time. This assumption is justified by [35].

6.5.2 Intracell mobility

The cell $\mathbb{D}$ is modeled by a disc of radius R which is divided into J rings. Each ring denoted by $j = 1, \ldots, J$ is delimited by discs with radii r_{j-1} and r_j where

$r_0 = 0$ and $r_J = R$. Let $A_j = \pi \left(r_j^2 - r_{j-1}^2 \right)$ be the surface of ring j. Of course J should be large enough to capture correctly the geometry of the problem.

Consider the case where mobility is within a given cell. Denote λ'_j the inverse of the average sojourn duration of users at ring j. Applying Equation (6.63) gives

$$\begin{array}{l} \lambda'_j = \frac{\upsilon}{\pi} \frac{L_j}{A_j} = 2\upsilon \frac{r_j + r_{j-1}}{A_j}, \quad j = 1, \ldots, J-1 \\ \lambda'_J = \frac{\upsilon}{\pi} \frac{L_J}{A_J} = 2\upsilon \frac{r_{J-1}}{A_J} \end{array}$$

A user finishing its sojourn at ring j is routed:

- either to ring $j-1$ or to ring $j+1$ with respective probabilities $p'_{j,j-1} = r_{j-1} / (r_j + r_{j-1})$ and $p'_{j,j+1} = r_j / (r_j + r_{j-1})$, if $j = 2, \ldots, J-1$;
- to ring 2 with probability 1, if $j = 1$;
- to ring $J-1$ with probability 1, if $j = J$.

Hence

$$\begin{array}{l} \lambda_{j,j-1} = 2\upsilon \frac{r_{j-1}}{A_j}, \quad j = 2, \ldots, J \\ \lambda_{j,j+1} = 2\upsilon \frac{r_j}{A_j}, \quad j = 1, \ldots, J-1 \end{array} \tag{6.64}$$

Proposition 31 *The mobility kernel $(\lambda_{jk}; j, k \in \{1, \ldots, J\})$ where the λ_{jk} are given by (6.64) admit*

$$\varrho_j = \frac{A_j}{\pi R^2}, \quad j = 1, \ldots, J \tag{6.65}$$

as invariant probability measure, i.e. solution of (6.61),

Proof. Equations (6.61) may be written as follows

$$\left\{ \begin{array}{ll} \varrho_j \left(\lambda_{j,j-1} + \lambda_{j,j+1} \right) = \varrho_{j-1} \lambda_{j-1,j} + \varrho_{j+1} \lambda_{j+1,j} & \text{for } j = 2, \ldots, J-1 \\ \varrho_1 \lambda_{1,2} = \varrho_2 \lambda_{2,1} & \\ \varrho_J \lambda_{J,J-1} = \varrho_{J-1} \lambda_{J-1,J} & \end{array} \right.$$

For the rates (6.64) we get

$$\left\{ \begin{array}{ll} \varrho_j \frac{r_j + r_{j-1}}{A_j} = \varrho_{j-1} \frac{r_{j-1}}{A_{j-1}} + \varrho_{j+1} \frac{r_j}{A_{j+1}} & \text{for } j = 2, \ldots, J-1 \\ \varrho_1 \frac{r_1}{A_1} = \varrho_2 \frac{r_1}{A_2} & \\ \varrho_J \frac{r_{J-1}}{A_J} = \varrho_{J-1} \frac{r_{J-1}}{A_{J-1}} & \end{array} \right.$$

which clearly admits ϱ given by (6.65) as solution. ■

Proposition 32 Reversibility of the routing kernel. *Consider the motion rates (6.64), and let $\lambda_{oj} > 0$ and $\lambda_{jo} > 0$ be the arrival and departure rates respectively. Then the routing kernel $(\lambda_{jk}; j, k \in \{0, 1, \ldots, J\})$ is reversible, iff*

$$\lambda_{oj} \lambda_{jk} \lambda_{ko} = \lambda_{ok} \lambda_{kj} \lambda_{jo} \tag{6.66}$$

which is equivalent to

$$\frac{\lambda_{j+1,o}}{\lambda_{jo}} = \frac{\lambda_{o,j+1}}{A_{j+1}} \frac{A_j}{\lambda_{oj}} \tag{6.67}$$

in which case λ is reversible with respect to ρ defined by

$$\rho_0 = 1, \quad \rho_j = \frac{\lambda_{oj}}{\lambda_{jo}} \tag{6.68}$$

In the particular case where the intensity of arrivals per unit surface, say λ, is constant, i.e. $\lambda_{oj} = \lambda A_j$, Condition (6.67) is equivalent to

$$\lambda_{j+1,o} = \lambda_{jo}$$

Proof. If λ is reversible, then by Kolmogorov criterion, we should have (6.66). Inversely if (6.66) holds true, then

$$\frac{\lambda_{oj}}{\lambda_{jo}}\lambda_{jk} = \frac{\lambda_{ok}}{\lambda_{ko}}\lambda_{kj}, \quad \text{and } \frac{\lambda_{oj}}{\lambda_{jo}}\lambda_{jo} = \lambda_{oj}$$

hence λ is reversible with respect to ρ defined by (6.68).

Taking $k = j + 1$ in Equation (6.66) we get

$$\lambda_{oj}\frac{r_j}{A_j}\lambda_{j+1,o} = \lambda_{o,j+1}\frac{r_j}{A_{j+1}}\lambda_{jo}$$

and taking $k = j - 1$ in Equation (6.66) we get

$$\lambda_{oj}\frac{r_{j-1}}{A_j}\lambda_{j-1,o} = \lambda_{o,j-1}\frac{r_{j-1}}{A_{j-1}}\lambda_{jo}$$

which is true iff (6.67) holds true. ∎

Proposition 33 *Consider the motion rates (6.64), and let $\lambda_{oj} > 0$ and $\lambda_{jo} > 0$ be the arrival and departure rates respectively. Then for each speed $\upsilon \geq 0$, the traffic equations associated to the routing kernel $(\lambda_{jk}; j, k \in \{0, 1, \ldots, J\})$*

$$\rho_0(\upsilon) = 1 \text{ and } \rho_j(\upsilon)\sum_{k=0}^{J}\lambda_{jk} = \sum_{k=0}^{J}\rho_k(\upsilon)\lambda_{kj}, \quad j \in \{1, \ldots, J\} \tag{6.69}$$

admit a unique solution. Moreover

$$\lim_{\upsilon\to\infty}\rho_j(\upsilon) = A_j\frac{\sum_{k=1}^{J}\lambda_{ok}}{\sum_{k=1}^{J}A_k\lambda_{ko}}$$

Proof. The routing kernel $(\lambda_{jk}; j, k \in \{0, 1, \ldots, J\})$ is irreducible by the positivity of the arrival and departure rates. Since the state space $\{0, 1, \ldots, J\}$ is finite, the Markov process associated to the routing kernel is positive recurrent and admits an invariant measure ρ with positive terms and unique up to a multiplicative factor. Hence (6.69) admit a unique solution.

It remains to show that $\lim_{\upsilon\to\infty}\rho_j(\upsilon)$ exists and is finite for all $j \in \{1, \ldots, J\}$. To this end we may write equations (6.69) in the matrix form $\rho(\upsilon)\Lambda = 0$ where

$\Lambda = (\lambda_{jk}; j, k \in \{0, 1, \ldots, J\})$ with the diagonal terms chosen such that the row sums are null. Note that Λ may be written in the form

$$\Lambda = \upsilon\Lambda^{(0)} + \Lambda^{(1)}$$

where

$$\Lambda^{(0)} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & * & \lambda_{12}/\upsilon & 0 & \\ 0 & \lambda_{12}/\upsilon & * & \lambda_{23}/\upsilon & \\ \vdots & & \ddots & \ddots & \ddots \\ 0 & & & \lambda_{J-1,J}/\upsilon & * \end{pmatrix}$$

and

$$\Lambda^{(1)} = \begin{pmatrix} * & \lambda_{01} & \ldots & \lambda_{0J} \\ \lambda_{10} & * & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{J0} & 0 & \ldots & * \end{pmatrix}$$

with the diagonal terms chosen such that the row sums are null. Hence $\rho(\upsilon)$ satisfies

$$\rho(\upsilon)\left(\Lambda^{(0)} + \frac{1}{\upsilon}\Lambda^{(1)}\right) = 0$$

In the case where υ is large, we say that $\Lambda^{(0)}$ and $\frac{1}{\upsilon}\Lambda^{(1)}$ are the unperturbed generator and the perturbation respectively and since the unperturbed generator is reducible, the perturbation is said to be singular. The existence and finiteness of $\lim_{\upsilon\to\infty} \rho_j(\upsilon)$ may be deduced from [43].

Note that we may get the expression of the limit from [43], but we give a direct proof. Let $\rho_j = \lim_{\upsilon\to\infty} \rho_j(\upsilon)$ for all $j \in \{1, \ldots, J\}$. Dividing (6.69) by υ and letting $\upsilon \to \infty$ shows $(\rho_j, j \in \{1, \ldots, J\})$ satisfies the traffic equations of the mobility kernel and hence by Proposition 31

$$\rho_j = \alpha\frac{A_j}{\pi R^2}, \quad j = 1, \ldots, J$$

for some constant α. Summing equations (6.69) over $j = 1, \ldots, J$ we get

$$\sum_{k=1}^{J} \lambda_{ok} = \sum_{k=1}^{J} \rho_k(\upsilon)\lambda_{ko}$$

letting $\upsilon \to \infty$ in the display above gives

$$\alpha = \frac{\sum_{k=1}^{J} \lambda_{ok}}{\sum_{k=1}^{J} \frac{A_k}{\pi R^2}\lambda_{ko}}$$

which proves the formula given in the proposition for $\lim_{\upsilon\to\infty} \rho_j(\upsilon)$.

(Historically the above problem is treated in [105], [39], [43] and references therein. We may view the limiting behavior as υ goes to ∞ in terms of the so called aggregated process on the set of the ergodic classes of $\Lambda^{(0)}$, say $\{0,1\}$ with generator [124, Theorem 7.4]

$$\bar{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \varrho \end{pmatrix} \Lambda^{(1)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\sum_{k=1}^{J} \lambda_{ok} & \sum_{k=1}^{J} \lambda_{ok} \\ \sum_{k=1}^{J} \varrho_k \lambda_{ko} & -\sum_{k=1}^{J} \varrho_k \lambda_{ko} \end{pmatrix}$$

The above perturbation problem with countable sate space, and in particular the problem of finding an analytic series of $\rho(\upsilon)$ in $\frac{1}{\upsilon}$ is treated in [5] or [123].)
■

6.5.3 Intercell mobility

Denote λ'_u the inverse of the average sojourn duration of users at cell u. Applying Equation (6.63) gives

$$\lambda'_u = \frac{\upsilon}{\pi} \frac{L}{A} = \frac{2\upsilon}{\pi R} \tag{6.70}$$

where the perimeter equals $L = 2\pi R$ and the area equals $A = \pi R^2$. If each base station has 6 neighbours as in the infinite or toric hexagonal model, then

$$\lambda_{u,w} = \frac{1}{6} \lambda'_u = \frac{\upsilon}{3\pi R}$$

for each pair of neighbours u, w.

Note that in the case of intracell mobility, the average sejourn duration in ring j, say τ_j, may be deduced from Equation (6.63)

$$\begin{aligned} \tau_j &= \tfrac{\pi}{2\upsilon} (r_j - r_{j-1}), \quad j = 1, \ldots, J-1 \\ \tau_J &= \tfrac{\pi}{2\upsilon} \tfrac{r_J^2 - r_{J-1}^2}{r_{J-1}} \end{aligned}$$

Then the average sejourn duration in the cell is given by

$$\sum_{j=1}^{J} \tau_j = \frac{\pi}{2\upsilon} \frac{r_J^2}{r_{J-1}} = \frac{\pi R}{2\upsilon} \frac{R}{r_{J-1}}$$

If J is sufficiently large, then $\frac{R}{r_{J-1}} \sim 1$ and the right hand side of the above display is almost equal to $\frac{\pi R}{2\upsilon}$ which is the sojourn duration in a cell deduced directly from Equation (6.63).

6.5.4 Complete mobility

Consider now a network of hexagonal cells such that each one has exactly 6 neighbours. Each cell is approximated by a disc and divided into J rings. The cells are indexed by $u \in \mathbf{U} = \{1, \ldots, U\}$, and the rings by $j \in \mathbf{J} = \{1, \ldots, J\}$. The ring j of the cell u is indexed by $uj \in \mathcal{D} = \mathbf{J} \times \mathbf{U}$.

Remark 19 *We may alternatively index the rings of the cell u as follows $(u-1)J+1,\dots(u-1)J+J$. Hence each ring is identified by some $x=\{1,\dots,U\times J\}$. From a given such x, we may retreive the index of the corresponding cell u and ring j by the Euclidean division*

$$x-1=(u-1)J+(j-1), \quad 1\leq j\leq J$$

Denote λ'_{uj} the inverse of the average sojourn duration of users at ring uj. Applying Equation (6.63) gives

$$\lambda'_{uj}=\frac{\upsilon}{\pi}\frac{L_j}{A_j}=2\upsilon\frac{r_j+r_{j-1}}{A_j}, \quad j\in\mathbf{J}$$

A user finishing its sojourn at ring uj is routed:

- to either ring $u(j-1)$ or ring $u(j+1)$ with respective probabilities $p'_{uj,u(j-1)}=r_{j-1}/(r_j+r_{j-1})$ and $p'_{uj,u(j+1)}=r_j/(r_j+r_{j-1})$, if $j=2,\dots,J-1$;
- to ring $u2$ with probability 1, if $j=1$;
- to either ring $u(J-1)$ or ring wJ, where w is a neighbour of u, with respective probabilities $p'_{uJ,u(J-1)}=r_{J-1}/(r_J+r_{J-1})$ and $p'_{uJ,wJ}=\frac{1}{6}r_J/(r_J+r_{J-1})$, if $j=J$.

Hence

$$\begin{cases}\lambda_{uj,u(j-1)}=2\upsilon\frac{r_{j-1}}{A_j}, & j=2,\dots,J\\ \lambda_{uj,u(j+1)}=2\upsilon\frac{r_j}{A_j}, & j=1,\dots,J-1\\ \lambda_{uJ,wJ}=\frac{1}{3}\upsilon\frac{r_J}{A_J}, & w\text{ is a neighbour of }u\end{cases} \tag{6.71}$$

The result of Proposition 31 may be easily extended to the complete mobility case as follows

Proposition 34 *The mobility kernel $(\lambda_{uj,wk};uj,wk\in\mathbf{J}\times\mathbf{U})$ given by (6.71) admit*

$$\varrho_{uj}=\varrho_j=\frac{A_j}{\pi R^2}, \quad j\in\mathbf{J},u\in\mathbf{U}$$

as invariant probability measure, i.e. solution of (6.61).

Proof. Besides the proof of Proposition 31, it remains to show that

$$\varrho_J\left[\lambda_{uJ,u(J-1)}+\sum_w\lambda_{uJ,wJ}\right]=\varrho_{J-1}\lambda_{u(J-1),uJ}+\varrho_J\sum_w\lambda_{wJ,uJ}$$

which is equivalent to

$$\varrho_J\lambda_{uJ,u(J-1)}=\varrho_{J-1}\lambda_{u(J-1),uJ}$$

which holds true. ∎

Relation to intracell traffic equations

We aim in the present section to study the relation between the traffic equations associated to complete mobility and those associated to intracell mobility.

If the arrival and departure rates don't depend on the particular cell but only on the ring in which they occur, then it is easy to extend the results of Proposition 32 and Proposition 33 to the complete mobility case. Let's for example verify the reversibility annouced in Proposition 32 with respect to

$$\rho_{uo} = \rho_o = 1, \quad \rho_{uj} = \rho_j = \frac{\lambda_{oj}}{\lambda_{jo}}$$

Besides the proof of Proposition 31, it remains to show that

$$\rho_J \lambda_{uJ,wJ} = \rho_J \lambda_{wJ,uJ}$$

which clearly holds true.

The following proposition shows the relation between the solutions of the traffic equations associated to intracell and complete mobility.

Proposition 35 *Assume that the arrival and departure rates don't depend on the particular cell but only on the ring in which they occur. If* $\left(\rho_j; j \in \bar{\mathbf{J}}\right)$ *is solution of the traffic equations associated to the routing kernel* $(\lambda_{jk}; j, k \in \mathbf{J})$*, then* $\left(\rho_{uj}; uj \in \bar{\mathcal{D}}\right)$ *defined by*

$$\rho_{uo} = \rho_o = 1, \quad \rho_{uj} = \rho_j, \quad u \in \mathbf{U}, j \in \mathbf{J} \tag{6.72}$$

is solution of the traffic equations associated to the routing kernel $\left(\lambda_{uj,uk}; uj, uk \in \bar{\mathcal{D}}\right)$.

Proof. Assume that $\left(\rho_j; j \in \bar{\mathbf{J}}\right)$ is a solution of the traffic equations associated to the routing kernel $(\lambda_{jk}; j, k \in \mathbf{J})$. Let's verify that $\left(\rho_{uj}; uj \in \bar{\mathcal{D}}\right)$ defined by (6.72) satisfy the traffic equations associated to the routing kernel $\left(\lambda_{uj,uk}; uj, uk \in \bar{\mathcal{D}}\right)$, i.e.

$$\begin{cases} \rho_j \left(\lambda_{uj,o} + \lambda_{uj,u(j-1)} + \lambda_{uj,u(j+1)}\right) = \rho_o \lambda_{o,uj} + \rho_{j-1} \lambda_{u(j-1),uj} + \rho_{j+1} \lambda_{u(j+1),uj} & \text{for } j = 2, \ldots \\ \rho_1 \left(\lambda_{u1,o} + \lambda_{u1,u2}\right) = \rho_o \lambda_{o,u1} + \rho_2 \lambda_{u2,u1} \\ \rho_J \left[\lambda_{uJ,o} + \lambda_{uJ,u(J-1)} + \sum_w \lambda_{uJ,wJ}\right] = \rho_o \lambda_{o,uJ} + \rho_{J-1} \lambda_{u(J-1),uJ} + \sum_w \rho_J \lambda_{wJ,uJ} \end{cases}$$

which are clearly satisfied since $\lambda_{uj,uk} = \lambda_{jk}$ and $\sum_w \lambda_{uJ,wJ} = \sum_w \lambda_{wJ,uJ}$. ■

Chapter 7

SMQ for elastic traffic

We consider here elastic traffic (i.e. messages having some amount of data to transmit at a bit rate which may be chosen by the network).

The classical results for *discrete* single server queues are recalled in §14.A.1. In particular analytical formulae for the throughput and delay for a M/GI/1 multiclass processor-sharing queue are recalled. The aim of the present chapter is to extend these results to the case where the set of classes is *continuous*, and where there is *mobility* between the different classes.

To this end we consider two Markovian models for wireless networks serving elastic traffic: the *Whittle model* proposed for such networks in [25, § 2]; and a *wireless model*, that we describe later. The fundamental difference between the two models is that in the wireless model each user has some given data-volume to transmit during the whole sojourn in the system, while in the Whittle of [25, § 2] each user has a different data-volume to transmit at each visited location, and he does not move from this location until the end of the transmission of the required volume. Consequently, when the Whittle model approaches congestion, user mobility is being frozen, whereas it is not influenced by a congestion in the wireless model.

Unfortunately the ergodicity conditions for established in § 6.3 are not helpfull in the case of elastic traffic, thus we need to establish ergodicity by alternative ways.

7.1 Whittle model

7.1.1 Model

The following model is classical for modelling wired networks serving elastic traffic [98] in the discrete case, [106]):

(a2) Exogenous arrivals come to dx as a Poisson process with intensity $\lambda(o, dx)$.

(b2) A user finishing its service at location x is routed to dy according to the probability Kernel $p(x, dy)$ (Bernoulli routing) where $p(x, \bar{\mathbb{D}}) = 1$. As usual $y = o$ corresponds to the case where the user exits the system.

(c2) Define the LOCATION-CALL-VOLUME at a location x as the residual amount of data to transmit or receive at location x before a motion to another location or the exit from the system. The location-call-volumes are assumed i.i.d. exponentially distributed with parameters $\lambda(x)$ (or equivalently with mean $1/\lambda(x)$) and independent from arrivals. (Note that the location-call-volume equals the service duration if the user is served at rate 1.)

(d2) The state of the system is $\nu \in \mathbb{M}$ counting the users at each location. The rate allocated to the users at location x when the state of the system is ν equals $\nu_x \psi_x(\nu)$ (i.e. the rate allocated for each user at location x is $\psi_x(\nu)$).

Note that the mobility process here is different from that given in Section 6.5. In particular the sojourn duration of some user at a given position x depends on the bit-rate allocated by the network, which in its turn depends on the number and locations of the other users.

Proposition 36 *The physical interpretation (a2)–(d2) corresponds to the mathematical description (a1)–(c1) in Section 6.2.3 and thus to the generator (6.8) if*

$$\begin{cases} \lambda(x) = \lambda(x, \bar{\mathbb{D}}), & x \in \mathbb{D} \\ p(x, A) = \lambda(x, A) / \lambda(x, \bar{\mathbb{D}}), & x \in \mathbb{D}, A \in \bar{\mathcal{D}} \end{cases}$$

and

$$r_{xy}(\nu) = \begin{cases} 1, & x = o, y \in \mathbb{D} \\ \psi_x(\nu), & x \in \mathbb{D}, y \in \bar{\mathbb{D}} \end{cases}$$

Hence we get a Whittle network.

Proof. Clearly (a2) is the same as (a1) and (b2) is equivalent to (b1) if

$$p(x, A) = \lambda(x, A) / \lambda(x, \bar{\mathbb{D}})$$

It remains to show that (c2)-(d2) imply (c1). At a given location x there are ν_x users, each of which, by (c2) and (d2), will depart from location x after an exponentially-distributed duration of rate $\lambda(x)\psi_x(\nu)$. The next departure will then take place at the minimum of these durations which is exponentially distributed with rate $\lambda(x)\nu_x\psi_x(\nu)$, hence we obtain (c1) if

$$\lambda(x) = \lambda(x, \bar{\mathbb{D}})$$

■

Hence the departure rate $\lambda(x) \times \nu_x \psi_x(\nu)$ comprises two factors. The first one $\lambda(x)$ may be interpreted as the rate of an exponentially-distributed location-call-volume at location x (which is related to the motion within the system and the exist from the system) of a given user and the second one $\nu_x \psi_x(\nu)$ may be interpreted as the total rate at location x when the state of the system is ν ($\nu_x \psi_x(\nu)$ is called capacity by some authors [98]).

The Whittle model is classically used for wired networks, but it may be used for wireless networks. We will discuss the relevance of the Whittle model for wireless networks in Remark 21.

7.1.2 Service rate balance

The following lemmas will be useful in the study of the ergodicity and balance of the service rates of the Whittle model for wireless networks.

Lemma 9 *Let $\gamma : \mathbb{D} \to \mathbb{R}_+^*$ be some measurable function. If, for all $x \in \mathbb{D}, y \in \bar{\mathbb{D}}, A \in \bar{\mathcal{D}}$, we replace $\lambda(x, A)$ by $\lambda(x, A)/\gamma(x)$ and $r_{xy}(\nu)$ by $r_{xy}(\nu)\gamma(x)$ the generator (6.8) remains invariant.*

Proof. The products $\lambda(x, \{o\}) r_{xo}(\nu)$ and $\lambda(x, dy) r_{xy}(\nu)$ remain invariant, then the generator (6.8) remains invariant. ■

Corollary 8 *Consider a Whittle network where, for all $x \in \mathbb{D}, \nu \in \mathbb{M}$, the service rate $\psi_x(\nu)$ is in the form $1/(h(\nu(\mathbb{D}))\gamma(x))$ for two positive measurable functions γ and h. By replacing $\lambda(x, A)$ by $\lambda(x, A)/\gamma(x)$ and $\psi_x(\nu)$ by $1/h(\nu(\mathbb{D}))$ (for all $x \in \mathbb{D}, \nu \in \mathbb{M}, A \in \bar{\mathcal{D}}$) the generator (6.8) remains invariant.*

Proof. Immediate consequence of Lemma 9. ■

Lemma 10 *The service rates $\psi_x(\nu) = 1/h(\nu(\mathbb{D}))$ are balanced by*

$$\Psi(\nu) = Z^{-1} \prod_{n=1}^{\nu(\mathbb{D})} h(n)$$

where Z is a normalizing constant. Let ρ be some measure on $\mathbb{D}$ and Π_ρ be the distribution of a Poisson process of intensity measure ρ. If we normalize Ψ such that $E_{\Pi_\rho}[\Psi] = 1$, then

$$Z = \mathbf{E}_{\Pi_\rho}\left[\prod_{n=1}^{\nu(\mathbb{D})} h(n)\right] = \mathbf{E}\left[\prod_{n=1}^{X} h(n)\right] \tag{7.1}$$

where X a Poisson random variable with mean $\rho(\mathbb{D})$. If $\psi_x(\nu) = 1/\nu(\mathbb{D})$ and if $\rho(\mathbb{D}) < 1$, then

$$Z = e^{-\rho(\mathbb{D})}(1 - \rho(\mathbb{D}))^{-1}$$

and

$$\Psi(\nu) = e^{\rho(\mathbb{D})}(1 - \rho(\mathbb{D}))\nu(\mathbb{D})!$$

Proof. We normalize Ψ such that $E_{\Pi_\rho}[\Psi] = 1$. If $\psi_x(\nu) = 1/\nu(\mathbb{D})$ then

$$Z = \mathbf{E}[X!] = \sum_{n=0}^{\infty} e^{-\rho(\mathbb{D})} \frac{\rho(\mathbb{D})^n}{n!} n! = e^{-\rho(\mathbb{D})}(1 - \rho(\mathbb{D}))^{-1}$$

■

7.1.3 Ergodicity

Particular case: no mobility

Proposition 37 *Consider the model described in Section 7.1.1 without mobility and with* $\psi_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$. *The corresponding SMQ* $\{N_t\}$ *has routing rates satisfying*

$$\lambda(x, \mathbb{D}) = 0, \quad x \in \mathbb{D}$$

and service rates given by

$$r_{xy}(\nu) = \begin{cases} 1 & \text{if } x = o, y \in \mathbb{D} \\ 1/h(\nu(\mathbb{D})) & \text{if } x \in \mathbb{D}, y \in \bar{\mathbb{D}} \end{cases}$$

The total number of users $\{N_t(\mathbb{D})\}$ *may be viewed as a generalized processor-sharing (GPS)* M/GI/1 *queue with arrival rate* $\lambda = \lambda(o, \mathbb{D})$*, with mean volume (i.e. amount of data to transmit or receive)*

$$\mu^{-1} = \frac{1}{\lambda(o, \mathbb{D})} \int_{\mathbb{D}} \frac{\lambda(o, dx)}{\lambda(x, \{o\})}$$

and such that, when there are n *users in the queue, each one is served at rate* $1/h(n)$.

If

$$\lambda\mu^{-1} = \int_{\mathbb{D}} \frac{\lambda(o, dx)}{\lambda(x, \{o\})} < \lim_{n\to\infty} n/h(n)$$

then the state $\emptyset$ *is positive recurrent for the process* $\{N_t\}$ *and the return time of* $\{N_t\}$ *to* $\emptyset$ *from any initial state is almost surely finite.*

Proof. A user arrives at dx with probability $\lambda(o, dx)/\lambda(o, \mathbb{D})$ and has a location-call-volume which is exponentially distributed with mean $1/\lambda(x, \{o\})$. Hence at his arrival, a user has a volume with mean

$$\mu^{-1} = \frac{1}{\lambda(o, \mathbb{D})} \int_{\mathbb{D}} \frac{\lambda(o, dx)}{\lambda(x, \{o\})}$$

The traffic intensity of the equivalent GPS queue is

$$\lambda\mu^{-1} = \rho(\mathbb{D})$$

The last part of the proposition is due to the properties of the GPS queue [78, Proposition 4.1 and Remark 4.1]. ■

Note that the traffic equations (6.47) admit the solution $\rho(dx) = \lambda(o, dx)/\lambda(x, \{o\})$, then

$$\rho(\mathbb{D}) = \lambda\mu^{-1}$$

and the ergodicity condition writes

$$\rho(\mathbb{D}) < \lim_{n\to\infty} n/h(n)$$

Particular case: volume mean independent from location

Proposition 38 *Consider the model described in Section 7.1.1 with location-call-volume rates* $p(x,\{o\})\lambda(x)=\mu$ *where* μ *is a given constant (i.e. independent from location* $x\in\mathbb{D}$*) and with* $\psi_x(\nu)=1/h(\nu(\mathbb{D})), x\in\mathbb{D},\nu\in\mathbb{M}$*. The corresponding SMQ* $\{N_t\}$ *has routing rates satisfying*

$$\lambda(x,\{o\})=\mu,\quad x\in\mathbb{D}$$

and service rates given by

$$r_{xy}(\nu)=\begin{cases}1 & \textit{if } x=o, y\in\mathbb{D}\\ 1/h(\nu(\mathbb{D})) & \textit{if } x\in\mathbb{D}, y\in\bar{\mathbb{D}}\end{cases}$$

The total number of users $\{N_t(\mathbb{D})\}$ *may be viewed as a processor-sharing* M/M/1 *queue with arrival rate* $\lambda=\lambda(o,\mathbb{D})$*, with mean volume* μ^{-1}*, and such that, when there are* n *users in the queue, each one is served at rate* $1/h(n)$*.*

If

$$\lambda\mu^{-1}<\lim_{n\to\infty}n/h(n)$$

then the state $\emptyset$ *is positive recurrent for the process* $\{N_t\}$ *and the return time of* $\{N_t\}$ *to* $\emptyset$ *from any initial state is almost surely finite.*

Proof. A user arrives at dx with probability $\lambda(o,dx)/\lambda(o,\mathbb{D})$ and has a location-call-volume which is exponentially distributed with mean $1/(p(x,\{o\})\lambda(x))=1/\mu$. Hence at his arrival, a user has a volume with mean

$$\int_{\mathbb{D}}\mu^{-1}\frac{\lambda(o,dx)}{\lambda(o,\mathbb{D})}=\mu^{-1}$$

The traffic intensity of the equivalent processor-sharing M/M/1 queue is

$$\lambda\mu^{-1}=\rho(\mathbb{D})$$

Another way to prove the equivalence of our system to a M/M/1 queue is to consider the subsets of the state space $\mathbb{M}$ defined by

$$\Gamma(\nu)=\{\nu'\in M;\nu'(\mathbb{D})=\nu(\mathbb{D})\},\quad \nu\in\mathbb{M}$$

We say that $\Gamma(\nu)$ is the class of $\nu\in\mathbb{M}$ and we show now that the aggregation of the states into these classes gives rise to a Markov process, say $\{\tilde{N}_t\}$. To this end we have to show that

$$\nu'\in\Gamma(\nu)\Rightarrow(q(\nu,\Gamma(\eta))=q(\nu',\Gamma(\eta)),\forall\eta\in\mathbb{M}\backslash\Gamma(\nu))\tag{7.2}$$

in which case the aggregated Markov process $\{\tilde{N}_t\}$ would have the generator $\tilde{q}$ defined on the state space $\tilde{\mathbb{M}}=\{\Gamma(\nu);\nu\in\mathbb{M}\}$ by

$$\tilde{q}(\Gamma(\nu),\Gamma(\eta))=q(\nu,\Gamma(\eta))\tag{7.3}$$

From (6.10) we get

$$q(\nu, \Gamma(\nu) + 1) = \lambda(o, \mathbb{D}), \quad q(\nu, \Gamma(\nu) - 1) = \frac{\nu(\mathbb{D})}{h(\nu(\mathbb{D}))}\mu$$

hence (7.2) is satisfied. Note that the state space $\tilde{\mathbb{M}}$ may be identified with the set of integers $\mathbb{N}$, and from (7.3) we get the aggregated generator

$$\tilde{q}(n, n+1) = \lambda(o, \mathbb{D}), \quad \tilde{q}(n, n-1) = \frac{n}{h(n)}\mu$$

which is the generator of a M/M/1 queue with the properties announced in the Proposition.

The second part of the proposition is due to the properties of the processor-sharing M/M/1 queue; see e.g. [37]. ■

Note that if the traffic equations (6.47) admit a solution $\rho(\cdot)$, then by Equation (6.48),

$$\rho(\mathbb{D}) = \lambda\mu^{-1}$$

and the ergodicity condition writes

$$\rho(\mathbb{D}) < \lim_{n\to\infty} n/h(n)$$

Particular case: the set $\mathbb{D}$ of locations is discrete

Proposition 39 *Consider the model described in Section 7.1.1 with* $\psi_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$. *Assume moreover that the set* $\mathbb{D}$ *of locations is discrete, the SMQ generator* q *is irreducible and that the traffic equations (6.47) have a solution some measure* ρ *on* $\bar{\mathbb{D}}$. *Then the condition*

$$\rho(\mathbb{D}) < \lim_{n\to\infty} n/h(n) \tag{7.4}$$

implies that the SMQ process $\{N_t\}$ *is ergodic.*

Proof. Since the Markov process $\{N_t\}$ has a discrete state space, it is enough to show that it admits a finite invariante measure. To this end, we show that the conditions of Proposition 28 hold true.

First note that (7.4) implies Condition (6.51). It remains to show that the condition (7.4) implies (6.52). Note that the service rates are balanced with respect to

$$\Psi(\nu) = \prod_{n=1}^{\nu(\mathbb{D})} h(n)$$

Since under Π_ρ, the total number of points $\nu(\mathbb{D})$ is a Poisson random variable with mean $\rho(\mathbb{D})$, we deduce that, for each fixed $b \in \mathbb{R}_+$,

$$\mathbf{E}_{\Pi_\rho}\left[b^{\nu(\mathbb{D})}\nu(\mathbb{D})!\right] = e^{-\rho(\mathbb{D})}\frac{1}{1 - b\rho(\mathbb{D})}$$

Denote $a = \lim_{n\to\infty} h(n)/n$ and fix some $\epsilon > 0$ such that $(a-\epsilon)\rho(\mathbb{D}) < 1$ (which is possible since $a\rho(\mathbb{D}) < 1$). There exists some n_ϵ such that, for all $n > n_\epsilon$, $h(n) \le (a-\epsilon)n$. Then, writing $\Psi(\nu) = 1\{\nu(\mathbb{D}) \le n_\epsilon\}\Psi(\nu) + 1\{\nu(\mathbb{D}) > n_\epsilon\}\Psi(\nu)$, we deduce that

$$\begin{aligned}
\mathbf{E}_{\Pi_\rho}[\Psi] &= \mathbf{E}_{\Pi_\rho}[1\{\nu(\mathbb{D}) \le n_\epsilon\}\Psi(\nu)] + \mathbf{E}_{\Pi_\rho}[1\{\nu(\mathbb{D}) > n_\epsilon\}\Psi(\nu)] \\
&\le \mathbf{E}_{\Pi_\rho}[1\{\nu(\mathbb{D}) \le n_\epsilon\}\Psi(\nu)] + \mathbf{E}_{\Pi_\rho}\left[1\{\nu(\mathbb{D}) > n_\epsilon\}\prod_{n=1}^{n_\epsilon} h(n) \prod_{n=n_\epsilon+1}^{\nu(\mathbb{D})} (a-\epsilon)n\right] \\
&\le \sup_{1\le k\le n_\epsilon}\left(\prod_{n=1}^{k} h(n)\right) + \left(\prod_{n=1}^{n_\epsilon} \frac{h(n)}{(a-\epsilon)n}\right)\mathbf{E}_{\Pi_\rho}\left[\prod_{n=1}^{\nu(\mathbb{D})}(a-\epsilon)n\right] \\
&\le \sup_{1\le k\le n_\epsilon}\left(\prod_{n=1}^{k} h(n)\right) + \left(\prod_{n=1}^{n_\epsilon} \frac{h(n)}{(a-\epsilon)n}\right)e^{-\rho(\mathbb{D})}\frac{1}{1-(a-\epsilon)\rho(\mathbb{D})} < \infty
\end{aligned}$$

which shows that Condition (6.52) holds true. ■

General case: Ergodicity condition

An interesting question for future work is whether Condition (7.4) is sufficient for ergodicity when the state space is not discrete.

7.1.4 Performance

Delay

We consider the model described in Section 7.1.1.

We assume that $\emptyset$ is positive recurrent for the SMQ process $\{N_t\}$, and that the return time of $\{N_t\}$ to $\emptyset$ from any initial state is almost surely finite. Then by Corollary 6 $\lim_{t\to\infty} P_t(\nu,\Gamma) = \Pi(\Gamma)$ for all $\nu \in \mathbb{M}, \Gamma \in \mathcal{M}$ where Π is a probability measure given by (6.30). By Proposition 25, there is a unique invariant probability measure which is the limiting distribution Π.

In the present section we extend the development of [25, §.4] to the spatial case. Assume from now that the system is at its stationary regime.

By Proposition 26, the arrival rate of users to some $A \in \mathcal{D}$, which we denote $\Lambda(A)$, equals

$$\begin{aligned}
\Lambda(A) &= \mathbf{E}_\Pi\left[q\left(\nu, T_{\bar{\mathbb{D}}\setminus A,A}\nu\right)\right] \\
&= \mathbf{E}_\Pi\left[\int_A r_{oy}(\nu)\lambda(o,dy) + \int_{(\mathbb{D}\setminus A)\times A} r_{xy}(\nu)\lambda(x,dy)\nu(dx)\right] \\
&= \lambda(o,A) + \int_{\mathbb{D}\setminus A}\lambda(x,A)\mathbf{E}_\Pi[\psi_x(\nu)\nu(dx)]
\end{aligned}$$

By Little's theorem (cf. [17]), the expected delay for users in $A \in \mathcal{D}$, denoted $\bar{T}(A)$, equals

$$\bar{T}(A) = \frac{\mathbf{E}_{\Pi}[\nu(A)]}{\Lambda(A)} = \frac{\mathbf{E}_{\Pi}[\nu(A)]}{\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\,\mathbf{E}_{\Pi}[\psi_x(\nu)\,\nu(dx)]} \tag{7.5}$$

In particular, the delay for users in all the area $\mathbb{D}$ equals

$$\bar{T}(\mathbb{D}) = \frac{\mathbf{E}_{\Pi}[\nu(\mathbb{D})]}{\lambda(o,\mathbb{D})}$$

Proposition 40 *Assume now that Conditions of Proposition 28 hold true, then the limiting distribution Π is a Gibbs distribution with density Ψ with respect to a Poisson weight process with finite measure ρ. In this case, for any $A \in \mathcal{D}$,*

$$\mathbf{E}_{\Pi}\left[\int_A \psi_x(\nu)\,\nu(dx)\right] = \rho(A) \tag{7.6}$$

then

$$\bar{T}(A) = \frac{\mathbf{E}_{\Pi}[\nu(A)]}{\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\,\rho(dx)} \tag{7.7}$$

Proof. (Cf. [106, p?] for the discrete case.) By Proposition 25, there is a unique invariant measure, then Π is a Gibbs distribution. The left hand side of (7.6) may be written in the form

$$\mathbf{E}_{\Pi}\left[\int_{\mathbb{D}} g(\nu - \delta_x, x)\,\nu(dx)\right]$$

where

$$g(\nu, x) = \psi_x(\nu + \delta_x)\,\mathbf{1}\{x \in A\}$$

From Proposition 20 we get

$$\begin{aligned}\mathbf{E}_{\Pi}\left[\int_{\mathbb{D}} g(\nu - \delta_x, x)\,\nu(dx)\right] &= \mathbf{E}_{\Pi}\left[\int_{\mathbb{D}} g(\nu, x)\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)}\rho(dx)\right] \\ &= \mathbf{E}_{\Pi}\left[\int_A \psi_x(\nu+\delta_x)\,\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)}\rho(dx)\right]\end{aligned}$$

But since the service rates ψ_x are Ψ-balanced we have

$$\psi_x(\nu+\delta_x)\,\frac{\Psi(\nu+\delta_x)}{\Psi(\nu)} = 1$$

which proves Equation (7.6).

Using Equations (7.5) and (7.6) we get (7.7). ■

Proposition 41 *Assume that the traffic equations (6.47) have a solution* ρ *which satisfies* $\rho(\mathbb{D}) < 1$*. If* $\psi_x(\nu) = 1/h(\nu(\mathbb{D}))$*, then*

$$\mathbf{E}_\Pi[\nu(A)] = \rho(A)\mathcal{H}(\rho(\mathbb{D})) \tag{7.8}$$

where the function $\mathcal{H}(s)$ *is defined for* $s > 0$ *by*

$$\mathcal{H}(s) = \frac{\mathbf{E}[H(X+1)]}{\mathbf{E}[H(X)]}, \quad H(k) = \prod_{n=1}^{k} h(k) \tag{7.9}$$

where X *is a Poisson random variable with parameter* s*. The delay for users in* $A \in \mathcal{D}$*,*

$$\bar{T}(A) = \frac{\rho(A)}{\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\rho(dx)} \mathcal{H}(\rho(\mathbb{D})) \tag{7.10}$$

In particular

$$\bar{T}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{\lambda(o,\mathbb{D})} \mathcal{H}(\rho(A)) \tag{7.11}$$

More particularly, if $\psi_x(\nu) = 1/\nu(\mathbb{D})$*, then*

$$\mathbf{E}_\Pi[\nu(A)] = \frac{\rho(A)}{1-\rho(\mathbb{D})} \tag{7.12}$$

and the delay for users in $A \in \mathcal{D}$*,*

$$\bar{T}(A) = \frac{\rho(A)}{(1-\rho(\mathbb{D}))\left[\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\rho(dx)\right]} \tag{7.13}$$

In particular

$$\bar{T}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{(1-\rho(\mathbb{D}))\lambda(o,\mathbb{D})} \tag{7.14}$$

Proof. Assume that $\psi_x(\nu) = 1/h(\nu(\mathbb{D}))$. Let $A \in \mathcal{D}$. Applying Proposition 20 with $g(\nu, x) = \mathbf{1}\{x \in A\}$ we obtain

$$\begin{aligned}
\mathbf{E}_\Pi[\nu(A)] &= \mathbf{E}_\Pi\left[\int_A \frac{\Psi(\nu+\epsilon_x)}{\Psi(\nu)}\rho(dx)\right] = \mathbf{E}_\Pi\left[\int_A h(\nu(\mathbb{D})+1)\rho(dx)\right] \\
&= \rho(A)\mathbf{E}_\Pi[h(\nu(\mathbb{D})+1)] \\
&= \rho(A)\mathbf{E}_{\Pi_\rho}[h(\nu(\mathbb{D})+1)\Psi(\nu)] \\
&= \rho(A)\mathcal{H}(\rho(\mathbb{D}))
\end{aligned}$$

where the function $\mathcal{H}(\cdot)$ is given by (7.9). This proves (7.8). From Equations (7.7) and (7.8) we get (7.10). Equation (7.11) is obtained by taking $A = \mathbb{D}$ in (7.10).

In the particular case $\psi_x(\nu) = 1/\nu(\mathbb{D})$ we have

$$\mathbf{E}[H(X+1)] = Z\mathbf{E}[(X+1)!] = \sum_{n=0}^{\infty} e^{-\rho(\mathbb{D})}\frac{\rho(\mathbb{D})^n}{n!}(n+1)! = e^{-\rho(\mathbb{D})}(1-\rho(\mathbb{D}))^{-2}$$

then

$$\mathcal{H}(\rho(\mathbb{D})) = \frac{1}{1-\rho(\mathbb{D})} \tag{7.15}$$

Hence we get (7.12), (7.13) and (7.14). ■

Remark 20 *When $\mathbb{D}$ is discrete our expression (7.8) is coherent with that given in [29, Proposition 3.1]. In deed our $h(n)$ is denoted $n/G(n)$ there, so the expression of the mean number of users in [29, Proposition 3.1] reads*

$$\mathbf{E}_\Pi[\nu(\mathbb{D})] = \frac{\sum_{n=0}^{\infty} n\rho(\mathbb{D})^n \frac{H(n)}{n!}}{\sum_{n=0}^{\infty} \rho(\mathbb{D})^n \frac{H(n)}{n!}} = \rho(\mathbb{D}) \frac{\sum_{n=0}^{\infty} \frac{\rho(\mathbb{D})^n}{n!} H(n+1)}{\sum_{n=0}^{\infty} \frac{\rho(\mathbb{D})^n}{n!} H(n)}$$

Throughput

Recall that the rate allocated to the users at location x when the state of the system is ν equals $\nu_x \psi_x(\nu)$, then, the average rate allocated to users in $A \in \mathcal{D}$ at the steady state is

$$\mathbf{E}_\Pi\left[\int_A \psi_x(\nu)\,\nu(dx)\right]$$

Then the expected throughput for users in $A \in \mathcal{D}$, denoted $\bar{r}(A)$, equals

$$\bar{r}(A) = \frac{\mathbf{E}_\Pi\left[\int_A \psi_x(\nu)\,\nu(dx)\right]}{\mathbf{E}_\Pi[\nu(A)]} \tag{7.16}$$

Proposition 42 *If Conditions of Proposition 28 hold true, then*

$$\bar{r}(A) = \frac{\rho(A)}{\mathbf{E}_\Pi[\nu(A)]} \tag{7.17}$$

If $\psi_x(\nu) = 1/h(\nu(\mathbb{D}))$, we get

$$\bar{r}(A) = \frac{1}{\mathcal{H}(\rho(\mathbb{D}))} \tag{7.18}$$

where the function $\mathcal{H}(\cdot)$ is given by (7.9). In the particular case $\psi_x(\nu) = 1/\nu(\mathbb{D})$, we get

$$\bar{r}(A) = 1-\rho(\mathbb{D}) \tag{7.19}$$

Proof. Equation (7.39) is immediate from (7.16) and (7.6). Equation (7.19) is obtained from (7.12). ■

Example 6 WHITTLE MODEL FOR HSDPA. *We consider the case where $\psi_x(\nu) = 1/(h(\nu(\mathbb{D}))\gamma(x))$ and denote $\{N_t\}$ the corresponding Whittle SMQ process. We consider the SMQ $\{N'_t\}$ where we modify the following rates*

$$\lambda'(x,A) = \lambda(x,A)/\gamma(x), \quad x \in \mathbb{D}, A \in \bar{\mathcal{D}}$$
$$\psi'_x(\nu) = 1/h(\nu(\mathbb{D})), \quad x \in \mathbb{D}, \nu \in \mathbb{M}$$

We denote with prime (′) the parameters which are specific to $\{N'_t\}$. *By Corollary 8, the processes* $\{N'_t\}$ *and* $\{N_t\}$ *are identical in law. Assume that the traffic equations for the routing rates* λ' *admit a solution* ρ', *then it is easy to see that* λ *admit a solution*

$$\rho(dx) = \rho'(dx)/\gamma(x) \tag{7.20}$$

Assume that $\rho'(\mathbb{D}) < \lim_{n\to\infty} n/h(n)$ *which is equivalent to*

$$\int_{\mathbb{D}} \gamma(x)\,\rho(dx) < \lim_{n\to\infty} n/h(n)$$

which assures the ergodicity of $\{N'_t\}$ *and thus also of* $\{N_t\}$. *Let* Π *be their common limiting distribution. From (7.8) we get*

$$\mathbf{E}_\Pi[\nu(A)] = \rho'(A)\,\mathcal{H}(\rho'(\mathbb{D}))$$

where the function $\mathcal{H}(\cdot)$ *is given by (7.9). From (7.10) we get*

$$\begin{aligned}\bar{T}(A) &= \frac{\rho'(A)}{\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda'(x,A)\,\rho'(dx)}\mathcal{H}(\rho'(\mathbb{D})) \\ &= \frac{\rho'(A)}{\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\,\rho(dx)}\mathcal{H}(\rho'(\mathbb{D}))\end{aligned}$$

From (7.6) we get

$$\mathbf{E}_\Pi\left[\int_A \psi_x(\nu)\,\nu(dx)\right] = \mathbf{E}_\Pi\left[\int_A \gamma(x)^{-1}\,\psi'_x(\nu)\,\nu(dx)\right] = \int_A \gamma(x)^{-1}\,\rho'(dx) = \rho(A)$$

and from (7.16) we get

$$\bar{r}(A) = \frac{\mathbf{E}_\Pi\left[\int_A \psi_x(\nu)\,\nu(dx)\right]}{\mathbf{E}_\Pi[\nu(A)]} = \frac{\mathbf{E}_\Pi\left[\int_A \psi_x(\nu)\,\nu(dx)\right]}{\mathbf{E}_\Pi[\nu(A)]} = \frac{\rho(A)}{\rho'(A)}\frac{1}{\mathcal{H}(\rho'(\mathbb{D}))}$$

Example 7 Whittle model for the shared channel in CDMA. *This is a particular case of Example 6 where* $h(n) = n$. *In this case, from (7.15) we get*

$$\mathcal{H}(\rho'(\mathbb{D})) = \frac{1}{1-\rho'(\mathbb{D})}$$

then

$$\mathbf{E}_\Pi[\nu(A)] = \frac{\rho'(A)}{1-\rho'(\mathbb{D})}$$

$$\bar{T}(A) = \frac{\rho'(A)}{(1-\rho'(\mathbb{D}))\left[\lambda(o,A) + \int_{\mathbb{D}\setminus A} \lambda(x,A)\,\rho(dx)\right]} \tag{7.21}$$

$$\bar{r}(A) = \frac{\rho(A)}{\rho'(A)}[1-\rho'(\mathbb{D})] \tag{7.22}$$

7.2 Wireless model

7.2.1 Model

We have sometimes to distinguish the motion dynamics from the service dynamics. This is particularly useful for wireless networks.

We describe first the service dynamics:

(a3) Exogenous arrivals come to dx as a Poisson process with intensity $\lambda(o, dx)$.

(b3) Define the CALL-VOLUME at a location x as the residual amount of data to transmit or receive at location x before the exit from the system. The call-volumes are assumed i.i.d. exponentially distributed with parameters $\lambda(x, \{o\})$ (or equivalently with mean $1/\lambda(x, \{o\})$) and independent from arrivals. (Note that the call-volume equals the service duration if the user is served at rate 1.)

(c3) The state of the system is $\nu \in \mathbb{M}$ counting the users at each location. The rate allocated to the users at location x when the state of the system is ν equals $\nu_x g_x(\nu)$ (i.e. the rate allocated for each user at location x is $g_x(\nu)$).

We describe now the motion dynamics:

(d3) The sojourn duration of users at location x are i.i.d. random variables which are exponentially distributed with rate $\lambda'(x)$ and independent of arrivals, call-volumes and the service rates.

(e3) A user finishing its sojourn at location x, is routed to dy according to the probability kernel $p'(x, dy)$ where $p'(x, \mathbb{D}) = 1$.

An example of the rates $\lambda'(x)$ and the kernel $p'(x, dy)$ is given in Section 6.5.

Proposition 43 *This physical interpretation (a3)–(e3) corresponds to the generator (6.8) if*

$$\begin{aligned} \lambda'(x) &= \lambda(x, \mathbb{D}), \quad x \in \mathbb{D} \\ p'(x, A) &= \lambda(x, A)/\lambda'(x), \quad x \in \mathbb{D}, A \in \mathcal{D} \end{aligned} \tag{7.23}$$

and

$$r_{xy}(\nu) = \begin{cases} 1 & \text{if } x \in \bar{\mathbb{D}}, y \in \mathbb{D} \\ g_x(\nu) & \text{if } x \in \mathbb{D}, y = o \end{cases} \tag{7.24}$$

Proof. Suppose that the state of the system at some instant t is some $\nu \in \mathbb{M}$. The next event may be a birth (which corresponds to transitions $\nu \to T_{ox}\nu$ for some $x \in \mathbb{D}$), a death (which corresponds to transitions $\nu \to T_{xo}\nu$ for some $x \in \nu$) or a motion (which corresponds to transitions $\nu \to T_{xy}\nu$ for some $x \in \nu, y \in \mathbb{D}$).

By (a3)

$$q(\nu, T_{oA}\nu) = \lambda(o, A), \quad A \in \mathcal{D}$$

At a given location x there are ν_x users, each of which, by (b3) and (c3), will exit the system after an exponentially-distributed duration of rate $\lambda(x, \{o\}) g_x(\nu)$. The next departure will take place at the minimum of these durations which is exponentially distributed with rate $\lambda(x, \{o\}) \nu_x g_x(\nu)$, hence

$$q(\nu, T_{xo}\nu) = \lambda(x, \{o\}) \nu_x g_x(\nu)$$

By (d3)-(e3), for each user at location x, the next transition from this location to $A \in \mathcal{D}$ will, take place after an exponentially-distributed duration of rate $\lambda'(x) p'(x, A)$. The next transition of a user from x to A will take place at the minimum of these durations which is exponentially distributed with rate $\lambda'(x) p'(x, A) \nu_x$, hence

$$q(\nu, T_{xA}\nu) = \lambda'(x) p'(x, A) \nu_x, \quad x \in \mathbb{D}, A \in \mathcal{D}$$

If we have the conditions of the Proposition we retrieve the generator (6.8). ■

Remark 21 *Note that we get a SMQ which is not Whittle. So the model for wireless networks serving elastic traffic described in the present section is different from the classical Whittle model described in Section 7.1.1 classically used for wired networks and sometimes used for wireless networks [25, §.2]. The fundamental difference between the two models is the following. At arrival epoch, a user requires some volume to transmit during* the whole call *in the wireless model; wheras he requires some volume to transmit* at each location *he visits in the Whittle model. Consequently in the wireless model, the motion is independent from transmission whereas in the Whittle model, the motion is related to the bit-rate allocated by the network. Hence when the system approaches congestion (instability), the users motion is frozen in the Whittle model whereas they continue their motion independently from the system congestion in the wireless model.*

7.2.2 Service rate balance

The following lemmas will be useful in the study of the ergodicity and balance of the service rates for the wireless model.

Lemma 11 *The service rates (7.24) are balanced iff $g_x(\nu)$ is a (measurable) function of $\nu(\mathbb{D})$, say $g_x(\nu) = 1/h(\nu(\mathbb{D}))$. In such case the results of Lemma 10 hold true.*

Proof. Suppose that the service rates (7.24) are balanced. Then there exists some positive function $\Psi : \mathbb{M} \to (0, \infty)$ such that

$$\Psi(\nu) r_{xy}(\nu) = \Psi(T_{xy}\nu) r_{yx}(T_{xy}\nu), \quad x \neq y \in \bar{\mathbb{D}}, \nu \in \mathbb{M}, \nu_x > 0$$

which is equivalent to

$$\Psi(\nu) g_x(\nu) = \Psi(T_{xo}\nu), \quad \text{when } y = o \tag{7.25}$$
$$\Psi(\nu) = \Psi(T_{oy}\nu) g_y(T_{oy}\nu) \quad \text{when } y \neq o, x = o \tag{7.26}$$
$$\Psi(\nu) = \Psi(T_{xy}\nu) \quad \text{when } y \neq o, x \neq o \tag{7.27}$$

From (7.27) we deduce that $\Psi(\nu)$ depends only on $\nu(\mathbb{D})$, and not in the particular positions of the points of ν. Equation (7.25) shows then that $g_x(\nu)$ is only a function of $\nu(\mathbb{D})$.

Inversely, suppose now that $g_x(\nu)$ is only a function of $\nu(\mathbb{D})$, say $1/h(\nu(\mathbb{D}))$. Consider

$$\Psi(\nu) = \prod_{n=1}^{\nu(\mathbb{D})} h(n)$$

It is clear that Equation (7.25), (7.26) and (7.27) are satisfied. Hence The service rates (7.24) are balanced. ■

Lemma 12 *Let $\gamma : \mathbb{D} \to \mathbb{R}_+^*$ be some measurable function. If, for all $x \in \mathbb{D}$, we replace $\lambda(x, \{o\})$ by $\lambda(x, \{o\}) / \gamma(x)$ and $r_{xo}(\nu)$ by $r_{xo}(\nu)\gamma(x)$ the generator (6.8) remains invariant.*

Proof. The product $\lambda(x, \{o\}) r_{xo}(\nu)$ remains invariant, then the generator (6.8) remains invariant. ■

Corollary 9 *Consider service rates in the form (7.24) where, for all $x \in \mathbb{D}, \nu \in \mathbb{M}$, the service rates $g_x(\nu)$ is in the form $1/(h(\nu(\mathbb{D}))\gamma(x))$ for two positive measurable functions γ and h. By replacing $\lambda(x, \{o\})$ by $\lambda(x, \{o\}) / \gamma(x)$ and $g_x(\nu)$ by $1/h(\nu(\mathbb{D}))$ (for all $x \in \mathbb{D}, \nu \in \mathbb{M}$) the generator (6.8) remains invariant, and the transformed service rates are balanced.*

Proof. Immediate consequences of Lemma 12 and Lemma 11. ■

In the rest of the present section we consider the context deduced from Corollary 9, where, if the state of the system is ν, then the service rate of each user is $1/h(\nu(\mathbb{D}))$, i.e. depends only on the number of users in the system and no on their particular positions (we assume that $\lambda(x, \{o\})$ has been correctly modified as indicated in Corollary 9).

7.2.3 Ergodicity

Particular case: no mobility

The model described in Section 7.2.1 without mobility (i.e. $\lambda'(x) = 0$, for all $x \in \mathbb{D}$) and with service rates $g_x(\nu)$ is equivalent to the model described in Section 7.1.1 without mobility and with $\psi_x(\nu) = g_x(\nu), x \in \mathbb{D}, \nu \in \mathbb{M}$. Thus, if $g_x(\nu) = 1/h(\nu(\mathbb{D}))$, then the results of Proposition 37 hold true.

Particular case: call-volume mean independent from location

Proposition 44 *Consider the model described in Section 7.2.1 with call-volume mean independent from location and with* $g_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$. *The corresponding SMQ* $\{N_t\}$ *has routing rates satisfying*

$$\lambda(x, \{o\}) = \mu, \quad x \in \mathbb{D}$$

where μ *is a given constant, and service rates given by*

$$r_{xy}(\nu) = \begin{cases} 1 & \text{if } x = \bar{\mathbb{D}}, y \in \mathbb{D} \\ 1/h(\nu(\mathbb{D})) & \text{if } x \in \mathbb{D}, y = o \end{cases}$$

The total number of users $\{N_t(\mathbb{D})\}$ *may be viewed as a processor-sharing* M/M/1 *queue with arrival rate* $\lambda = \lambda(o, \mathbb{D})$, *with mean call-volume* μ^{-1}, *and such that, when there are* n *users in the queue, each one is served at rate* $1/h(n)$.

If

$$\lambda\mu^{-1} < \lim_{n\to\infty} n/h(n)$$

then the state $\emptyset$ *is positive recurrent for the process* $\{N_t\}$ *and the return time of* $\{N_t\}$ *to* $\emptyset$ *from any initial state is almost surely finite.*

Proof. A user arrives at dx with probability $\lambda(o, dx)/\lambda(o, \mathbb{D})$ and has a call-volume which is exponentially distributed with mean $1/\lambda(x, \{o\}) = 1/\mu$. The rest of the proof is similar to that of Proposition 38. ■

Infinite mobility case

Consider the model described in Section 7.2.1 with service rates $g_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$ and where the motion rates $\{\lambda(x, A), x \in \mathbb{D}, A \in \mathcal{D}\}$ are given in Section 6.5.

In order to simplify the study the process at hand, we may aggregate its states, by grouping the states with the same number of users, i.e. let

$$\mathbb{M}_n = \{\nu \in \mathbb{M}; \nu(\mathbb{D}) = n\}, \quad n = 0, 1, 2, \ldots$$

The aggregated process has state space $\mathbb{N} = \{0, 1, 2, \ldots\}$. Generally, the aggregated process is not necessarily Markov, but we will see that the limit process when the user's average speed υ goes to infinity is Markovian.

To this end note that the Markov generator at hand may be decomposed in two terms

$$q = \upsilon q^{(0)} + q^{(1)}$$

where the generator $\upsilon q^{(0)}$ corresponds to the motion

$$q^{(0)}(\nu, \Gamma) = \int_{\mathbb{D}\times\mathbb{D}} 1_\Gamma(T_{xy}\nu) \frac{\lambda(x, dy)}{\upsilon} \nu(dx)$$

and the generator $q^{(1)}$ corresponds to the births and deaths

$$q^{(1)}(\nu,\Gamma)=\int_{\mathbb{D}}1_{\Gamma}(T_{oy}\nu)\,\lambda(o,dy)+\frac{1}{h(\nu(\mathbb{D}))}\int_{\mathbb{D}}1_{\Gamma}(T_{xo}\nu)\,\lambda(x,\{o\})\,\nu(dx)$$

If υ is very large, then the transitions due to motion are much more frequent then the births and deaths. A rich literature treated the limiting behavior as υ goes to ∞ when the state space is either finite (cf. for example [43]) or countable (cf. for example [5]). If $\mathbb{D}$ is countable (which is the case if we divide the cell into a countable set of rings), then our sate space $\mathbb{M}$ is countable. In this case, previous works (cf. for example [5]) shows that the limiting behavior as υ goes to ∞ may be studied in terms of the so called aggregated generator defined on $\mathbb{N}=\{0,1,2,\ldots\}$ by

$$\bar{q}(n,m)=\mathbf{E}_{\Pi_n}\left[q^{(1)}(\nu,\mathbb{M}_m)\right]$$

where Π_n designates the invariant probability measure of $q^{(0)}$ on class $\mathbb{M}_n$.

Lemma 13 *Assume that the motion generator $\{\lambda(x,A),x\in\mathbb{D},A\in\mathcal{D}\}$ is positive recurrent. Let Π_n be the invariant probability measure of $q^{(0)}$ on class $\mathbb{M}_n$. Then*

$$\mathbf{E}_{\Pi_n}\left[q^{(1)}(\nu,\mathbb{M}_m)\right]=1_{m=n+1}\lambda(o,\mathbb{D})+1_{m=n-1}\frac{n}{h(n)}\int_{\mathbb{D}}\lambda(x,\{o\})\,\varrho(dx)$$

where ϱ is the solution of the traffic equations associated to the motion generator $\{\lambda(x,A),x\in\mathbb{D},A\in\mathcal{D}\}$.

Proof. We should first calculate the invariant probability measures Π_n of $q^{(0)}$ on each class $\mathbb{M}_n$. It is not difficult to see that Π_n is the distribution of n i.i.d. points on $\mathbb{D}$, each one having distribution ϱ. That is,

$$\Pi_n(d\nu)=n!\prod_{x\in\text{supp}(\nu)}\frac{\varrho_x^{\nu_x}}{\nu_x!}d\nu$$

or equivalently, for each measurable function $f:\mathbb{M}_n\to\mathbb{R}_+$,

$$\mathbf{E}_{\Pi_n}[f]=\int_{\mathbb{M}_n}f\left(\sum_{j=1}^{n}\delta_{x_j}\right)\varrho(dx_1)\ldots\varrho(dx_n)$$

in particular

$$\mathbf{E}_{\Pi_n}[\nu(A)]=n\varrho(A),\quad A\in\mathcal{D}$$

From the expression of the generator $q^{(1)}$, we get

$$\mathbf{E}_{\Pi_n}\left[q^{(1)}(\nu,\mathbb{M}_m)\right]=1_{m=n+1}\lambda(o,\mathbb{D})+1_{m=n-1}\frac{n}{h(n)}\int_{\mathbb{D}}\lambda(x,\{o\})\,\varrho(dx)$$

■

The above analysis justifies the model for infinite mobility described in the following proposition.

Proposition 45 *Consider the model described in Section 7.2.1 with* $g_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$. *Suppose that the mobility of users is so fast that we can reasonably assume that during the periods of time when the number of users* $\nu(\mathbb{D})$ *is constant, the call-volume rate* $\lambda(x, \{o\})$ *is averaged over mobility; i.e., assume that the call-volume rate is*

$$\int_{\mathbb{D}} \lambda(x, \{o\}) \varrho(dx) \tag{7.28}$$

where ϱ *is the solution of the traffic equations associated to the motion generator* $\{\lambda(x, A), x \in \mathbb{D}, A \in \mathcal{D}\}$.

The total number of users $\{N_t(\mathbb{D})\}$ *may be viewed as a* M/GI/1 *processor-sharing queue with arrival rate* $\lambda = \lambda(o, \mathbb{D})$, *with mean call-volume (7.28), and such that, when there are* n *users in the queue, each one is served at rate* $1/h(n)$.

If

$$\frac{\lambda(o, \mathbb{D})}{\int_{\mathbb{D}} \lambda(x, \{o\}) \varrho(dx)} < \lim_{n\to\infty} n/h(n) \tag{7.29}$$

then the state $\emptyset$ *is positive recurrent for the process* $\{N_t\}$ *and the return time of* $\{N_t\}$ *to* $\emptyset$ *from any initial state is almost surely finite.*

Proof. The mean call-volume (7.28) is clearly independent from location, then the result follows from Proposition 44. ■

General case

An interesting subject for future research is to determine the ergodicity condition for the model described in Section 7.2.1 with $g_x(\nu) = 1/h(\nu(\mathbb{D})), x \in \mathbb{D}, \nu \in \mathbb{M}$. TO COMPLETE.

7.2.4 Performance

Delay

We assume that $\emptyset$ is positive recurrent for the SMQ process $\{N_t\}$, and that the return time of $\{N_t\}$ to $\emptyset$ from any initial state is almost surely finite. Then by Corollary 6 $\lim_{t\to\infty} P_t(\nu, \Gamma) = \Pi(\Gamma)$ for all $\nu \in \mathbb{M}, \Gamma \in \mathcal{M}$ where Π is a probability measure given by (6.30). By Proposition 25, there is a unique invariant probability measure which is the limiting distribution Π.

Now assume that the system is at its stationary regime.

The arrival rate of users to some $A \in \mathcal{D}$, which we denote $\Lambda(A)$, equals

$$\begin{aligned}
\Lambda(A) &= \mathbf{E}_\Pi \left[q\left(\nu, T_{\bar{\mathbb{D}}\setminus A, A}\nu\right)\right] \\
&= \mathbf{E}_\Pi \left[\int_A r_{oy}(\nu) \lambda(o, dy) + \int_{(\mathbb{D}\setminus A)\times A} r_{xy}(\nu) \lambda(x, dy) \nu(dx)\right] \\
&= \lambda(o, A) + \int_{\mathbb{D}\setminus A} \lambda(x, A) \mathbf{E}_\Pi[\nu(dx)]
\end{aligned}$$

By Little's theorem (cf. [17]), the expected delay for users in $A \in \mathcal{D}$, denoted $\bar{T}(A)$, equals

$$\bar{T}(A) = \frac{\mathbf{E}_\Pi[\nu(A)]}{\Lambda(A)} = \frac{\mathbf{E}_\Pi[\nu(A)]}{\lambda(o,A) + \int_{\mathbb{D}\setminus A}\lambda(x,A)\,\mathbf{E}_\Pi[\nu(dx)]} \tag{7.30}$$

In particular, the delay for users in all the area $\mathbb{D}$ equals

$$\bar{T}(\mathbb{D}) = \frac{\mathbf{E}_\Pi[\nu(\mathbb{D})]}{\lambda(o,\mathbb{D})}$$

Proposition 46 *Assume that the traffic equations (6.47) have a solution ρ which satisfies $\rho(\mathbb{D}) < 1$ and that $g_x(\nu) = 1/h(\nu(\mathbb{D}))$ and let*

$$\Psi(\nu) = Z^{-1}\prod_{n=1}^{\nu(\mathbb{D})} h(n)$$

where Z is a normalizing constant (see Lemma 10). Assume moreover that the Gibbs distribution with density Ψ with respect to a Poisson weight process with finite measure ρ is invariant (which is the case when for example the routing kernel λ is reversible). Then the limiting distribution Π is this Gibbs distribution and

$$\mathbf{E}_\Pi[\nu(A)] = \rho(A)\,\mathcal{H}(\rho(\mathbb{D})) \tag{7.31}$$

where the function $\mathcal{H}(\cdot)$ is given by (7.9). The delay for users in $A \in \mathcal{D}$, is given by

$$\bar{T}(A) = \frac{\rho(A)}{\frac{1}{\mathcal{H}(\rho(\mathbb{D}))}\lambda(o,A) + \int_{\mathbb{D}\setminus A}\lambda(x,A)\,\rho(dx)} \tag{7.32}$$

In particular

$$\bar{T}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{\lambda(o,\mathbb{D})}\mathcal{H}(\rho(\mathbb{D})) \tag{7.33}$$

More particularly, if $\psi_x(\nu) = 1/\nu(\mathbb{D})$, then

$$\mathbf{E}_\Pi[\nu(A)] = \frac{\rho(A)}{1-\rho(\mathbb{D})} \tag{7.34}$$

and the delay for users in $A \in \mathcal{D}$,

$$\bar{T}(A) = \frac{\rho(A)}{(1-\rho(\mathbb{D}))\,\lambda(o,A) + \int_{\mathbb{D}\setminus A}\lambda(x,A)\,\rho(dx)} \tag{7.35}$$

In particular

$$\bar{T}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{(1-\rho(\mathbb{D}))\,\lambda(o,\mathbb{D})} \tag{7.36}$$

Proof. The proof of (7.31) is analogous to that of (7.8). We deduce (7.32) from (7.30) and (7.31).

In the particular case $\psi_x(\nu) = 1/\nu(\mathbb{D})$, the function $\mathcal{H}(\cdot)$ is given by (7.15), hence we get (7.34), (7.35) and (7.36). ■

Throughput

Recall that the rate allocated to the users at location x when the state of the system is ν equals $\nu_x g_x(\nu)$, then, the average rate allocated to users in $A \in \mathcal{D}$ at the steady state is

$$\mathbf{E}_\Pi\left[\int_A g_x(\nu)\,\nu(dx)\right]$$

Then the expected throughput for users in $A \in \mathcal{D}$, denoted $\bar{r}(A)$, equals

$$\bar{r}(A) = \frac{\mathbf{E}_\Pi\left[\int_A g_x(\nu)\,\nu(dx)\right]}{\mathbf{E}_\Pi[\nu(A)]} \tag{7.37}$$

Example 8 *If for the wireless model the stationary distribution* Π *is a Gibbs distribution with density* Ψ *with respect to a Poisson weight process with finite measure* ρ*, then, for any* $A \in \mathcal{D}$*,*

$$\mathbf{E}_\Pi\left[\int_A g_x(\nu)\,\nu(dx)\right] = \rho(A) \tag{7.38}$$

and

$$\bar{r}(A) = \frac{\rho(A)}{\mathbf{E}_\Pi[\nu(A)]} \tag{7.39}$$

If $\psi_x(\nu) = 1/h(\nu(\mathbb{D}))$*, then*

$$\bar{r}(A) = \frac{1}{\mathcal{H}(\rho(\mathbb{D}))}$$

where the function $\mathcal{H}(\cdot)$ *is given by (7.9). In the particular case* $g_x(\nu) = 1/\nu(\mathbb{D})$*,*

$$\bar{r}(A) = 1 - \rho(\mathbb{D}) \tag{7.40}$$

Proof. The proof of (7.38) is similar to the proof of Equation (7.6). Equation (7.39) is then obtained from (7.37). Equation (7.40) is finally obtained from (7.34). ■

Example 9 Wireless model for HSDPA. *We consider the case where* $g_x(\nu) = 1/(h(\nu(\mathbb{D}))\gamma(x))$ *and denote* $\{N_t\}$ *the corresponding SMQ process as described in Proposition 43. We consider the SMQ* $\{N'_t\}$ *where we modify the following rates*

$$\lambda'(x,\{o\}) = \lambda(x,\{o\})/\gamma(x), \quad x \in \mathbb{D}$$
$$g'_x(\nu) = 1/h(\nu(\mathbb{D})), \quad x \in \mathbb{D}, \nu \in \mathbb{M}$$

We denote with prime (′) the parameters which are specific to $\{N'_t\}$*. By Corollary 9, the processes* $\{N'_t\}$ *and* $\{N_t\}$ *are identical in law. Assume that the traffic equations for the routing rates* λ' *admit a solution* ρ'*. Assume that* $\{N'_t\}$ *is ergodic (which is the case if* $\rho'(\mathbb{D}) < \lim_{n\to\infty} n/h(n)$ *either in the case without* *mobility by Proposition 37 or in the case where the call-volume mean is independent from location by Proposition 44) and thus* $\{N_t\}$ *is also ergodic. Let* Π *be their common limiting distribution. Assume moreover that the Gibbs distribution with density* Ψ' *(which balances* $g'_x(\nu) = 1/h(\nu(\mathbb{D}))$*) with respect to a Poisson weight process with finite measure* ρ' *is invariant for* $\{N'_t\}$ *(which is the case when for example the routing kernel* λ' *is reversible). Then the limiting distribution* Π *is this Gibbs distribution. From (7.31) we get*

$$\mathbf{E}_\Pi[\nu(A)] = \rho'(A)\,\mathcal{H}(\rho'(\mathbb{D}))$$

where the function $\mathcal{H}(\cdot)$ *is given by (7.9). From (7.32) we get*

$$\bar{T}(A) = \frac{\rho'(A)}{\frac{\lambda(o,A)}{\mathcal{H}(\rho'(\mathbb{D}))} + \int_{\mathbb{D}\setminus A} \lambda(x,A)\,\rho'(dx)}$$

From (7.38) we get

$$\mathbf{E}_\Pi\left[\int_A g_x(\nu)\,\nu(dx)\right] = \mathbf{E}_\Pi\left[\int_A \gamma(x)^{-1} g'_x(\nu)\,\nu(dx)\right] = \int_A \gamma(x)^{-1}\rho'(dx) = \rho(A)$$

and from (7.37) we get

$$\bar{r}(A) = \frac{\mathbf{E}_\Pi\left[\int_A g_x(\nu)\,\nu(dx)\right]}{\mathbf{E}_\Pi[\nu(A)]} = \frac{\rho(A)}{\rho'(A)}\frac{1}{\mathcal{H}(\rho'(\mathbb{D}))}$$

Example 10 Wireless model for the shared channel in CDMA. *This is a particular case of Example 9 where* $h(n) = n$*. In this case, from (7.15) we get*

$$\mathcal{H}(\rho'(\mathbb{D})) = \frac{1}{1-\rho'(\mathbb{D})}$$

then

$$\mathbf{E}_\Pi[\nu(A)] = \frac{\rho'(A)}{1-\rho'(\mathbb{D})} \tag{7.41}$$

$$\bar{T}(A) = \frac{\rho'(A)}{(1-\rho'(\mathbb{D}))\,\lambda(o,A) + \int_{\mathbb{D}\setminus A}\lambda'(x,A)\,\rho'(dx)} \tag{7.42}$$

$$\bar{r}(A) = \frac{\rho(A)}{\rho'(A)}\left[1-\rho'(\mathbb{D})\right] \tag{7.43}$$

Remark 22 *It is surprising that even if the Whittle model and our wireless model are different mathematically and physically, the delay and throughput expressions for the two models (in the particular case where* $\psi_x(\nu) = g_x(\nu) = 1/\nu(\mathbb{D})$*) are the same. In fact we should notice that the solution* $\rho(\cdot)$ *of the traffic equations accounts for users speed in our wireless model whereas it is independent of the users speed in the Whittle model (in fact the Whittle model doesn't account for the users speed; see Remark 21).*

In the case of Whittle model for the shared channel in CDMA (Example 7) we have the relation (7.20), while this relation doesn't hold for the wireless model (Example 10) except for the particular cases when there is no motion, i.e. $\lambda(x,\mathbb{D}) = 0, \forall x \in \mathbb{D}$*, or when* $\gamma(x) = 1$.

Chapter 8

SMQ for streaming traffic

We consider here streaming traffic (i.e. calls requiring to be served for a given duration).

Consider a M/GI/∞ queue where arrivals have intensity λ and call duration has mean $1/\mu$. As recalled in Lemma 24, the stationary distribution, say Π, of the number of calls in progress of such a FREE (i.e. unconstrained) system is a Poisson with mean λ/μ. Assume now that the number of servers is some positive integer C. We need to describe an *admission control* policy which specifies how the model performs when the C limit is reached. Suppose that we simply drop the calls that arrive when there are C calls already in progress. We get the M/GI/C loss queue whose properties are recalled in Lemma 25. In particular, the *blocking probability* b is given by the Erlang's formula (14.3) which may be written as follows

$$b = \frac{\Pi\{C\}}{\Pi(\mathbb{M}^{\mathrm{f}})}$$

where $\mathbb{M}^{\mathrm{f}} = \{0, 1, \ldots, C\}$ is the state space. Erlang's formula shows that in spite of the differences between the dynamics of the free and that of the loss system, the blocking probability can be expressed in terms of the stationary distribution of calls in progress of the free system. This is the surprising fact that we will attempt to extend to the spatial case.

As well known, Erlang published this formula in 1917 [44], and since that time, the statistical equilibria of much more complicated loss networks have been found to coincide with the truncation of the stationary distribution of some free system to some polytope[1]. This lead to the calculation of the associated blocking probabilities in explicit form for large classes of networks. For an exhaustive survey on loss systems, see [75]. Classical loss models are well adapted to wired communication networks, where the spatial component of the model is typically represented by some *graph of links*, and where the coexistence of calls on a common link is modeled by the occupancy of a *discrete number of circuits* available on this link. In wireless communication, one needs to take into account

[1] A POLYTOPE is a generalization to $\mathbb{R}^n$ ($n \in \mathbb{N}$) of the notion of polygon in $\mathbb{R}^2$.

the spatial characteristics of the network in a more thorough way because *it is the relative location of the radio channels which determines their joint feasibility.* This is especially important for CDMA and other so called *interference limited systems*. One of the additional difficulties then stems from the fact that the spatial component of the model is subject to changes due to the mobility of users and instantaneous changes of radio conditions.

The classical results for *discrete* multiserver queues are recalled in §14.A.2. In particular analytical formulae for the blocking probability for a M/GI multi-class loss queue are recalled. The aim of the present chapter is to extend these formulae to the case where the set of classes is *continuous*, and where there is *mobility* between the different classes. In particular we establish in this setting a relation between the *blocking probability* in the loss system, and the stationary distribution of calls in progress of the free system. This relation may be seen as a *spatial* extension of the *Erlang's formula.*

8.1 Whittle and wireless models

In Section 6.2.3, we give a mathematical description (a1)-(c1) of a Whittle SMQ. We aim here to give a physical description of networks serving streaming traffic.

8.1.1 Whittle model

The following model is classical for modelling wired networks serving streaming traffic ([98] in the discrete case, [106]):

(a2) Exogenous arrivals come to dx as a Poisson process with intensity $\lambda(o, dx)$.

(b2) A user finishing its service at location x is routed to dy according to the probability Kernel $p(x, dy)$ (Bernoulli routing) where $p\left(x, \bar{\mathbb{D}}\right) = 1$. As usual $y = o$ corresponds to the case where the user exits the system.

(c2) Define the LOCATION-CALL-DURATION at a location x as the residual duration at location x before a motion to another location or the exit from the system. The location-call-durations are assumed i.i.d. exponentially distributed with parameters $\lambda(x)$ (or equivalently with mean $1/\lambda(x)$) and independent from arrivals.

(d2) At a given time t, some users are being served and others are waiting. The state of the system is $\nu \in \mathbb{M}$ counting all the users (being served or waiting) at each location. The number of active servers at location x when the state of the network is ν equals $\nu_x \psi_x(\nu) \leq \nu_x$ (since there are at most ν_x streaming messages served simultaneously). The $\nu_x - \nu_x \psi_x(\nu)$ users which are not being served are placed in a waiting room at note x which is assumed infinitely large.

Proposition 47 *The physical interpretation (a2)–(d2) above corresponds to a Whittle SMQ if*

$$\begin{cases} \lambda(x) = \lambda\left(x, \bar{\mathbb{D}}\right), & x \in \mathbb{D} \\ p(x, A) = \lambda(x, A) / \lambda\left(x, \bar{\mathbb{D}}\right), & x \in \mathbb{D}, A \in \bar{\mathcal{D}} \end{cases}$$

and

$$r_{xy}(\nu) = \begin{cases} 1, & x = o, y \in \bar{\mathbb{D}} \\ \psi_x(\nu), & x \in \mathbb{D}, y \in \bar{\mathbb{D}} \end{cases}$$

Proof. Clearly (a2) is the same as (a1) and (b2) is equivalent to (b1) if

$$p(x, A) = \lambda(x, A) / \lambda\left(x, \bar{\mathbb{D}}\right)$$

It remains to show that (c2)-(d2) imply (c1). Suppose that the state of the system at some instant t is some $\nu \in \mathbb{M}$. Recall that the minimum of independent exponential random variables is an exponential random with rate equal to the sum of the rates. Suppose that the network state at some instant is ν. At a given location x, by (d2), there are $\nu_x \psi_x(\nu)$ active users (the term active means that the user is served by the network) each of which, by (c2), will depart from location x after an exponentially-distributed duration of rate $\lambda(x)$. The next departure from location x will then take place at the minimum of these durations which is exponentially distributed with rate $\lambda(x)\nu_x\psi_x(\nu)$, hence we obtain (c1) if

$$\lambda(x) = \lambda\left(x, \bar{\mathbb{D}}\right)$$

Note that since the exogenous arrival rate equals $q(\nu, T_{o\bar{\mathbb{D}}}\nu) = \lambda\left(o, \bar{\mathbb{D}}\right)$ all the arriving users are admitted to the system, but there are not necessarily served immediately. The $\nu_x - \nu_x\psi_x(\nu)$ users which are not being served are placed in a waiting room with has an infinite capacity as assumed in (d2). ■

Hence the departure rate $\lambda(x) \times \nu_x\psi_x(\nu)$ comprises two factors. The first one $\lambda(x)$ may be interpreted as the rate of an exponentially-distributed *service duration at location x* (which is related to the motion within the network and the exit from the network) of a given user and the second one $\nu_x\psi_x(\nu)$ may be interpreted as the number of active servers at location x when the state of the network is ν (called capacity by some authors [98]).

The following examples illustrate the decomposition of the departure rate.

Example 11 *Consider the case where users are motionless.*

- *If there is an infinite number of servers at each location, then the departure rate from location x is $\lambda(x, \{o\}) \times \nu_x$ where the first term $\lambda(x, \{o\})$ is the rate of the exponentially-distributed time until the exist from the network (call duration or total service time has exponential distribution with mean $\lambda(x, \{o\})^{-1}$); and the second term ν_x is the number of active servers at location x when the state of the network is ν. In the particular case where there is a single location, i.e. $|\mathbb{D}| = 1$, we get the* M/M/∞ *queue.*

- *Suppose now that there is a single server at each location, then the departure rate from location x is $\lambda(x, \{o\}) \times 1\{\nu_x > 0\}$ where the first term $\lambda(x, \{o\})$ has the same interpretation as previously and since there is a single server at each location, $\nu_x \psi_x(\nu) = \mathbf{1}\{\nu_x > 0\}$. If there is a single location, i.e. $|\mathbb{D}| = 1$, we get the $M/M/1/\infty$ queue.*

- *Suppose now that each location x consists of $s \in [1, \infty]$ exponential servers, then the departure rate from location x is $\lambda(x, \{o\}) \times \min(\nu_x, s)$ where the first term $\lambda(x, \{o\})$ has the same interpretation as previously and the number of active servers equals $\nu_x \psi_x(\nu) = \min(\nu_x, s)$. If $|\mathbb{D}| = 1$, we get the* M/M/s/∞ *queue.*

Example 12 *Consider now a network consisting of two locations $\mathbb{D} = \{x, y\}$ where motion between locations is possible. We study the three above case:*

- *If there is an infinite number of servers at each location, then the departure rate from location x is $(\lambda(x, \{o\}) + \lambda(x, \{y\})) \times \nu_x$ where the first term $\lambda(x, \{o\}) + \lambda(x, \{y\})$ comprises $\lambda(x, \{o\})$ which is the rate of the exponential time until the exist from the network and $\lambda(x, \{y\})$ which is the rate of the exponential time until the departure to $\{y\}$ and the second term ν_x is the number of active servers at location x when the state of the network is ν.*

- *Suppose now that there is a single server at each location, then the departure rate from location x is $(\lambda(x, \{o\}) + \lambda(x, \{y\})) \times 1\{\nu_x > 0\}$ where the first term $\lambda(x, \{o\}) + \lambda(x, \{y\})$ has the same interpretation as previously and since there is a single server at each location, $\nu_x \psi_x(\nu) = \mathbf{1}\{\nu_x > 0\}$.*

- *Suppose now that each location x consists of $s \in [1, \infty]$ exponential servers, then the departure rate from location x is $(\lambda(x, \{o\}) + \lambda(x, \{y\})) \times \min(\nu_x, s)$ where the first term $\lambda(x, \{o\}) + \lambda(x, \{y\})$ has the same interpretation as previously and the number of active servers is $\nu_x \psi_x(\nu) = \min(\nu_x, s)$.*

8.1.2 Wireless model

We have sometimes to distinguish the motion dynamics from the service dynamics. This is particularly useful for wireless networks.

We describe first the service dynamics:

(a3) Exogenous arrivals come to dx as a Poisson process with intensity $\lambda(o, dx)$.

(b3) Define the CALL-DURATION at a location x as the residual duration at location x before the exit from the system. The call-durations are assumed i.i.d. exponentially distributed with parameters $\lambda(x, \{o\})$ (or equivalently with mean $1/\lambda(x, \{o\})$) and independent from arrivals.

(c3) At a given time t, some users are being served and others are waiting. The state of the system is $\nu \in \mathbb{M}$ counting all the users (being served or waiting) at each location. The number of active servers at location x when the state of the network is ν equals $\nu_x\psi_x(\nu) \leq \nu_x$ (since there are at most ν_x streaming messages served simultaneously). The $\nu_x - \nu_x\psi_x(\nu)$ users which are not being served are placed in a waiting room at note x which is assumed infinitely large.

We describe now the motion dynamics:

(d3) The sojourn duration of users at location x are i.i.d. random variables which are exponentially distributed with rate $\lambda'(x)$ and independent from arrivals and from residual service duration in the system.

(e3) A user finishing its sojourn at location x, is routed to $dy \subset \mathbb{D}$ according to the probability Kernel $p'(x, dy)$ where $p'(x, \mathbb{D}) = 1$.

Proposition 48 *Assume that only active users are moving (waiting users are motionless), then the physical interpretation (a3)–(e3) corresponds to the network generator (6.8) if*

$$\begin{cases} \lambda'(x) = \lambda(x, \mathbb{D}), & x \in \mathbb{D} \\ p'(x, A) = \lambda(x, A)/\lambda'(x), & x \in \mathbb{D}, A \in \mathcal{D} \end{cases} \tag{8.1}$$

and

$$r_{xy}(\nu) = \begin{cases} 1, & x = o, y \in \bar{\mathbb{D}} \\ \psi_x(\nu), & x \in \mathbb{D}, y \in \bar{\mathbb{D}} \end{cases}$$

If all the users (active and waiting ones) are moving, then we have to add the condition

$$\psi_x(\nu) = 1 \tag{8.2}$$

i.e. $r_{xy}(\nu) \equiv 1$ *which corresponds to the Markov Poisson location (MPL) process is defined in Example 3.*

Proof. Suppose that the state of the system at some instant t is some $\nu \in \mathbb{M}$. The next event may be a birth (which corresponds to transitions $\nu \to T_{ox}\nu$), a death (which corresponds to transitions $\nu \to T_{ox}\nu$) or a motion (which corresponds to transitions $\nu \to T_{xy}\nu$).

By (a3)

$$q(\nu, T_{oA}\nu) = \lambda(o, A), \quad A \in \mathcal{D}$$

By (c3), at a given location x, there are $\nu_x\psi_x(\nu)$ active users. By (b3), each of the active users at location x, will exit the network after an exponentially-distributed duration of rate $\lambda(x, \{o\})$. The next death (exit from the network) will take place at the minimum of these durations which is exponentially distributed with rate $\lambda(x, \{o\})\nu_x\psi_x(\nu)$, hence

$$q(\nu, T_{x0}\nu) = \lambda(x, \{o\})\nu_x\psi_x(\nu)$$

By (d3)-(e3), for each user at location x, the next transition from this location to $A \in \mathcal{D}$ will take place after an exponentially-distributed duration of rate $\lambda'(x)p'(x, A)$.

1. Assume that only active users are moving. The next transition of an active user from x to A will take place at the minimum of $\nu_x\psi_x(\nu)$ independent exponentially-distributed durations with rates $\lambda'(x)p'(x,A)$, hence

$$q(\nu, T_{xA}\nu) = \lambda'(x)p'(x,A)\nu_x\psi_x(\nu), \quad x \in \mathbb{D}, A \in \mathcal{D}$$

If we have the conditions (8.1) we retrieve the generator (6.8).

2. Assume that all the users are moving. The next transition of an active user from x to A will take place at the minimum of ν_x independent exponentially-distributed durations with rates $\lambda'(x)p'(x,A)$, hence

$$q(\nu, T_{xA}\nu) = \lambda'(x)p'(x,A)\nu_x, \quad x \in \mathbb{D}, A \in \mathcal{D}$$

If we have the conditions (8.1) and (8.2) we retrieve the generator (6.8).

■

Note that if only active users are moving (waiting users are motionless), then we get a Whittle network as in Section 8.1.1. Hence wired and wireless networks serving streaming traffic may be modelled by the same model in this case. This is also the case if all the users (active and waiting ones) are moving and Condition (8.2) holds true.

Remark 23 *If we assume that $\lambda(x,\{x\}) = 0$, then there is no feedback at the locations. If $\lambda(x,\{x\}) > 0$, then the network is equivalent to another network with the same service rates as the initial one and with routing rates $\tilde{\lambda}(x,A) = \lambda(x, A\backslash\{x\})$. This is due the property of Markov process which says that if we get rid of pseudo-transitions from a Markov process (with stable q-matrix and unique transition functions), we obtain a process equivalent in law with the initial process.*

The correspondent parameters in the above descriptions are

$$\begin{aligned}
\tilde{\lambda}(x) &= \lambda(x) - \lambda(x,\{x\}) = \lambda(x)(1 - p(x,\{x\})), \quad x \in \mathbb{D} \\
\tilde{p}(x,A) &= p(x, A\backslash\{x\}) / (1 - p(x,\{x\})), \quad x \in \mathbb{D}, y \in \bar{\mathbb{D}} \\
\tilde{\lambda}'(x) &= \lambda'(x) - \lambda(x,\{x\}) = \lambda'(x)(1 - p'(x,\{x\})), \quad x \in \mathbb{D} \\
p'(x,A) &= p'(x, A\backslash\{x\}) / (1 - p'(x,\{x\})), \quad x, y \in \mathbb{D}
\end{aligned}$$

8.2 Two spatial loss wireless models

Classical loss models are well adapted to wired communication networks. In wireless communication models, we have to take into account two important aspects, absent in the classical models. The *spatial geometry of the network* cannot be longer reduced to an abstract graph of links but has to capture the relative location of radio channels, which determines their joint feasibility. This spatial component of the model is subject to *changes due to the mobility of users* and also instantaneous changes of radio conditions. One of the consequences of the above circumstances, is that a call can be rejected not only when it is arriving to the network but also when a user is changing its geometrical location while his communication being in progress. Note that this latter can happen even if the user displacement is the only change in the configuration of calls in progress.

Consider a SMQ generator q introduced in Section 6.2. We suppose that it is regular and ergodic. We call the corresponding SMQ process $\{N_t\}$ the *free process* and consider it as describing the evolution of a system without capacity constraints. Thus, q describes arrivals of calls, service demands, service discipline and the mobility of calls. Suppose now that the evolution of the free process is subject to some constraints, which can be expressed as the limitation of the original state space $\mathbb{M}$ to a given fixed measurable subset $\mathbb{M}^{\mathrm{f}} \subset \mathbb{M}$ of feasible states. The constrained (restricted) process, started off at an initial state in $\mathbb{M}^{\mathrm{f}}$ follows the same dynamic as the free process as long as it stays in $\mathbb{M}^{\mathrm{f}}$, and will be forced to modify its behavior each time an attempt of a transition from $\mathbb{M}^{\mathrm{f}}$ to $\mathbb{M} \setminus \mathbb{M}^{\mathrm{f}}$ occurs. We will consider two possible behaviors adopted at such epochs. They lead to two following different models.

- *Transition blocking model.* In this model we suppose that all the transitions from a state $\nu \in \mathbb{M}^{\mathrm{f}}$ to a state $\mathbb{M} \setminus \mathbb{M}^{\mathrm{f}}$ are "blocked", which means that the process remains in the state ν and continues its evolution driven by q. The dynamics of the restricted process $\{N_t^{\mathrm{tb}}\}$ in this model is described by a generator q^{tb} where

$$q^{\mathrm{tb}}(\nu, \Gamma) = \begin{cases} q(\nu, \Gamma \cap \mathbb{M}^{\mathrm{f}}) & \text{if } \nu \in \mathbb{M}^{\mathrm{f}} \\ q(\nu, \Gamma) & \text{if } \nu \notin \mathbb{M}^{\mathrm{f}} \end{cases} \tag{8.3}$$

$$q^{\mathrm{tb}}(\nu) = \begin{cases} q(\nu, \mathbb{M}^{\mathrm{f}}) & \text{if } \nu \in \mathbb{M}^{\mathrm{f}} \\ q(\nu, \mathbb{M}) & \text{if } \nu \notin \mathbb{M}^{\mathrm{f}} \end{cases}$$

Note that q^{tb} is also a SMQ generator, with the same routing kernel λ and the service rates

$$r_{xy}^{\mathrm{tb}}(\nu) = \begin{cases} r_{xy}(\nu) 1(T_{xy}\nu \in \mathbb{M}^{\mathrm{f}}) & \text{if } \nu \in \mathbb{M}^{\mathrm{f}} \\ r_{xy}(\nu) & \text{if } \nu \notin \mathbb{M}^{\mathrm{f}} \end{cases} \tag{8.4}$$

and we will always assume that it is regular and ergodic. Moreover q^{tb} is the so called *truncation* of q and its invariant measure is equal to the truncation of the invariant measure of q to $\mathbb{M}^{\mathrm{f}}$, at least if q^{tb} is reversible. This makes possible an expression of blocking probabilities in relatively simple formula, which we call the spatial Erlang formula.

Not that in the transition blocking model there are no losses of calls in progress: an unauthorized displacement is blocked and the call in question rests at its previous location until the next event. This might be seen as a not very realistic assumption in the context of modeling of voice calls.

- *Forced termination model.* In this model we suppose that all the call arrivals that would result in taking the process to a state outside $\mathbb{M}^{\mathrm{f}}$ are blocked, as in the transition blocking model, however an attempt of the displacement of a call in progress that would take the process to $\mathbb{M} \setminus \mathbb{M}^{\mathrm{f}}$ lead to the forced termination of this call. The evolution of the process is thus described by the following generator

$$q^{\mathrm{ft}}(\nu, \Gamma) = \begin{cases} q(\nu, \Gamma \cap \mathbb{M}^{\mathrm{f}}) & \text{for } \Gamma \subset T_{\bar{\mathbb{D}}\mathbb{D}}\nu, \nu \in \mathbb{M}^{\mathrm{f}} \\ q(\nu, \Gamma) + q(\nu, T_{A\mathbb{D}}\nu \setminus \mathbb{M}^{\mathrm{f}}) & \text{for } \Gamma = T_{Ao}\nu, A \in \mathcal{D}, \nu \in \mathbb{M}^{\mathrm{f}} \\ q(\nu, \Gamma) & \text{for } \Gamma \in \mathcal{M}, \nu \notin \mathbb{M}^{\mathrm{f}} \end{cases} \tag{8.5}$$

$$q^{\mathrm{ft}}(\nu) = \begin{cases} q(\nu, \mathbb{M}^{\mathrm{f}} \cup T_{\mathbb{D}\bar{\mathbb{D}}}\nu) & \text{if } \nu \in \mathbb{M}^{\mathrm{f}} \\ q(\nu, \mathbb{M}) & \text{if } \nu \notin \mathbb{M}^{\mathrm{f}} \end{cases}$$

and we assume that $\mathbb{M}^{\mathrm{f}}$ is *closed* with respect to transition $T_{xo}\nu$ for all $x \in \mathbb{D}$. Note that q^{ft} is also a SMQ generator, with the same routing kernel λ and the service rates $r^{\mathrm{ft}}_{xy}(\nu) = r^{\mathrm{tb}}_{xy}(\nu)$ for $y \neq o$ and $r^{\mathrm{ft}}_{xo}(\nu) = r_{xo}(\nu) + \int_{\mathbb{D}} r_{xy}(\nu) 1(T_{xy}\nu \notin \mathbb{M}^{\mathrm{f}})\, \lambda(x, dy)/\lambda(x, o)$ if $\nu \in \mathbb{M}^{\mathrm{f}}$ and $r_{xo}(\nu)$ otherwise. We will always assume that q^{ft} it is regular and ergodic. However it cannot be seen as a truncation of q and typically its invariant measure is not explicitly known even if the invariant measure of q is known.

In the remaining part of this section we will study loss probabilities in the above models.

8.2.1 Transition blocking model

In this section we will concentrate on the transition blocking model. Suppose q is a regular, ergodic SMQ generator, as described in Section 6.2, and call its unique invariant probability measure Π. Consider the SMQ process $\{N_t\}$ corresponding to q as the free process (without capacity constraints; see the discussion above). Fix a measurable subset $\mathbb{M}^{\mathrm{f}}$ of its state space $\mathbb{M}$ as the subspace of all feasible states of the restricted process $\{N^{\mathrm{tb}}_t\}$ that evolves according the generator q^{tb} given by (8.3). In what follows we assume that the restricted process is also ergodic and has a particular form of the limiting distribution Π^{tb} being the *truncation* of Π to $\mathbb{M}^{\mathrm{f}}$. This truncation property does not always hold, and one simple sufficient condition for it to hold is as follows (cf [106, Proposition 3.14]):

Lemma 14 *Suppose that* $\Pi(\mathbb{M}^{\mathrm{f}}) > 0$. *The invariant probability measure* Π^{tb} *of the restricted process* $\{N^{\mathrm{tb}}_t;\, t \geq 0\}$ *is given by*

$$\Pi^{\mathrm{tb}}(\Gamma) = \frac{\Pi(\Gamma \cap \mathbb{M}^{\mathrm{f}})}{\Pi(\mathbb{M}^{\mathrm{f}})} \tag{8.6}$$

if and only if Π *satisfies the following balance equation*

$$q(\nu, \mathbb{M}^{\mathrm{f}})\Pi(d\nu) = \int_{\mathbb{M}^{\mathrm{f}}} q(\mu, d\nu)\, \Pi(d\mu), \quad \nu \in \mathbb{M}^{\mathrm{f}}$$

The truncation property (8.6) holds in particular if $\{N_t;\, t \ge 0\}$ *is* reversible *on* $\mathbb{M}^{\rm f}$ *or* $\mathbb{M} \setminus \mathbb{M}^{\rm f}$ *with respect to* Π*, meaning that* q *satisfies the following* detailed balance equation

$$q(\nu, d\mu)\Pi(d\nu) = q(\mu, d\nu)\Pi(d\mu)$$

for, either $\nu, \mu \in \mathbb{M}^{\rm f}$ *or* $\nu, \mu \in \mathbb{M} \setminus \mathbb{M}^{\rm f}$.

In what follows we assume that (8.6) is true, in particular $\Pi(\mathbb{M}^{\rm f}) > 0$, and we call $\Pi(\mathbb{M}^{\rm f})$ the *feasibility probability.* Note that $\Pi(\mathbb{M}^{\rm f})$ is the probability that the free process in steady state takes its value in the feasible (for the restricted process) part of the space.

Blocking probabilities

For given subsets $A \in \bar{\mathcal{D}}$, $B \in \mathcal{D}$ we are interested, roughly speaking, in the "ergodic frequency" $p^{\rm tb}_{AB}$ of "blocked transitions" $\nu \to T_{xy}\nu$ for $x \in A$, $y \in B$ of the process driven by $q^{\rm tb}$. (The quantity $p^{\rm tb}_{AB}$ will be defined precisely later; cf. Equation (8.7).) Actually $p^{\rm tb}_{AB}$ cannot be well defined given realizations of the process $N^{\rm tb}_t$ because "we do not observe" blocked transitions there. In order to formalize this notion note that the time-epochs and departure-arrival locations of the *blocked transitions* can be modeled by a double stochastic Poisson point process $\Phi^{\rm tb}_0 = \sum_i \delta_{(t_i, x_i, y_i)}$ driven by $N^{\rm tb}_t$, where t_i, x_i, y_i denote, respectively, the time-epochs, departure and arrival locations of blocked transition of $N^{\rm tb}_t$. Given a realization $N^{\rm tb}_{\cdot} = \{N^{\rm tb}_t, t \ge 0\}$, $\Phi^{\rm tb}_0$ is a Poisson point process with intensity measure $\Lambda^{\rm tb}_{N^{\rm tb}_{\cdot}}$ on $(0,\infty) \times (\bar{\mathbb{D}})^2$, given by $\Lambda^{\rm tb}_{N_{\cdot}}(D \times A \times B) = \int_D q(N^{\rm tb}_t, T_{AB}N^{\rm tb}_t \setminus \mathbb{M}^{\rm f})dt$. Denote also by $\Phi^{\rm tb}_1$ the point process on $(0,\infty) \times (\bar{\mathbb{D}})^2$ associated to ("true") transitions of $N^{\rm tb}_t$; i.e., $\Phi^{\rm tb}_1(D \times A \times B) = \sum_{s>0} 1(s \in D, N^{\rm tb}_s = T_{xy}N^{\rm tb}_{s-}, x \in A, y \in B)$. Let $\Phi^{\rm tb} = \Phi^{\rm tb}_0 + \Phi^{\rm tb}_1$ be the superposition of $\Phi^{\rm tb}_i$, $i = 0, 1$. Finally define the blocking probability for the transitions $\nu \to T_{AB}(\nu)$ for some $A, B \in \bar{\mathbb{D}}$ and $\nu \in \mathbb{M}^{\rm f}$ (will call them transitions from A to B for short) as the following limiting ratio of blocked transitions to all transitions

$$p^{\rm tb}_{AB} = \lim_{t\to\infty} \frac{\Phi^{\rm tb}_0((0,t] \times A \times B)}{\Phi^{\rm tb}((0,t] \times A \times B)} \tag{8.7}$$

The above limit exists by the following result.

Lemma 15 *Suppose that* $\emptyset$ *is a positive recurrent state for* $q^{\rm tb}$ *with the limiting distribution* $\Pi^{\rm tb}$*. If*

$$\mathbf{E}_{\Pi^{\rm tb}}[q(N, \mathbb{M})] < \infty \tag{8.8}$$

then

$$\lim_{t\to\infty} \frac{1}{t}\Phi^{\rm tb}_0((0,t] \times A \times B) = \mathbf{E}_{\Pi^{\rm tb}}[q(N, T_{AB}N \setminus \mathbb{M}^{\rm f})]$$
$$\lim_{t\to\infty} \frac{1}{t}\Phi^{\rm tb}_1((0,t] \times A \times B) = \mathbf{E}_{\Pi^{\rm tb}}[q(N, T_{AB}N \cap \mathbb{M}^{\rm f})]$$

a.s. for any initial value $N^{\rm tb}_0 = \nu$ *for which the return time to* $\emptyset$ *is a.s. finite.*

Proof. Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P}_\Pi)$ on which $\{N_t\}_{t\in\mathbb{R}}$ and both point processes Φ_i^{tb} $(i=0,1)$ are (time) stationary. Note that the expectation corresponding to the distribution of $\{N_t\}_{t\geq 0}$ under $\mathbf{P}_\Pi$ is $\mathbf{E}_\Pi$. Condition (8.8) implies

$$\lambda_1 = \mathbf{E}_{\Pi^{\mathrm{tb}}}[\Phi_1^{\mathrm{tb}}((0,1]\times\bar{\mathbb{D}}\times\bar{\mathbb{D}})] = \mathbf{E}_{\Pi^{\mathrm{tb}}}[q^{\mathrm{tb}}(N_0^{\mathrm{tb}})] \leq \mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N_0^{\mathrm{tb}}, \mathbb{M})] < \infty$$

where the second equality is by Lévy's formula (6.44). Similarly, since Φ_0^{tb} is a doubly stochastic Poisson point process,

$$\lambda_0 = \mathbf{E}_{\Pi^{\mathrm{tb}}}[\Phi_0^{\mathrm{tb}}((0,1]\times\bar{\mathbb{D}}\times\bar{\mathbb{D}})] = \int_0^1 \mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N_t^{\mathrm{tb}}, \mathbb{M}\backslash\mathbb{M}^{\mathrm{f}})]\,dt \leq \mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N_0^{\mathrm{tb}}, \mathbb{M})] < \infty$$

For given $A, B \in \bar{\mathcal{D}}$ the processes $X_t^i = \Phi_i^{\mathrm{tb}}((0,t]\times A\times B)$ $(i=1,2)$ are cumulative with the imbedded renewal process being the epochs of successive visits of N_t^{tb} at $\emptyset$. By (6.43) we have then

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}\Phi_1^{\mathrm{tb}}((0,t]\times A\times B) &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[\Phi_1^{\mathrm{tb}}((0,1]\times A\times B)]\\ &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N_0^{\mathrm{tb}}, T_{AB}N_0^{\mathrm{tb}}\cap\mathbb{M}^{\mathrm{f}})]\end{aligned}$$

where the second equality follows from Lévy's formula (6.44). Similarly, by the fact that Φ_0^{tb} is a doubly stochastic Poisson point process

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}\Phi_0^{\mathrm{tb}}((0,t]\times A\times B) &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[\Phi_0^{\mathrm{tb}}([0,1]\times A\times B)]\\ &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[\Lambda_{N.}^{\mathrm{tb}}((0,1]\times A\times B)]\\ &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N_0^{\mathrm{tb}}, T_{AB}N_0^{\mathrm{tb}}\setminus\mathbb{M}^{\mathrm{f}})]\end{aligned}$$

This completes the proof. ■

The following result immediately follows from Lemma 15.

Proposition 49 *If the conditions of Lemma* 15 *are satisfied, then*

$$p_{AB}^{\mathrm{tb}} = \frac{\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N, T_{AB}N\setminus\mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N, T_{AB}N)]}$$

Corollary 10 *If the conditions of Lemma* 15 *are satisfied then for* $B\in\mathcal{D}$*,*

$$p_{oB}^{\mathrm{tb}} = \frac{\int_B p^{\mathrm{tb}}(o,y)\mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)]\lambda(o,dy)}{\int_B \mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)]\lambda(o,dy)}$$

where

$$p^{\mathrm{tb}}(o,y) = \frac{\mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)1(T_{oy}N\notin\mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)]} \qquad y\in\mathbb{D} \tag{8.9}$$

called the blocking rate *on the transition from* o *to* y*. Suppose moreover that (8.6) holds. Then*

$$p^{\mathrm{tb}}(o,y) = \frac{\mathbf{E}_\Pi[r_{oy}(N)1(N\in\mathbb{M}^{\mathrm{f}}, T_{oy}N\notin\mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_\Pi[r_{oy}(N)1(N\in\mathbb{M}^{\mathrm{f}})]} \qquad y\in\mathbb{D} \tag{8.10}$$

If $r_{oy}(\nu) \equiv 1$ for Π-almost all $\nu \in \mathbb{M}^{\mathrm{f}}$, in particular if q is a MPL generator, then we have

$$p^{\mathrm{tb}}(o,y) = \frac{\Pi(N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}})}{\Pi(N \in \mathbb{M}^{\mathrm{f}})} \tag{8.11}$$

If the conditions of Proposition 28 (or Proposition 29) for the free generator q hold, then

$$p^{\mathrm{tb}}(o,y) = \frac{\mathbf{E}_{\Pi_\rho}[r_{oy}(N)\Psi(N)1(N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi_\rho}[r_{oy}(N)\Psi(N)1(N \in \mathbb{M}^{\mathrm{f}})]} \quad y \in \mathbb{D} \tag{8.12}$$

Proof. The first equation follows from Proposition 49 and the fact that

$$\begin{aligned}\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{oB}N)] &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[\int_B r_{oy}(N)\,\lambda(o,dy)] \\ &= \int_B \mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)]\lambda(o,dy)\end{aligned}$$

$$\begin{aligned}\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{oB}N \setminus \mathbb{M}^{\mathrm{f}})] &= \mathbf{E}_{\Pi^{\mathrm{tb}}}[\int_B r_{oy}(N)\,1\left(T_{oy}N \notin \mathbb{M}^{\mathrm{f}}\right)\lambda(o,dy)] \\ &= \int_B \mathbf{E}_{\Pi^{\mathrm{tb}}}[r_{oy}(N)\,1\left(T_{oy}N \notin \mathbb{M}^{\mathrm{f}}\right)]\lambda(o,dy)\end{aligned}$$

Equation (8.10) is immediate from (8.9) and (8.6). Equation (8.11) is immediate from (8.10). Equation (8.12) is a consequence of the fact that the distribution Π is Gibbs with density Ψ with respect to Π_ρ. ■

Remark 24 *Note that formula (8.11) might be seen as a* spatial extension of the classical Erlang formula.

Corollary 11 *If the conditions of Proposition 29 for the free generator q hold as well as Lemma 15 and condition (8.6), then for $A, B \in \mathcal{D}$,*

$$p^{\mathrm{tb}}_{AB} = \frac{\int_A\int_B p^{\mathrm{tb}}(x,y)\mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\,1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]\rho(dx)\,\lambda(x,dy)}{\int_A\int_B \mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\,1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]\rho(dx)\,\lambda(x,dy)}$$

where

$$p^{\mathrm{tb}}(x,y) = \frac{\mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\,1\left(T_{ox}N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}}\right)\Psi(T_{ox}N)]}{\mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\,1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]} \quad x,y \in \mathbb{D} \tag{8.13}$$

called the blocking rate *on the transition from x to y. If $r_{xy}(\nu) \equiv 1$ for Π-almost all $\nu \in \mathbb{M}^{\mathrm{f}}$, in particular if q is a MPL generator, then we have*

$$p^{\mathrm{tb}}(x,y) = \frac{\Pi(T_{ox}N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}})}{\Pi(T_{ox}N \in \mathbb{M}^{\mathrm{f}})} \quad x \in \mathbb{D}, y \in \mathbb{D}. \tag{8.14}$$

Proof. From Proposition 49 we deduce that

$$
\begin{aligned}
p^{\mathrm{tb}}_{AB} &= \frac{\mathbf{E}_{\Pi^{\mathrm{tb}}}[\int_A \int_B r_{xy}(N)\, N(dx)\, 1\left(T_{xy}N \notin \mathbb{M}^{\mathrm{f}}\right) \lambda(x, dy)]}{\mathbf{E}_{\Pi^{\mathrm{tb}}}[\int_A \int_B r_{xy}(N)\, N(dx)\, \lambda(x, dy)]} \\
&= \frac{\mathbf{E}_{\Pi}[\int_A \int_B r_{xy}(N)\, 1\left(N \in \mathbb{M}^{\mathrm{f}}, T_{xy}N \notin \mathbb{M}^{\mathrm{f}}\right) N(dx)\, \lambda(x, dy)]}{\mathbf{E}_{\Pi}[\int_A \int_B r_{xy}(N)\, 1(N \in \mathbb{M}^{\mathrm{f}}) N(dx)\, \lambda(x, dy)]} \\
&= \frac{\mathbf{E}_{\Pi}[\int_A \int_B r_{xy}(T_{ox}N)\, 1\left(T_{ox}N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}}\right) \Psi(T_{ox}N)/\Psi(N)\rho(dx)\, \lambda(x, dy)]}{\mathbf{E}_{\Pi}[\int_A \int_B r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)/\Psi(N)\rho(dx)\, \lambda(x, dy)]} \\
&= \frac{\int_A \int_B \mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\, 1\left(T_{ox}N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}}\right) \Psi(T_{ox}N)]\rho(dx)\, \lambda(x, dy)}{\int_A \int_B \mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]\rho(dx)\, \lambda(x, dy)} \\
&= \frac{\int_A \int_B p^{\mathrm{tb}}(x, y)\mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]\rho(dx)\, \lambda(x, dy)}{\int_A \int_B \mathbf{E}_{\Pi_\rho}[r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\Psi(T_{ox}N)]\rho(dx)\, \lambda(x, dy)}
\end{aligned}
$$

where for the third equality we use Proposition 20 for

$$g(N, x) = 1(x \in A) \int_B r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}})\lambda(x, dy)$$

for the denominator and

$$g(N, x) = 1(x \in A) \int_B r_{xy}(T_{ox}N)\, 1(T_{ox}N \in \mathbb{M}^{\mathrm{f}}, T_{oy}N \notin \mathbb{M}^{\mathrm{f}})\lambda(x, dy)$$

for the numerator. ■

Average number of users

In some particular cases, the average number of users in the stationary state is related explicitly to the blocking rates as shown in the following lemma.

Lemma 16 *If $r_{xy}(\nu) \equiv 1$ for Π_ρ-almost all $\nu \in \mathbb{M}^{\mathrm{f}}$, in particular if q is a MPL generator, then we have*

$$\mathbf{E}_{\Pi^{\mathrm{tb}}}[N(B)] = \int_B \left[1 - p^{\mathrm{tb}}(o, x)\right] \rho(dx), \quad B \in \mathcal{D} \tag{8.15}$$

Assume that $\lambda(x, o)$ is constant. If moreover $\rho(dx) = \lambda(o, dx)/\lambda(x, o)$ in particular if q is a SBD generator, then

$$\mathbf{E}_{\Pi^{\mathrm{tb}}}[N(B)] = (1 - p_{oB})\, \rho(B) \tag{8.16}$$

Proof. We have

$$\begin{aligned}
\mathbf{E}_{\Pi^{\rm tb}}\left[N\left(B\right)\right] &= \int_{\mathbb{M}} N\left(B\right)\Pi^{\rm tb}\left(dN\right) \\
&= \Pi(\mathbb{M}^{\rm f})^{-1}\int_{\mathbb{M}} 1\left\{N \in \mathbb{M}^{\rm f}\right\} N\left(B\right)\Pi\left(dN\right) \\
&= \Pi(\mathbb{M}^{\rm f})^{-1}\mathbf{E}_{\Pi}\left[1\left\{N \in \mathbb{M}^{\rm f}\right\} N\left(B\right)\right] \\
&= \Pi(\mathbb{M}^{\rm f})^{-1}\int_{B}\mathbf{E}_{\Pi}\left[1\left\{N \in \mathbb{M}^{\rm f}\right\} N\left(dx\right)\right] \\
&= \Pi(\mathbb{M}^{\rm f})^{-1}\int_{B}\mathbf{E}_{\Pi}\left[1\left\{N + \delta_x \in \mathbb{M}^{\rm f}\right\} \rho\left(dx\right)\right] \\
&= \int_{B}\Pi^{\rm tb}\left(N + \delta_x \in \mathbb{M}^{\rm f}\right)\rho\left(dx\right)
\end{aligned}$$

where the second equality is due to the truncation property and for the fifth one we use Proposition 20 for $g\left(N,x\right) = 1\left(N + \delta_x \in \mathbb{M}^{\rm f}\right)$.

If $\lambda(x,o)$ is constant, then

$$p_{oB} = \frac{1}{\lambda(o,B)}\int_B p^{\rm tb}\left(o,x\right)\lambda(o,dx) = \frac{1}{\rho(B)}\int_B p^{\rm tb}\left(o,x\right)\rho(dx)$$

from which we get (8.16). ■

8.2.2 Forced termination model

In this section we will consider the forced termination model and will clearly distinguish between *blocking of new arrivals* and *cutting of existing calls in progress.*

We suppose that $q^{\rm ft}$ is regular and ergodic, and it has the an invariant distribution that we denote by $\Pi^{\rm ft}$. Note that $\Pi^{\rm ft}$ is a probability distribution on $\mathbb{M}$.

Blocking probabilities

As in the previous section, the *blocked arrivals* can be modeled by a double stochastic Poisson point process $\Phi_0^{\rm ft} = \sum_i \delta_{(t_i y_i)}$ driven by $N_t^{\rm ft}$, where t_i, y_i denote, respectively, the time-epochs and arrival locations of blocked transition of $N_t^{\rm ft}$. Given a realization $N_\cdot^{\rm ft} = \{N_t^{\rm ft}, t \geq 0\}$, $\Phi_0^{\rm ft}$ is a Poisson point process with intensity measure $\Lambda_{N_\cdot^{\rm ft}}^{\rm ft}$ on $(0,\infty)\times\mathbb{D}$, given by $\Lambda_{N_\cdot}^{\rm ft}(D\times B) = \int_D q(N_t^{\rm ft}, T_{oB}N_t^{\rm ft} \setminus \mathbb{M}^{\rm f})dt$. Denote also by $\Phi_1^{\rm tb}$ the point process on $(0,\infty)\times\mathbb{D}$ associated to ("true") arrivals of $N_t^{\rm ft}$; i.e., $\Phi_1^{\rm ft}(D\times B) = \sum_{s>0} 1(s \in D, N_s^{\rm ft} = T_{oy}N_{s-}^{\rm ft}, y \in B)$. Let $\Phi^{\rm ft} = \Phi_0^{\rm ft} + \Phi_1^{\rm ft}$ be the superposition of $\Phi_i^{\rm ft}$, $i = 0,1$. Finally define the blocking probability for the transitions $\nu \to T_{oB}(\nu)$ for some $B \in \mathbb{D}$ and $\nu \in \mathbb{M}^{\rm f}$ (will call them arrivals to B for short) as the following limiting ratio of blocked transitions to all transitions

$$p_{oB}^{\rm ft} = \lim_{t\to\infty}\frac{\Phi_0^{\rm ft}((0,t]\times B)}{\Phi^{\rm ft}((0,t]\times B)}$$

The following result can be proved along the same lines as Lemma 15.

Proposition 50 *Suppose that $\emptyset$ is a positive recurrent state for q^{ft} with the limiting distribution Π^{ft}. If*

$$\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N,\mathbb{M})] < \infty \tag{8.17}$$

then

$$\lim_{t\to\infty} \frac{1}{t}\Phi_0^{\mathrm{ft}}((0,t]\times B) = \mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{oB}N \setminus \mathbb{M}^{\mathrm{f}})]$$
$$\lim_{t\to\infty} \frac{1}{t}\Phi_1^{\mathrm{ft}}((0,t]\times B) = \mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{oB}N \cap \mathbb{M}^{\mathrm{f}})]$$

a.s. for any initial value $N_0^{\mathrm{ft}} = \nu$ for which the return time to $\emptyset$ is a.s. finite. Moreover

$$p_{oB}^{\mathrm{ft}} = \frac{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{oB}N \setminus \mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{oB}N)]} = \frac{\int_B p^{\mathrm{ft}}(o,y)\mathbf{E}_{\Pi^{\mathrm{ft}}}[r_{oy}(N)]\lambda(o,dy)}{\int_B \mathbf{E}_{\Pi^{\mathrm{ft}}}[r_{oy}(N)]\lambda(o,dy)}$$

where

$$p^{\mathrm{ft}}(o,y) = \frac{\mathbf{E}_{\Pi^{\mathrm{ft}}}[r_{oy}(N)1(T_{oy}N \notin \mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{ft}}}[r_{oy}(N)]} \qquad y \in \mathbb{D} \tag{8.18}$$

Cut probabilities

We are now interested in forced terminations (cuts) of the service. Looking at the form of the generator q^{ft} we see that each transition $\nu \to T_{xo}\nu$ for some $x \in \nu$ can be, independently of everything else, either a "regular termination" with probability

$$\tau_{\nu,x}(\{o\}) = \frac{r_{xo}(\nu)}{r_{xo}^{\mathrm{ft}}(\nu)}$$

or a "forced termination" due to an unsuccessful displacement to $B \in \mathbb{D}$ with probability

$$\tau_{\nu,x}(B) = \frac{\int_B r_{xy}(\nu)1(T_{xy}\nu \notin \mathbb{M}^{\mathrm{f}})\,\lambda(x,dy)}{r_{xo}^{\mathrm{ft}}(\nu)\lambda(x,o)}$$

So we can model different terminations of the process N^{ft} by a marked point process $\Phi_2^{\mathrm{ft}} = \sum_i \delta_{(t_i,x_i,y_i)}$ where t_i are termination epochs, $x_i \in \mathbb{D}$ denote the departure locations and $y_i \in \bar{\mathbb{D}}$ denote the termination status: $y_i = o$ if it is a regular one and $y_i \in \mathbb{D}$ if it is caused by an unsuccessful displacement from x_i to y_i. Note that given a realization of N_t^{ft} the points t_i and marks x_i are known, and we assume that y_i are independently chosen with the distribution $\tau_{N_{t_i}^{\mathrm{ft}},x_i}(B)$. Considering marked point processes of epochs and departure-arrival location of all transition we can express various ergodic limit fractions of users cut on when trying to move from some given $A \in \mathcal{D}$ to $B \in \mathcal{D}$. Here we will show only how to treat ergodic limit fractions c of users that are forced to terminate during their sojourn in the network

$$c = \lim_{t\to\infty} \frac{\Phi_2^{\mathrm{ft}}((0,t]\times\mathbb{D}\times\mathbb{D})}{\Phi_1^{\mathrm{ft}}((0,t]\times\mathbb{D})}$$

Proposition 51 *Suppose that $\emptyset$ is a positive recurrent state for q^{ft} with the limiting distribution Π^{ft}. If*

$$\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{o\mathbb{D}}N \cap \mathbb{M}^{\mathrm{f}})] < \infty \tag{8.19}$$

and

$$\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{\mathbb{D}\mathbb{D}}N \setminus \mathbb{M}^{\mathrm{f}})] < \infty \tag{8.20}$$

then the limit c exists almost surely for any initial value $N_0^{\mathrm{ft}} = \nu$ for which the return time to $\emptyset$ is finite and

$$c = \frac{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{\mathbb{D}\mathbb{D}}N \setminus \mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{o\mathbb{D}}N \cap \mathbb{M}^{\mathrm{f}})]} \tag{8.21}$$

$$= 1 - \frac{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{\mathbb{D}o}N)]}{\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{o\mathbb{D}}N \cap \mathbb{M}^{\mathrm{f}})]} \tag{8.22}$$

Proof. Note that the process $X_t = \Phi_2^{\mathrm{ft}}([0,t] \times \mathbb{D} \times \mathbb{D})$ is cumulative with the imbedded renewal process being the epochs of successive visits of N_t^{ft} at $\emptyset$. Following the same lines as in the proof of Lemma 15 we find under condition (8.20) that $\lim_{t\to\infty} 1/t\ \Phi_2^{\mathrm{ft}}((0,t] \times \mathbb{D}^2) = \mathbf{E}_{\Pi^{\mathrm{ft}}}[\Phi_2^{\mathrm{ft}}((0,1] \times \mathbb{D}^2)]$. Similarly, under condition (8.19) $\lim_{t\to\infty} 1/t\ \Phi_1^{\mathrm{ft}}((0,t] \times \mathbb{D}) = \mathbf{E}_{\Pi^{\mathrm{ft}}}[\Phi_1^{\mathrm{ft}}((0,1] \times \mathbb{D})]$. By Lévy's formula we obtain (8.21).

Writing that the arrival rate equals the departure rate, gives

$$\mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{\mathbb{D}\mathbb{D}}N \setminus \mathbb{M}^{\mathrm{f}})] + \mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{\mathbb{D}o}N)] = \mathbf{E}_{\Pi^{\mathrm{ft}}}[q(N, T_{o\mathbb{D}}N \cap \mathbb{M}^{\mathrm{f}})]$$

which together with (8.21) imply (8.22). ■

8.2.3 Approximation of the cut probability

We don't have an explicit expression of the stationary distribution of the forced termination model. Hence Formula (8.21) doesn't give an explicit way to calculate the cut probability. We define for the transition blocking model a *fictitious*-cut probability which will be shown to be a *good approximation* of the cut probability of the *forced termination* model in §10.4.

For the transition blocking model, we define the MOTION-BLOCKING as the ratio of *blocked motions* to *regular terminations*

$$d^{\mathrm{tb}} = \lim_{t\to\infty} \frac{\Phi_0^{\mathrm{tb}}((0,t] \times \mathbb{D} \times \mathbb{D})}{\Phi^{\mathrm{tb}}((0,t] \times \mathbb{D} \times \{o\})} \tag{8.23}$$

The FICTITIOUS-CUT probability is defined as

$$c^{\mathrm{tb}} = \frac{d^{\mathrm{tb}}}{1 + d^{\mathrm{tb}}}$$

The following proposition permits to relate the motion-blocking to the stationary distribution of the transition blocking model.

Proposition 52 *The motion-blocking equals*

$$d^{\mathrm{tb}}=\frac{\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{\mathbb{D}\mathbb{D}}N\setminus\mathbb{M}^{\mathrm{f}})]}{\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{\mathbb{D}o}N)]} \tag{8.24}$$

If $r_{xy}(\nu)\equiv 1$ for Π_ρ-almost all $\nu\in\mathbb{M}^{\mathrm{f}}$, then

$$d^{\mathrm{tb}}=\frac{\int_{\mathbb{D}\times\mathbb{D}}\lambda(x,dy)\,\Pi^{\mathrm{tb}}\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}},N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)\rho(dx)}{\int_{\mathbb{D}}\lambda(x,\{o\})\left[1-p^{\mathrm{tb}}(o,x)\right]\rho(dx)} \tag{8.25}$$

Proof. From Lemma 15 we get (8.24). Assume that $r_{xy}(\nu)\equiv 1$ for Π_ρ-almost all $\nu\in\mathbb{M}^{\mathrm{f}}$, in particular if q is a MPL generator. For the denominator of the right-hand side of the above display we have

$$\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{\mathbb{D}o}N)]=\int_{\mathbb{D}}\lambda(x,\{o\})\,\mathbf{E}_{\Pi^{\mathrm{tb}}}[N(dx)]=\int_{\mathbb{D}}\lambda(x,\{o\})\left[1-p^{\mathrm{tb}}(o,x)\right]\rho(dx)$$

where the last equality is due to Lemma 16.

For the numerator, we have

$$\mathbf{E}_{\Pi^{\mathrm{tb}}}[q(N,T_{\mathbb{D}\mathbb{D}}N\setminus\mathbb{M}^{\mathrm{f}})]=\int_{\mathbb{D}\times\mathbb{D}}\lambda(x,dy)\,\mathbf{E}_{\Pi^{\mathrm{tb}}}[1(T_{xy}N\notin\mathbb{M}^{\mathrm{f}})N(dx)]$$

We calculate

$$\begin{aligned}\mathbf{E}_{\Pi^{\mathrm{tb}}}[1(T_{xy}N\notin\mathbb{M}^{\mathrm{f}})N(dx)]&=\Pi(\mathbb{M}^{\mathrm{f}})^{-1}\mathbf{E}_{\Pi}\left[1(N\in\mathbb{M}^{\mathrm{f}},T_{xy}N\notin\mathbb{M}^{\mathrm{f}})N(dx)\right]\\&=\Pi(\mathbb{M}^{\mathrm{f}})^{-1}\Pi\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}},N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)\rho(dx)\\&=\Pi^{\mathrm{tb}}\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}},N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)\rho(dx)\end{aligned}$$

where the first equality is due to the truncation property and for the second one we use Proposition 20 for $g(N,x)=1\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}},N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)$ (note that $g(N-\delta_x,x)=1(N\in\mathbb{M}^{\mathrm{f}},N-\delta_x+\delta_y\notin\mathbb{M}^{\mathrm{f}})=1(N\in\mathbb{M}^{\mathrm{f}},T_{xy}N\notin\mathbb{M}^{\mathrm{f}})$). ■

Example 13 *Assume that*

$$\Pi^{\mathrm{tb}}\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}},N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)=\Pi^{\mathrm{tb}}\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}}\right)\Pi^{\mathrm{tb}}\left(N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)$$

(which is the case if $\mathbb{M}^{\mathrm{f}}$ is in the form (8.27)) then

$$\begin{aligned}\mathbf{E}_{\Pi^{\mathrm{tb}}}[1(T_{xy}N\notin\mathbb{M}^{\mathrm{f}})N(dx)]&=\Pi^{\mathrm{tb}}\left(N+\delta_x\in\mathbb{M}^{\mathrm{f}}\right)\Pi^{\mathrm{tb}}\left(N+\delta_y\notin\mathbb{M}^{\mathrm{f}}\right)\rho(dx)\\&=p^{\mathrm{tb}}(o,y)\left[1-p^{\mathrm{tb}}(o,x)\right]\rho(dx)\end{aligned}$$

Hence, from (8.25), we get

$$d^{\mathrm{tb}}=\frac{\int_{\mathbb{D}\times\mathbb{D}}\lambda(x,dy)\,p^{\mathrm{tb}}(o,y)\left[1-p^{\mathrm{tb}}(o,x)\right]\rho(dx)}{\int_{\mathbb{D}}\lambda(x,\{o\})\left[1-p^{\mathrm{tb}}(o,x)\right]\rho(dx)} \tag{8.26}$$

Assume now that each location plays the same role then

$$\begin{aligned} d^{\mathrm{tb}} &= p^{\mathrm{tb}}(o,\cdot)\frac{\int_{\mathbb{D}\times\mathbb{D}}\lambda(x,dy)\,\rho(dx)}{\int_{\mathbb{D}}\lambda(x,\{o\})\,\rho(dx)} \\ &= p^{\mathrm{tb}}(o,\cdot)\frac{\lambda(\cdot,\mathbb{D})}{\lambda(\cdot,o)} \end{aligned}$$

Assume moreover that $\mathbb{D}$ *is discrete and*

$$\mathbb{M}^{\mathrm{f}} = \{\nu \in \mathbb{M}; \forall x \in \mathbb{D}, \nu_x \leq \Gamma\} \tag{8.27}$$

then

$$p^{\mathrm{tb}}(o,\cdot) = \mathrm{Erl}(\rho.,\Gamma)$$

where Z *is a Poisson random variable with mean* $\rho(\mathbb{D})$*. If, for each* $x \in \mathbb{D}$*, we take* $\lambda_{x,x_k} = \lambda$ *for some constant* λ *and some neighbours* $x_1,\ldots,x_n$ *of* x*, then*

$$d^{\mathrm{tb}} = \mathrm{Erl}(\rho.,\Gamma)\frac{\lambda}{\lambda(\cdot,o)}n \tag{8.28}$$

In this case, the fictitious-*cut probability equals*

$$\begin{aligned} c^{\mathrm{tb}} &= \frac{d^{\mathrm{tb}}}{1+d^{\mathrm{tb}}} \\ &= 1-\frac{1}{1+d^{\mathrm{tb}}} = 1-\frac{1}{1+\mathrm{Erl}(\rho.,\Gamma)\frac{\lambda}{\lambda(\cdot,o)}n} \end{aligned}$$

Part III

Performance evaluation

Chapter 9

Introduction

The present part focuses on third objective described in Section 0.2, i.e. to apply the results of the previous two parts to the performance evaluation of wireless cellular networks.

Our motivation is to build explicit expressions for quality of service indicators in large cellular systems. A particular effort is made to go further than pure simulations, i.e. to build explicit expressions. This is a crucial issue because explicit expressions are more effective than simulations for the optimization of the capacity of the network and for the study the interactions between the link level and the system level.

Wireless networks carry both streaming and elastic traffic. The service requirements are specific to each type of service. Streaming users require some connection duration with a fixed bit-rate, whereas elastic users have some amount of data to send. Quality of service may be expressed in terms of the blocking and cut probabilities for streaming traffic, and in terms of the expected throughput and delay for elastic traffic.

The results of Chapter 6 are helpfull to prove the regularity and the ergodicity of the process modellling the evolution of the streaming calls: as well as the regularity of the process modellling the evolution of the elastic calls.

We show that the tools developped in the first two parts permit to study systems with different multiple access schemes: CDMA such as UMTS Release 99, TDMA such as HSDPA, and FDMA such as GSM in Chapters 10, 11 and 12 respectively.

We shall use the notation described in Section 2.3 (with some adaptation for HSDPA and GSM).

Chapter 10

UMTS Release 99

10.1 Introduction

We consider a UMTS[1] Release 99 network serving both streaming and elastic traffic. Each streaming user is served by a Dedidated CHannel (DCH). The elastic users in a given cell are served simultaneously by the Downlink Shared CHannel (DSCH).

The present chapter is organized as follows. Section 10.2 considers the case where there are only elastic traffic and gives explicit expressions of the throughput per user. Section 10.3 considers a network carrying streaming traffic only and give explicit approximate expressions of the blocking probability for such services. Mixed scenarios with streaming and elastic traffic are studied in section 10.5. We shall describe the related works and our contribution within each section.

10.2 Elastic traffic on DSCH

We consider in the present section a UMTS network carrying elastic traffic only. Hence each user has some volume of data to transmit (or receive) with a flexible bit-rate. We assume that active users are served simultaneously by the DSCH. Our objective is to evaluate the quality of service perceived by a user for this type in terms of the EXPECTED THROUGHPUT (i.e. the average effective transmission or reception bit-rate) and DELAY (i.e. transfer time).

10.2.1 Related works

The major part of the existing literature treating these issues in CDMA networks make some simplifying assumptions about the interference. In [2, 62] the uplink of a CDMA network is treated with the assumption that the inter-cell interference is proportional to intra-cell interference. An analogous assumption

[1] Universal Mobile Telecommunications System

is made for the downlink in [110]. The authors of [6] make a statistical independence assumption between the inter-cell interference perceived by the different cells.

For elastic traffic, the mean number of users may grow unboundedly in the long run of the system; in which case the system is said to be UNSTABLE. This situation has to be avoided, which means that the stochastic process describing the evolution of the system should be ergodic and that the mean number of users under the limiting distribution should be finite. The literature say in this case that the system is STABLE. (Note that this notion of *stability of a system* is different from the notion of *stability of a generator* of a Markov process defined in §6.1.2.)

By definition a necessary condition for the system to be stable is that the stochastic process describing its evolution is ergodic. The conditions for ergodicity depend on the congestion control algorithm. [28] establishes the ergodicity condition for feasibility based congestion control algorithms in CDMA networks. The author investigates also the macro-diversity effects. Ergodicity issues of TDMA networks are addressed in [24] for a single cell network. The impact of mobility within the cell on ergodicity is studied in [22]. Ergodicity issues are also addressed for queueing networks without the spatial component (cf. for example [9]).

Another concern of congestion control algorithms is to assure FAIRNESS among users. We discuss this issue in Annex 14.B.

10.2.2 Our contribution

We describe the temporal evolution of a large CDMA network serving elastic traffic only by a spatial processor-sharing Markov queuing process, and obtain in this framework the explicit expressions of the expected throughput (and delay) for the congestion control scheme based on the decentralized feasibility condition, denoted FC.

Contrarily to the models in [2, 6], our model takes into account the exact representation of the geometry of inter-cell and intra-cell interferences.

Our model is more restrictive than that considered in [28] where the arrival process is not necessarily Poisson. In [28], an ergodicity condition is established, whereas in the present work we go further by establishing explicit expressions for expected throughput (and delay).

10.2.3 Congestion control algorithms

Recall that the feasibility condition FC may be written in the general form

$$\sum_{m\in u}\varphi\left(m\right)<C,\quad u\in\mathbf{U}$$

where u designates a base station, the notation $m\in u$ means that user m is served by base station u and the formulae for the function $\varphi(\cdot)$ and the parameter C are given by (4.14) and (4.12). Note that the function $\varphi(.)$ may be decomposed as follows

$$\varphi(m) = X_m \xi'_m, \quad m \in u$$

where

- ξ'_m designates the modified SINR $\xi'_m = \xi_m / (1 + \alpha \xi_m)$;
- ξ_m designates the signal-to-interference-and-noise ratio (SINR) threshold;
- α is the orthogonality factor, which we assume constant;
- X_m is some factor characterizing the geometry

Recall that C depends on the case considered: uplink or downlink, with or without power limitation. We assume from now that C is independent of the SINR thresholds $\{\xi_m\}$. (This is true except for the uplink with power limitation. This particular case needs further investigations which may be carried in future studies.)

As in §2.1, we take for DSCH modulation a single representative value of the E_b/N_0 and assume that the set of possible rates is *continuous*: $\mathbb{R}_+$. Then the bit-rate r_m and the SINR ξ_m are related by Equation (2.1) which may be written as follows

$$r_m = \frac{W'}{E_b/N_0} \xi_m \tag{10.1}$$

where E_b/N_0 designates the energy-per-bit to noise-density ratio and W' designates the chip-rate.

If the bit-rates $\{r_m; m \in u\}$ satisfy

$$\sum_{m \in u} \gamma_m r_m = 1 \tag{10.2}$$

where the weighting coefficients γ_m are given by

$$\gamma_m = \frac{E_b/N_0}{CW'} X_m \tag{10.3}$$

then

$$\sum_{m \in u} X_m \xi_m = C \tag{10.4}$$

which implies the feasibility condition (4.13), since $\xi'_m < \xi_m$. (There will be a loss of capacity, but this loss is small if $\xi'_m \simeq \xi_m$;or equivalently $\alpha \xi_m \ll 1$.)

Egalitarian congestion control

In [16, 13] an egalitarian scheme where all users in a given cell are given the same bit-rate is proposed. In order to satisfy (10.2), the allocated bit-rates are given by

$$r_m = \frac{1}{\sum_{m \in u} \gamma_m}, \quad m \in u \tag{10.5}$$

Proposed congestion control

The performance of the egalitarian congestion algorithm (10.5) are very difficult to calculate and it is probably hopeless to obtain analytical results. In the present section we consider the same congestion control algorithm as that considered in [28]. This algorithm allocates bit-rates to users as follows

$$r_m = \frac{1}{M\gamma_m}, \quad m \in u \tag{10.6}$$

where M designates the number of users in progress in cell u. (We shall compare different bit-rate allocations in Annex 14.B.)

Remark 25 *Note that we consider the case with no admission control, where an increase of the number of users in a cell is just coped with via a reduction of the bit rates of the users of this cell. Our case is more like TCP where the increase of the number of competitors eventually results in a decreased bit rate for all, and where no user is ever rejected. We model the user bit-rate as a fluid whose rate adjusts immediately in response to changes in the number and positions of users in progress.*

10.2.4 Associated SMQ process

Traffic model

New elastic users arrive to the network as a Poisson process with intensity $\lambda_x \times dx$ in any region of surface dx. The required volumes (amount of data to transmit or receive) are i.i.d. exponentially distributed with mean μ_x^{-1} and independent from arrivals. We may have different SERVICE CLASSES, such as http, ftp, etc., with specific values of the arrival and volume parameters for each service class (arrival and volume processes for different service classes are assumed independent). In this case the parameter x designates both the geographic position and the service class. We consider a set $\mathbb{D}$ designating the possible values of the service class and position of users. Hence a value $x \in \mathbb{D}$, called a LOCATION, corresponds to a given *service class* and a given *geographic position*. We assume that

$$\lambda_x > 0, \mu_x > 0, \quad \forall x \in \mathbb{D}$$

We assume that the users don't move from one cell to another during their calls (mobility may eventually occur within each cell). Since moreover our congestion control algorithm is decentralized, we may study each cell independently from the other cells. Hence the location set $\mathbb{D}$ corresponds to a given cell.

Generator

We assume that the users in a location $x \in \mathbb{D}$ are assigned the same bitrate. We model the evolution of the system by a continuous-time Markov process $\{N_t; t \geq 0\}$ where N_t is a finite counting measures on $\mathbb{D}$ such that $N_t(A)$

designates the number of users in location $x \in A$ at time t. The state space, denoted $\mathbb{M}$, is the space of all finite counting measures on $\mathbb{D}$ and a typical element of $\mathbb{M}$ is denoted ν.

Equation (10.2) may be written as follows

$$\int_{\mathbb{D}} \gamma(x)\, r_x \nu(dx) = 1 \tag{10.7}$$

where $\nu \in \mathbb{M}$ is the state of the system, i.e. $N_t = \nu$; $\gamma(x)$ are some coefficients pondering the capacity consumption of the different locations; and r_x is the bit-rate allocated to the users of class x. The bit-rate allocation (10.6) takes the following form

$$r_x = \nu(\mathbb{D})^{-1} \gamma(x)^{-1} \tag{10.8}$$

We shall now determine the *generator*, say q, of the Markov process $\{N_t; t \geq 0\}$. The birth rate at dx equals $\lambda_x dx$ where λ_x is the intensity of arrival of users of class x. From Equation (10.8) we deduce that the death/service rate at $dx \in \mathbb{D}$ equals $\mu_x r_x \nu(dx) = \mu_x \gamma(x)^{-1} \nu(dx)\, \nu(\mathbb{D})^{-1}$ where μ_x^{-1} designates the mean volume that users of class x require to transmit.

From Proposition 43, we deduce that the generator q of the Markov process $\{N_t; t \geq 0\}$ may be viewed as a SMQ with birth rates and death rates given respectively by

$$\lambda(o, dx) = \lambda_x dx, \quad \lambda(x, \{o\}) = \mu_x$$

motion rates, say $\lambda(x, A), x \in \mathbb{D}, A \in \mathcal{D}$, given in section 6.5, and service rates

$$r_{xy}(\nu) = \begin{cases} 1 & \text{if } x \in \bar{\mathbb{D}}, y \in \mathbb{D} \\ \gamma(x)^{-1} \nu(\mathbb{D})^{-1} & \text{if } y = o, x \in \mathbb{D}; \nu_x > 0 \end{cases}$$

By Lemma 12, it is equivalent to consider another SMQ with routing rates

$$\lambda'(o, dx) = \lambda_x dx, \quad \lambda'(x, \{o\}) = \mu_x \gamma(x)^{-1} \tag{10.9}$$

motion rates $\lambda'(x, A) = \lambda(x, A), x \in \mathbb{D}, A \in \mathcal{D}$; and service rates

$$r'_{xy}(\nu) = \begin{cases} 1 & \text{if } x \in \bar{\mathbb{D}}, y \in \mathbb{D} \\ \nu(\mathbb{D})^{-1} & \text{if } y = o, x \in \mathbb{D}; \nu_x > 0 \end{cases} \tag{10.10}$$

(where we replaced $\lambda(x, \{o\})$ by $\lambda(x, \{o\})\, \gamma(x)^{-1}$ and $r_{xo}(\nu)$ by $r_{xo}(\nu)\, \gamma(x)$). We denote with prime ($'$) the parameters which are specific to this new SMQ. Of course, the generators are identical, $q' = q$.

10.2.5 No mobility case

Here q' is a spatial birth death (SBD) generator (which is a particular case of SMQ studied in Part II).

Regular. The generator q' is stable iff Condition (6.12) is satisfied, which is equivalent to

$$\lambda'(o,\mathbb{D}) = \int_{\mathbb{D}} \lambda_x dx < \infty$$

what will be assumed from now on. Consider the birth rates $\{b'_n\}$ and death rates $\{d'_n\}$ defined by (6.16), i.e.

$$b'_n = \sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q'(\nu, T_{o\mathbb{D}}\nu) = \lambda'(o,\mathbb{D})$$

$$d'_n = \inf_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q(\nu, T_{\mathbb{D}o}\nu)$$

Observe that $q'(\nu, T_{o\mathbb{D}}\nu) = \lambda'(o,\mathbb{D})$ hence Condition (6.17) is satisfied. Moreover we have $\sup_{\nu\in\mathbb{M}} q'(\nu, T_{o\mathbb{D}}\nu) = \lambda'(o,\mathbb{D}) < \infty$, hence by Lemma 4, the generator q' is regular. Let $\{N'_t; t \geq 0\}$ be a Markov jump process associated to the generator q'.

Limiting distribution. Consider the associated M/GI/1 queue as described in Proposition 37. The traffic demand of this M/GI/1 queue is

$$\rho'(\mathbb{D}) = \int_{\mathbb{D}} \frac{\lambda'(o,dx)}{\lambda'(x,o)} = \int_{\mathbb{D}} \lambda_x \gamma(x)\, \mu_x^{-1} dx \tag{10.11}$$

The quantity $\rho'(\mathbb{D})$ is called MODIFIED TRAFFIC DEMAND in order to distinguish it from the TRAFFIC DEMAND of the original system, denoted $\rho(\mathbb{D})$, and given by

$$\rho(\mathbb{D}) = \int_{\mathbb{D}} \frac{\lambda(o,dx)}{\lambda(x,o)} = \int_{\mathbb{D}} \lambda_x \mu_x^{-1} dx$$

From Propositions 37 and 23 we deduce that a sufficient condition for the ergodicity of the Markov process $\{N'_t; t \geq 0\}$ is

$$\rho'(\mathbb{D}) < 1 \tag{10.12}$$

what will be assumed from now on.

Invariant probability measure. The traffic equations (6.47) have solution

$$\rho'\{o\} = 1, \quad \rho'(dx) = \frac{\lambda'(o,dx)}{\lambda'(x,\{o\})} = \frac{\gamma(x)\,\lambda_x}{\mu_x} dx, \quad x \in \mathbb{D} \tag{10.13}$$

The measure ρ' satisfies Condition (6.51). The service rates are balanced by

$$\Psi'(\nu) = e^{\rho'(\mathbb{D})}(1-\rho'(\mathbb{D}))\,\nu(\mathbb{D})!$$

which satisfies Condition (6.52). Let $\Pi_{\Psi'}$ be the Gibbs distribution on $\mathbb{D}$ having density Ψ' with respect to the Poisson process with intensity measure ρ'. Since the routing kernel λ' is reversible with respect to ρ', then, by Proposition 29, $\Pi_{\Psi'}$ is reversible with respect to q. In particular $\Pi_{\Psi'}$ is invariant with respect to the generator q'. By Proposition 25, there is a unique such probability measure which is the limiting distribution.

Analytical expressions

In the case without mobility, the Whittle and wireless models described in Sections 7.1.1 and 7.2.1 respectively are equivalent. Hence we may deduce from (7.21) or (7.42) that the expected delay for users in $A \in \mathcal{D}$, denoted $\bar{T}(A)$, equals

$$\bar{T}(A) = \frac{\rho'(A)}{\lambda(o,A)(1-\rho'(\mathbb{D}))} \tag{10.14}$$

From Equation (7.22) or (7.43) we deduce that the expected throughput for users in $A \in \mathcal{D}$, denoted $\bar{r}(A)$, equals

$$\bar{r}(A) = \frac{\rho(A)}{\rho'(A)}[1-\rho'(\mathbb{D})]$$

where

$$\rho(A) = \int_A \frac{\lambda_x}{\mu_x} dx \tag{10.15}$$

In particular, the expected throughput for users in all the area $\mathbb{D}$ equals

$$\bar{r}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{\rho'(\mathbb{D})}[1-\rho'(\mathbb{D})] \tag{10.16}$$

10.2.6 Infinite mobility case

Consider now the infinite mobility case described in Proposition 45 and recall that we consider the birth rates $\lambda'(o,dx)$ and death rates $\lambda'(x,\{o\})$ given by Equation (10.9). The total number of users $\{N'_t(\mathbb{D})\}$ may be viewed as a M/GI/1 processor-sharing queue with arrival rate $\lambda' = \lambda'(o,\mathbb{D})$ and with call-volume rate $\int_{\mathbb{D}} \lambda'(x,\{o\})\,\varrho(dx)$, i.e. with traffic demand, say ρ', given by

$$\rho' = \frac{\lambda'(o,\mathbb{D})}{\int_{\mathbb{D}} \lambda'(x,\{o\})\,\varrho(dx)} = \frac{\lambda(o,\mathbb{D})}{\int_{\mathbb{D}} \mu_x \gamma(x)^{-1}\,\varrho(dx)} \tag{10.17}$$

where ϱ is the stationary distribution of the location of individual user in $\mathbb{D}$ (i.e. ϱ is solution of Equation (6.61)). The ergodicity condition (7.29) writes

$$\rho' < 1 \tag{10.18}$$

From Lemma 22 we deduce that, in steady state, the mean number of users denoted $\mathbf{E}[\pi']$, equals

$$\mathbf{E}[\pi'] = \frac{\rho'}{1-\rho'} \tag{10.19}$$

the expected delay, denoted $\bar{T}$, equals

$$\bar{T} = \frac{\rho'}{\lambda(o,\mathbb{D})(1-\rho')} \tag{10.20}$$

The above two displays may also be deduced from (7.41) and (7.42) respectively.

From Lemma 22 we deduce also that the expected throughput of the M/GI/1 processor-sharing queue equals $1-\rho'$. This is not the expected throughput of our system, since we have modified the service rates (we transferred the term $\gamma(x)$ from the service rates to the death rates). Observe that the user volume rate of our system is $\int_{\mathbb{D}} \lambda(x,\{o\})\,\varrho(dx)$, then the expected throughput of our system, denoted $\bar{r}$, equals

$$\bar{r} = \frac{1}{\bar{T}\int_{\mathbb{D}} \lambda(x,\{o\})\,\varrho(dx)} = \frac{\rho}{\rho'}(1-\rho') \tag{10.21}$$

where

$$\rho = \frac{\lambda(o,\mathbb{D})}{\int_{\mathbb{D}} \lambda(x,\{o\})\,\varrho(dx)} = \frac{\lambda(o,\mathbb{D})}{\int_{\mathbb{D}} \mu_x \varrho(dx)} \tag{10.22}$$

10.2.7 Numerical results

In this section, we assume that the birth rates λ_x and the death rates μ_x don't depend on x.

Unless otherwise specified, all the numerical applications are made using the default values specified in Section 2.2.5.

No mobility case

The traffic intensities $\rho'(\mathbb{D})$ and $\rho(\mathbb{D})$ are deduced from Equations (10.11) and (10.15)

$$\rho'(\mathbb{D}) = \bar{\gamma}\rho(\mathbb{D}), \quad \rho(\mathbb{D}) = \pi R^2 \lambda \mu^{-1}$$

where $\bar{\gamma}$ is the average of $\gamma(x)$ over the cell $\mathbb{D}$. Hence the ergodicity condition (10.12) may be written as follows

$$\rho(\mathbb{D}) < \bar{\gamma}^{-1}$$

Note that $\rho(\mathbb{D})$ is the traffic demand per cell, hence the ergodicity condition says that the traffic demand per cell should be less than some critical value, denoted $\rho_c(\mathbb{D})$, that is

$$\rho(\mathbb{D}) < \rho_c(\mathbb{D}) \tag{10.23}$$

where the CRITICAL TRAFFIC DEMAND is given by

$$\rho_c(\mathbb{D}) = \bar{\gamma}^{-1} \quad (= \frac{\rho(\mathbb{D})}{\rho'(\mathbb{D})}) \tag{10.24}$$

The expected delay for users in all the area $\mathbb{D}$ is given by Equation (10.14)

$$\bar{T}(\mathbb{D}) = \frac{\bar{\gamma}}{\mu(1-\bar{\gamma}\rho(\mathbb{D}))} \tag{10.25}$$

The expected throughput for users in all the area $\mathbb{D}$ is given by Equation (10.16)

$$\bar{r}(\mathbb{D}) = \bar{\gamma}^{-1} - \rho(\mathbb{D}) \tag{10.26}$$

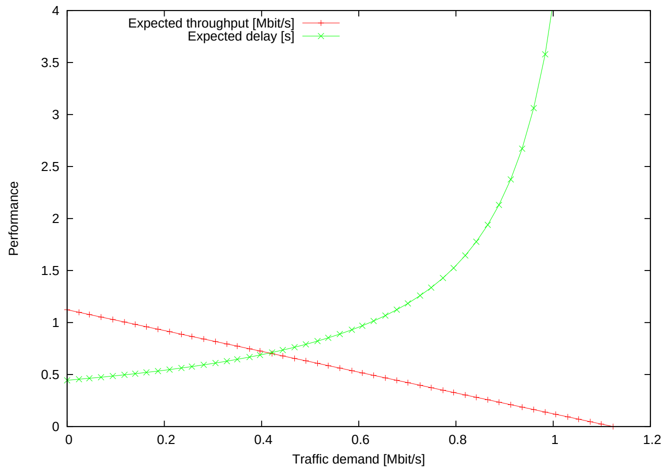

Figure 10.1: Expected throughput and delay for DSCH.

DSCH. Recall that the weighting coefficient $\gamma(x)$ for DSCH modulation is given by (10.3). Then $\bar{\gamma} = \frac{E_b/N_0}{CW'}\bar{X}$. Hence, we get the critical traffic from Equation (10.24)

$$\rho_c(\mathbb{D}) = \frac{1}{\bar{\gamma}} = \frac{CW'}{\bar{X}E_b/N_0} \tag{10.27}$$

The expression of $\bar{X}$ may be deduced from Appendix 13.D. For example for DFC (i.e. the downlink without power limitation) we have

$$\bar{X} = \alpha + \frac{0.94}{\eta - 2}$$

which gives

$$\rho_c(\mathbb{D}) \simeq 1.1\,\text{Mbit}\,/\,\text{s}$$

If the traffic demand $\rho(\mathbb{D})$ exceeds this critical value 1.1 Mbit / s, then the system becomes unstable (cf. (10.23)).

Fix the value of the call-volume average $\mu^{-1} = 0.5\,\text{Mbit}$. Figure 10.1 represents the expected throughput and delay for DSCH.

In particular for a traffic demand $\rho(\mathbb{D}) \simeq 0.5\,\text{Mbit}\,/\,\text{s}$, which corresponds to a mean interarrival duration $\lambda^{-1} = 1\,\text{s}$, we get an expected throughput and delay respectively equal to

$$\bar{r}(\mathbb{D}) = 0.6\,\text{Mbit}\,/\,\text{s}, \quad \bar{T}(\mathbb{D}) = 0.8\,\text{s}$$

We see in Figure 10.1 that the expected throughput decreases and the delay increases when the traffic demand increases. At the limit when the traffic demand $\rho(\mathbb{D})$ tends to the critical value $\rho_c(\mathbb{D})$, the throughput goes to 0 and the delay goes to infinity.

Shannon. The approximation (2.4) for the Shannon's bound may be written as follows

$$r_m \simeq \frac{W}{\ln(2)}\xi_m$$

In this case the expression (10.3) should be replaced by the following one

$$\gamma_m = \frac{\ln(2)}{CW} X_m$$

We get a critical traffic for Shannon's bound $\rho_c(\mathbb{D}) \simeq 6.7\,\mathrm{Mbit}/\mathrm{s}$ (for the downlink without power limitation) which is about six times the critical traffic for DSCH.

Infinite mobility case

The traffic intensities ρ' and ρ are deduced from Equations (10.17) and (10.22)

$$\rho' = \frac{\rho}{\overline{\gamma^{-1}}}, \quad \rho = \pi R^2 \lambda \mu^{-1}$$

where $\overline{\gamma^{-1}}$ is the average of $\gamma^{-1}(x)$ over the cell $\mathbb{D}$. Hence the ergodicity condition (10.18) may be written as follows

$$\rho < \rho_c = \overline{\gamma^{-1}}$$

where ρ_c is called critical traffic.

The expected delay is deduced from Equation (10.20)

$$\bar{T} = \frac{1}{\left(\overline{\gamma^{-1}} - \rho\right)\mu}$$

The expected throughput is given by Equation (10.21)

$$\bar{r} = \overline{\gamma^{-1}} - \rho$$

We compare the expected throughputs for DSCH in the no mobility and the infinite mobility cases in Figure 10.2. We deduce that the mobility increases the expected throughput by a factor of about 1.5.

Intermediate mobility

We recall the ingredients of the generator of our wireless model. The birth rates $\lambda'(o, dx)$ and death rates $\lambda'(o, dx)$ are given by (10.9) and the motion rates $\lambda'(x, A) = \lambda(x, A), x \in \mathbb{D}, A \in \mathcal{D}$ are given in section 6.5. The service rates $r'_{xy}(\nu)$ are given by (10.10).

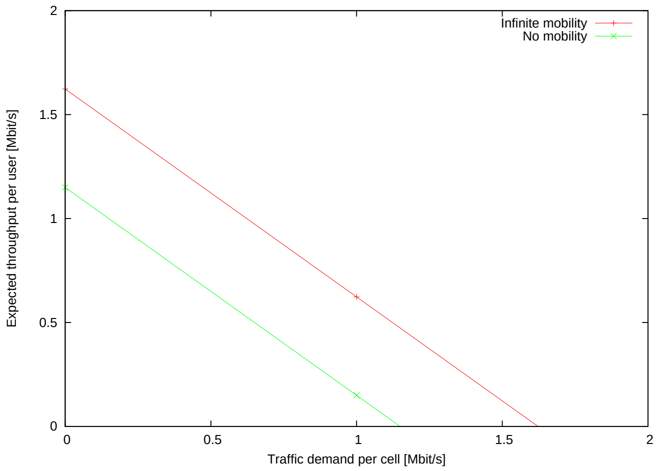

Figure 10.2: Expected throughputs for DSCH in the no mobility and the infinite mobility cases.

Trafic equations. We assume now an average speed υ which is finite and non-null. The cell $\mathbb{D}$ is modeled by a disc of radius R which is divided into J rings. Each ring denoted by $j \in \mathbf{J} = \{1, \ldots, J\}$ is delimited by discs with radii r_{j-1} and r_j where $r_0 = 0$ and $r_J = R$.

Let ρ denote the traffic demand, i.e. $\rho_j = \frac{\lambda_{oj}}{\lambda_{jo}}$ for $j \in \mathbf{J}$, and let ρ' be the solution of the traffic equations, which is function of ρ and of the speed υ denoted $\rho'(\rho, \upsilon)$. We call $\rho'(\rho, \upsilon)$ the MODIFIED TRAFFIC.

For a given speed υ, if we multiply the traffic demands by $a \in \mathbb{R}_+$, i.e. ρ_j is replaced by $a\rho_j$ for $j \in \mathbf{J}$, then ρ' is multiplied by the same factor, i.e.

$$\rho'(a\rho, \upsilon) = a\rho'(\rho, \upsilon)$$

Proof. Multiplying the original the traffic equations by a, gives

$$\begin{cases} a\rho'_j\left(\lambda_{jo} + \lambda_{j(j-1)} + \lambda_{j(j+1)}\right) = a\lambda_{oj} + a\rho'_{j-1}\lambda_{(j-1)j} + a\rho'_{j+1}\lambda_{(j+1)j} & \text{for } j = 2, \ldots, J-1 \\ a\rho'_1\left(\lambda_{1o} + \lambda_{12}\right) = a\lambda_{o1} + a\rho'_2\lambda_{21} & \\ a\rho'_J\left(\lambda_{Jo} + \lambda_{J(J-1)}\right) = a\lambda_{oJ} + a\rho'_{J-1}\lambda_{(J-1)J} & \end{cases}$$

which shows the desired result. ■

Hence

$$\rho'(\rho, \upsilon) = \rho_{\mathbb{D}} \times \rho'(\rho/\rho_{\mathbb{D}}, \upsilon)$$

and in particular

$$\rho'_{\mathbb{D}}(\rho, \upsilon) = \rho_{\mathbb{D}} \times \rho'_{\mathbb{D}}(\rho/\rho_{\mathbb{D}}, \upsilon) \tag{10.28}$$

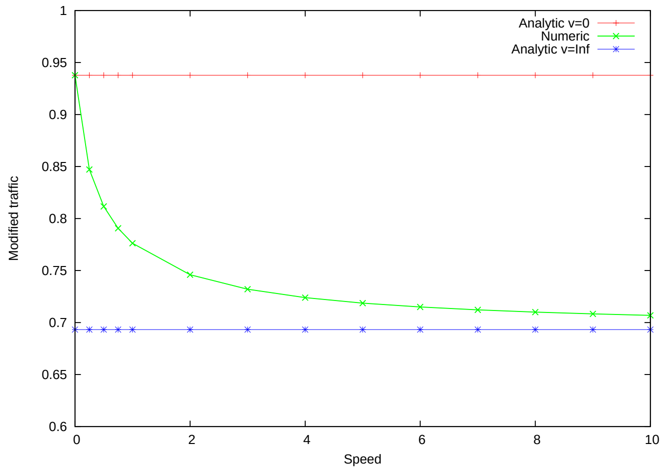

Figure 10.3: Monotonicity of the modified traffic with speed.

Whittle model. We construct a Whittle model by replacing *artificially* the service rates (10.10) of the wireless model by

$$r'_{xy}(\nu) = \begin{cases} \nu(\mathbb{D})^{-1} & \text{if } x = o, y \in \mathbb{D} \\ \nu(\mathbb{D})^{-1} & \text{if } y = o, x \in \mathbb{D}; \nu_x > 0 \end{cases}$$

and keeping the same routing rates $\lambda'(x, dy)$. In particular, the traffic equations are the same for both the wireless and the Whittle models.

We know that the wireless and Whittle models are identical at null and infinite speed υ. We will compare numerically the results of the two models for non-null finite speeds.

Monotonicity with speed. In the simulations we take the call-volume average $\mu^{-1} =$ 1 Mbit. For other values of the call-volume average μ, we can always retreive this particular case, since we may replace the generator q by q/μ without altering the ergodicity and invariant distribution. In doing so, we should replace the speed υ by υ/μ. In other words we may view our speed as expressed in kilometers per $\mu\times$ seconds.

Figure 10.3 represents the modified traffic on the cell $\rho'_{\mathbb{D}}(\rho/\rho_{\mathbb{D}}, \upsilon)$ as function of the speed. This figure shows that the modified traffic per cell is a decreasing function of the speed. (Nevertheless, we observe numerically that the modified traffic at a given ring $\rho'_j(\rho/\rho_{\mathbb{D}}, \upsilon)$ is not always monotonous with speed.) From this numerical observation, we propose the following conjectures:

Conjecture 1 *The modified traffic per cell decreases with the speed.*

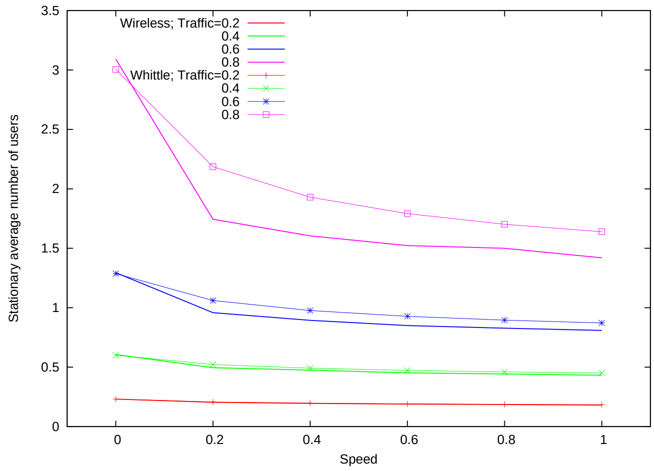

Figure 10.4: Monotonicity of the mean number of users in a cell with speed.

Let $\bar{N}_{\mathbb{D}}$ be the mean number of users in a cell. Since in the Whittle model, we have

$$\bar{N}_{\mathbb{D}} = \frac{\rho'_{\mathbb{D}}}{1-\rho'_{\mathbb{D}}} \tag{10.29}$$

we deduce that when the speed increases, the mean number of users per cell decreases. (Therefore the delay decreases and the throughput increases.)

Figure 10.4 represents the mean number of users per cell as function of the average speed υ for different values of the traffic demand $\rho_{\mathbb{D}} = 0.2, 0.4, 0.6, 0.8\,\text{Mbit}\,/\,\text{s}$. We observe numerically that the mean number of users per cell of the Whittle model is an upper bound of that of the wireless model; and that this upper bound is tight for small traffic demand.

From the above numerical observations, we propose the following conjecture:

Conjecture 2 *The mean number of users per cell of the Whittle model is an upper bound of that of the wireless model. This upper bound is tight for small traffic demand.*

Stability. The stability condition for the Whittle model is

$$\rho'_{\mathbb{D}}(\rho, \upsilon) < 1$$

or equivalently

$$\rho_{\mathbb{D}} < 1/\rho'_{\mathbb{D}}(\rho/\rho_{\mathbb{D}}, \upsilon)$$

The right-hand side of the above inequality is called CRITICAL TRAFFIC DEMAND.

Figure 10.5 represents the critical traffic demand as function of the speed.

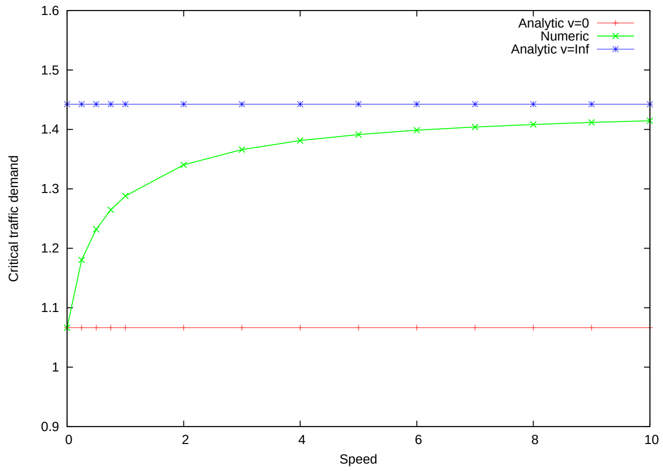

Figure 10.5: Critical traffic demand

We investigate numerically the stability of the wireless model. Figure 10.6 represents the mean number of users per cell as function of the traffic demand for different values of the average speed $\upsilon = 0, 0.2, 0.4, 0.6, 0.8, 1$. From this numerical result, we propose the following conjecture:

Conjecture 3 *The stability condition of the wireless model is*

$$\begin{cases} \rho'_{\mathbb{D}}(\rho, 0) < 1 & \textit{for } \upsilon = 0 \\ \rho'_{\mathbb{D}}(\rho, \infty) < 1 & \textit{for } \upsilon > 0 \end{cases}$$

10.3 Streaming traffic on DCH without mobility

We consider in the present section a UMTS network carrying streaming traffic only. Hence each user requires some transmission duration at fixed bit-rate and is served by a specific DCH. Our objective is to evaluate the quality of service perceived by a user for this type in terms of the blocking probability.

10.3.1 Related works

We make here a short survey of the literature on performance evaluation of load control schemes for CDMA networks. The QoS indicators introduced for semi-static models in [52, 121, 84, 47] correspond to the probability that the SINR is less than some threshold, when users, modeled as a spatial Poisson point process, are all accepted. In [84] and [47] this indicator is called the *outage probability*. The authors of [121] call it the blocking probability, but as mentioned in [84], the term outage probability is more appropriate. We

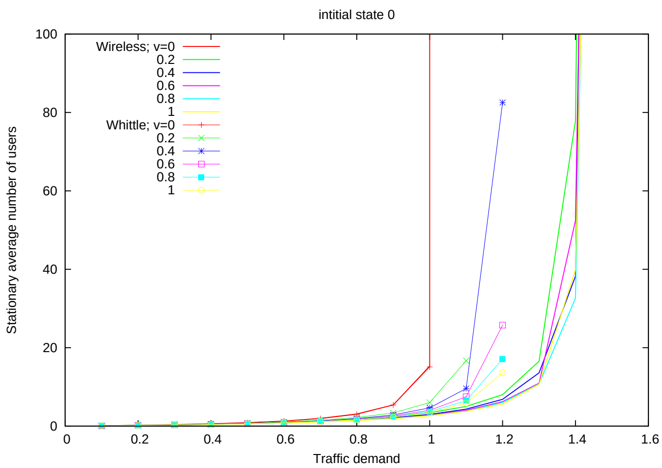

Figure 10.6: Numerical investigation of the stability of the wireless model.

propose to make the following distinction between OUTAGE and INFEASIBILITY probabilities, both being defined for a semi-static model: the former is related to the event that the transmission quality of service is not attained for *given transmission powers*, whereas the latter corresponds to the situation when *there is no solution to the power control problem.* Hence both the outage and the infeasibility probabilities are defined for a semi-static model and are related to "the probability that the transmission quality of service is not attained". But the outage probability depends on the transmission powers of the users and the base stations; whereas the infeasibility probability corresponds to an intrinsic characterization of power allocation feasibility, and consequently doesn't depend on transmission powers. The infeasibility probability is then a more appropriate performance indicator.

The authors of [84] define the *blocking probability* in a semi-dynamic model and give simulation results, which show that the outage and blocking probabilities are different in general. In [46] it is argued that "the outage probability may easily be computed whereas the blocking probability, even in the particular case where a product-form is obtained, requires methods such as Monte-Carlo acceptance-rejection technique or approximation techniques such as Erlang fixed point."

In analytical studies of the blocking probabilities in CDMA networks, the geometry of interferences specific to CDMA is often absent or seriously reduced. These studies make the distinction between blocking of new calls and of handoff calls. Examples of such studies are [100, 61], which consider a single cell and [89, 111], which consider a multi-cell scenario. In [100, 61], blocking probabilities are calculated via the classical Erlang formula. In [111] explicit expressions of blocking probabilities are given for two limiting regimes of the dynamic model:

no mobility and infinite mobility. In [89], Erlang fixed point approximations are used to calculate blocking probabilities.

10.3.2 Our contribution

We apply the spatial Erlang formula established in §8.2 to built approximate explicit expressions of the blocking probability of the feasibility based load control algorithms. It is, to our knowledge, the first time that an explicit expression of the blocking probability, taking into account in an accurate manner the interference in large UMTS networks, is established. Such an expression is very useful for operators since it opens the way to efficient and rapid capacity, dimensioning and cost evaluation methods.

10.3.3 Associated SMQ process

Traffic model

The definition of the location set $\mathbb{D}$ is similar to §10.2.4, except that now we consider streaming traffic.

New streaming users arrive to the network as a Poisson process with intensity $\lambda_x \times dx$ in any region of surface dx. The required transmission durations are i.i.d. exponentially distributed with mean μ_x^{-1} and independent from arrivals. We may have different SERVICE CLASSES, such as voice, streaming video, etc., with specific values of the arrival and duration parameters for each service class (arrival and duration processes for different service classes are assumed independent). In this case the parameter x designates both the geographic position and the service class. We consider a set $\mathbb{D}$ designating the possible values of the service class and position of users. Hence a value $x \in \mathbb{D}$, called a LOCATION, corresponds to a given *service class* and a given *geographic position.* We assume that

$$\lambda_x > 0, \mu_x > 0, \quad \forall x \in \mathbb{D}$$

We assume that the users don't move during their calls. Since moreover our admission control algorithm is decentralized, we may study each cell independently from the other cells. Hence the location set $\mathbb{D}$ corresponds to a given cell.

Generator

We model the evolution of the system by a continuous-time Markov process $\{N_t; t \geq 0\}$ where N_t is a finite counting measures on $\mathbb{D}$ such that $N_t(A)$ designates the number of users in location $x \in A$ at time t. The state space, denoted $\mathbb{M}$, is the space of all finite counting measures on $\mathbb{D}$ and a typical element of $\mathbb{M}$ is denoted ν.

We consider two types of systems as described below.

- Free system. In this case the capacity of each cell is supposed to be infinite.

- Loss system. We assume the following admission control:
  - We consider the FC admission criterion (4.13) which may be written as follows
  $$\int_{\mathbb{D}} \varphi(x) N_t(dx) < C$$
  or
  $$\int_{\mathbb{D}} \varphi(x) \nu(dx) < C \tag{10.30}$$
  where C is a given constant, $\varphi : \mathbb{D} \to \mathbb{R}_+$ is a given measurable function and $N_t = \nu$ is the current system state.
  - For a new user arrival, we account for the new user in the left-hand side of (10.30). If the inequality (10.30) is respected, the user is admitted, otherwise the user is blocked, i.e. definitely lost (i.e. the system remains in its previous state).

Free system

We shall now determine the *generator*, say q, of the free system. The birth rate at dx equals $\lambda_x dx$ where λ_x is the intensity of arrival of users of class x. The death/service rate at $dx \in \mathbb{D}$ equals $\mu_x \nu(dx)$ where μ_x^{-1} designates the mean required duration for users of class x.

Hence the generator q of the free system may be viewed as a SMQ with birth rates and death rates given respectively by

$$\lambda(o, dx) = \lambda_x dx, \quad \lambda(x, \{o\}) = \mu_x$$

motion rates $\lambda(x, A) = 0$, for all $x \in \mathbb{D}, A \in \mathcal{D}$, and service rates $r_{xy}(\nu) = 1$, for all $x, y \in \bar{\mathbb{D}}, \nu \in \mathbb{M}$. In fact, our process is a SBD.

Regular. The generator q is stable iff Condition (6.12) is satisfied, which is equivalent to

$$\lambda(o, \mathbb{D}) = \int_{\mathbb{D}} \lambda_x dx < \infty$$

what will be assumed from now on. Consider the birth rates $\{b_n\}$ and death rates $\{d_n\}$ defined by (6.16), i.e.

$$b_n = \sup_{\nu \in \mathbb{M}: \nu(\mathbb{D}) = n} q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o, \mathbb{D})$$

$$d_n = \inf_{\nu \in \mathbb{M}: \nu(\mathbb{D}) = n} q(\nu, T_{\mathbb{D}o}\nu)$$

Observe that $q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o, \mathbb{D})$ hence Condition (6.17) is satisfied. Moreover we have $\sup_{\nu \in \mathbb{M}} q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o, \mathbb{D}) < \infty$, hence by Lemma 4, q is regular. Let $\{N_t; t \geq 0\}$ be a Markov jump process associated to the generator q.

Limiting distribution. Observe that

$$d_n = \inf_{\nu \in \mathbb{M}:\nu(\mathbb{D})=n} \sum_{x\in\nu} \nu_x \lambda(x,\{o\}) \geq n \inf_{x\in\mathbb{D}} \mu_x$$

If

$$\inf_{x\in\mathbb{D}} \mu_x > 0$$

what will be assumed from now on, then Condition (6.33) is satisfied which implies by Lemma 6 and Proposition 23 that the process $\{N_t; t \geq 0\}$ is ergodic.

Invariant probability measure. The traffic equations (6.47) have solution

$$\rho\{o\} = 1, \quad \rho(dx) = \frac{\lambda(o,dx)}{\lambda(x,\{o\})} = \frac{\lambda_x}{\mu_x}dx, \quad x \in \mathbb{D}$$

The measure ρ satisfies Condition (6.51). The service rates are balanced by $\Psi \equiv 1$, hence (6.52) is satisfied. Let Π_ρ be the distribution of the Poisson process on $\mathbb{D}$ with intensity measure ρ. As the routing kernel λ is reversible with respect to ρ, then, by Proposition 29, Π_ρ is reversible with respect to q. In particular Π_ρ is invariant with respect to q. By Proposition 25, there is a unique such probability measure which is the limiting distribution.

Loss system

Let $\mathbb{M}^{\mathrm{f}}$ be the feasibility set corresponding to the admission criterion (10.30)

$$\mathbb{M}^{\mathrm{f}} = \left\{\nu \in \mathbb{M} : \int_{\mathbb{D}} \varphi d\nu < C\right\}$$

Consider the truncated generator q^{tb} on $\mathbb{M}^{\mathrm{f}}$ given by (8.3). The truncated generator q^{tb} is also a SMQ process with the service rates deduced from Equation (8.4)

$$r^{\mathrm{tb}}_{xy}(\nu) = 1\{\nu \in \mathbb{M}^{\mathrm{f}}, T_{xy}\nu \in \mathbb{M}^{\mathrm{f}}\} + 1\left\{\nu \notin \mathbb{M}^{\mathrm{f}}\right\}$$

Regular. Note that $q^{\mathrm{tb}}(\nu) \leq q(\nu)$; then the generator q^{tb} is stable. Observe that $q^{\mathrm{tb}}(\nu, T_{o\mathbb{D}}\nu) \leq q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o,\mathbb{D})$, hence Condition (6.17) is satisfied. Moreover we have $\sup_{\nu\in\mathbb{M}} q^{\mathrm{tb}}(\nu, T_{o\mathbb{D}}\nu) \leq \lambda(o,\mathbb{D}) < \infty$, hence by Lemma 4, q^{tb} is regular. Let $\left\{N^{\mathrm{tb}}_t; t \geq 0\right\}$ be a Markov jump process associated to the generator q^{tb}.

Limiting distribution. Note that $\mathbb{M}^{\mathrm{f}}$ is closed with respect to transition $T_{xo}\nu$ for all $x \in \mathbb{D}$ (we say that $\mathbb{M}^{\mathrm{f}}$ is stable by the deaths). Consider the birth rates $\{b^{\mathrm{tb}}_n\}$ and death rates $\{d^{\mathrm{tb}}_n\}$ defined by (6.16), i.e.

$$b^{\mathrm{tb}}_n = \sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q^{\mathrm{tb}}(\nu, T_{o\mathbb{D}}\nu)$$

$$d_n^{\mathrm{tb}} = \inf_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q^{\mathrm{tb}}\left(\nu, T_{\mathbb{D}o}\nu\right)$$

Observe that $\forall\nu \in \mathbb{M}, q^{\mathrm{tb}}\left(\nu, T_{o\mathbb{D}}\nu\right) \leq q\left(\nu, T_{o\mathbb{D}}\nu\right)$ then $b_n^{\mathrm{tb}} \leq b_n$. On the other hand

$$\begin{aligned} d_n^{\mathrm{tb}} &\geq \inf_{\nu\in\mathbb{M}^{\mathrm{f}}:\nu(\mathbb{D})=n} q^{\mathrm{tb}}\left(\nu, T_{\mathbb{D}o}\nu\right) = \inf_{\nu\in\mathbb{M}^{\mathrm{f}}:\nu(\mathbb{D})=n} q\left(\nu, T_{\mathbb{D}o}\nu \cap \mathbb{M}^{\mathrm{f}}\right) \text{ (by (8.3))} \\ &= \inf_{\nu\in\mathbb{M}^{\mathrm{f}}:\nu(\mathbb{D})=n} q\left(\nu, T_{\mathbb{D}o}\nu\right) \text{ (by the stability of } \mathbb{M}^{\mathrm{f}} \text{ by the deaths)} \\ &\geq d_n \end{aligned}$$

Hence $b_n^{\mathrm{tb}} \leq b_n$ and $d_n^{\mathrm{tb}} \geq d_n$. This implies by Lemma 6 and Proposition 23 that the process $\left\{N_t^{\mathrm{tb}}; t \geq 0\right\}$ is ergodic.

Invariant probability measure. Let Π_ρ^{tb} be the *truncation* of Π_ρ to $\mathbb{M}^{\mathrm{f}}$. As Π_ρ is reversible with respect to q, then Π_ρ^{tb} is reversible with respect to q^{tb}. In particular Π_ρ^{tb} is invariant with respect to q^{tb}. By Proposition 25, there is a unique such probability measure which is the limiting distribution.

Blocking probability. The following lemma proves a condition needed to apply Lemma 15 when calculating the blocking probability.

Lemma 17 *We have*

$$\int_{\mathbb{M}} q(\nu)\,\Pi_\rho(d\nu) < \infty \textit{ and } \int_{\mathbb{M}^{\mathrm{f}}} q^{\mathrm{tb}}(\nu)\,\Pi_\rho^{\mathrm{tb}}(d\nu) < \infty$$

Proof. Note first that $q^{\mathrm{tb}}(\nu) = q\left(\nu, \mathbb{M}^{\mathrm{f}}\right) \leq q(\nu)$ then, by the truncation property, it is enough to show the first inequality. We have

$$\begin{aligned} \int q\,d\Pi_\rho &= \mathbf{E}_{\Pi_\rho}\left[q\left(\nu, T_{o\mathbb{D}}\nu\right)\right] + \mathbf{E}_{\Pi_\rho}\left[q\left(\nu, T_{\mathbb{D}o}\nu\right)\right] \\ &= \lambda(o, \mathbb{D}) + \mathbf{E}_{\Pi_\rho}\left[\int_{\mathbb{D}} \lambda(x, \{o\})\,\nu(dx)\right] \\ &= \lambda(o, \mathbb{D}) + \int_{\mathbb{D}} \mathbf{E}_{\Pi_\rho}\left[\nu(dx)\right]\lambda(x, \{o\}) \\ &= \lambda(o, \mathbb{D}) + \int_{\mathbb{D}} \rho(dx)\,\lambda(x, \{o\}) = 2\times\lambda(o, \mathbb{D}) \end{aligned}$$

which is finite. ∎

From the above lemma and Corollary 10, we deduce that the blocking rate is given by

$$p^{\mathrm{tb}}(o, x) = 1 - \frac{\Pi_\rho\left(\mathbb{M}_x^{\mathrm{f}}\right)}{\Pi_\rho\left(\mathbb{M}^{\mathrm{f}}\right)} \tag{10.31}$$

where

$$\mathbb{M}_x^{\mathrm{f}} = \left\{\nu \in \mathbb{M} : \int_{\mathbb{D}} \varphi\,d\nu < C - \varphi(x)\right\}$$

Note that the infeasibility probability equals $P_i = 1 - \Pi_\rho\left(\mathbb{M}^{\mathrm{f}}\right)$. Note also that the set $\mathbb{M}^{\mathrm{f}}$ defined in the above display is analogous to $\mathbb{M}^{\mathrm{f}}$; the only difference is in the constant on the right-hand side of the inequality defining the set (C in $\mathbb{M}^{\mathrm{f}}$ is replaced by $C - \varphi(x)$ in $\mathbb{M}^{\mathrm{f}}$). Then Equation (10.31) relates the blocking rate and the infeasibility probability.

We deduce from Equation (10.31) that the blocking rate at some location x may be related formally to the infeasibility probability as follows

$$p^{\mathrm{tb}}(o,x) = 1 - \frac{1 - P_i\,[\text{with new call at } x]}{1 - P_i\,[\text{without new call}]}$$

10.3.4 Calculation methods

Note that the parameter C is constant for DFC, UFC and EDFC. Considering the version of EUFC described in Remark 4 assures that the parameter C is also constant for EUFC. Here are some methods to calculate the infeasibility probability:

Dynamic simulation

We simulate the Markov process and compute the blocking probability as the ratio of the number of blocked arrivals to the total number of arrivals in a time interval sufficiently large.

Gaussian approximation

From the Gaussian approximation of the infeasibility probability given by Equation (4.17) we deduce the following approximations

$$\Pi_\rho\left(\mathbb{M}^{\mathrm{f}}\right) = 1 - P_i \simeq 1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]$$

$$\Pi_\rho\left(\mathbb{M}^{\mathrm{f}}_x\right) \simeq 1 - Q\left[\left(C - \varphi(x) - \bar{S}\right)/\sigma_S\right]$$

where $\bar{S}$ and σ_S are the expectation and the standard deviation of the random variable S appearing in the left hand-side of the feasibility condition (4.13). Hence

$$p^{\mathrm{tb}}(o,x) \simeq 1 - \frac{1 - Q\left[\left(C - \varphi(x) - \bar{S}\right)/\sigma_S\right]}{1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]}$$

In order to gain insight on the variations of the blocking rate with the user's location, we make the following approximation

$$\begin{aligned} p^{\mathrm{tb}}(o,x) &\simeq \frac{Q\left[\left(C - \varphi(x) - \bar{S}\right)/\sigma_S\right] - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]}{1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]} \\ &\simeq \varphi(x)\,\frac{\exp\left[-\left(\left(C - \bar{S}\right)/\sigma_S\right)^2/2\right]}{\sigma_S\sqrt{2\pi}\left\{1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]\right\}} \end{aligned} \tag{10.32}$$

The blocking probability in the cell, say $\bar{b}$, is given by

$$\bar{b} = \frac{1}{|\mathbb{D}|} \int_{\mathbb{D}} p^{\mathrm{tb}}(o, x) dx$$

where $\mathbb{D}$ and $|\mathbb{D}|$ designate the cell its area respectively. In order to make faster the blocking probability calculation, we make the following approximation

$$\begin{aligned} \bar{b} &\simeq 1 - \frac{1 - Q\left[\left(C - \bar{\varphi} - \bar{S}\right)/\sigma_S\right]}{1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]} \\ &\simeq \bar{\varphi} \frac{\exp\left[-\left(\left(C - \bar{S}\right)/\sigma_S\right)^2/2\right]}{\sigma_S\sqrt{2\pi}\left\{1 - Q\left[\left(C - \bar{S}\right)/\sigma_S\right]\right\}} \end{aligned}$$

Erlang approximation

From Lemma 2, we deduce that the blocking probability may be approximated by the classical Erlang formula with traffic demand $\bar{M}$ for a queue with $\Gamma = C/\bar{\varphi}$ servers. (Note that Γ is the pole capacity defined by (4.4) for the downlink and by (4.10) for the uplink.)

Kauffman-Roberts algorithm

We divide the cell into a finite number of rings and use Kauffman-Roberts algorithm [73, 104] (described in Algorithm 1).

10.3.5 Numerical results

Unless otherwise specified, all the numerical applications are made using the default values specified in Section 2.2.5.

Figure 10.7 represents the blocking probability obtained with dynamic simulation, Gaussian approximation, Erlang approximation and Kauffman-Roberts algorithm for EDFC respectively. Visual inspection of this figure shows that the Kauffman-Roberts algorithm performs very well whereas the gaussian and Erlang approximations performs well only for small blocking probabilities (typically smaller than 0.3).

10.4 Streaming traffic on DCH with mobility

We consider in the present section a UMTS network carrying streaming traffic only in a dynamic context (i.e. users may move during their calls). Our objective is to build expressions for blocking and cut probabilities. The term BLOCKING designates the NON-ADMISSION of new arriving jobs due to the capacity saturation by the active ones. The term CUT designates the INTERRUPTION of active jobs due to capacity saturation induced by their mobility. Hence we define two quality of service indicators, the so called *blocking probability* and *cut*

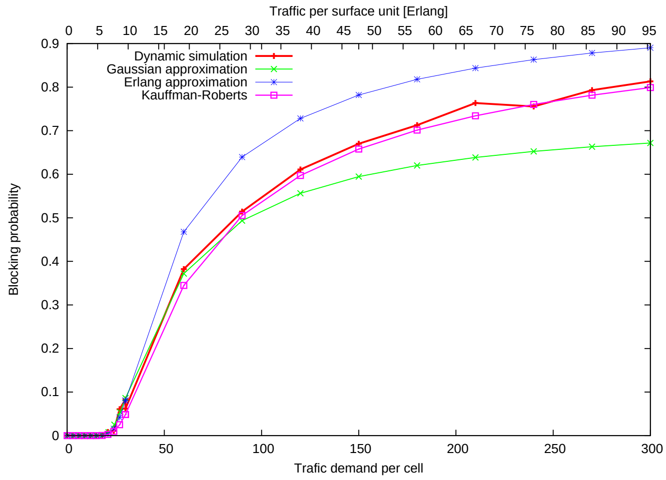

Figure 10.7: EDFC, Cell radius 1Km. (Similar results are obtained for other cell radii and for DFC,UFC and EUFC.)

probability. The BLOCKING PROBABILITY designates the fraction of calls that are not admitted, whereas the CUT PROBABILITY designates the fraction of calls that are interrupted in the long run by the system. The blocking of arriving jobs is a well known problem in fixed networks whereas the cut of active jobs is a new problem induced by mobility which is specific to cellular networks.

10.4.1 Related works

A large part of the existing literature on performance of cellular networks considers either a semi-static traffic model [52, 121] or a semi-dynamic traffic model [84, 47, 2]. In such context the mobility of users during their calls is not taken into account; and hence the cut probability, which is the central issue in the present study, may not be defined.

The literature considering a dynamic model where cut probability may be defined and going further than pure simulations makes some simplifying assumption about the interference. The most common assumption is that inter-cell interference is proportional to intra-cell interference. This leads to models where only the number of jobs per cell (and not their geographic positions) is relevant. Examples of such studies are [111, 89]. In [111] two QoS indicators are defined: new call blocking probability and handoff blocking probability. Explicit expressions for these indicators are given for two limiting regimes: no mobility and infinite mobility. (Note that the infinite mobility regime is too constraining as the capacity of the whole network reduces in this case to that of a single cell without mobility.) In [89] two QoS indicators are defined: blocking probability and forced termination probability. Erlang fix point approximations are used to calculate these probabilities numerically.

10.4.2 Our contribution

The major difference of our work with the existing literature is that our model takes into account the exact representation of the geometry of inter-cell and intra-cell interferences. In this context, we establish an explicit approximate expression of the cut probability as function of the average speed of the users, which permits to study analytically how does the cut probability varies when speed increases. It is, to our knowledge, the first time that an explicit expression of the cut probability, taking into account in an accurate manner the interference in large cellular networks, is established.

In the present work, we define the *cut* probability as the ratio of the number of calls which are cut to the total number of call terminations. The authors of [89] and [65] calculate the cut probability by considering a mobile moving along the network and assuming that the mobile don't affect the network state. We don't make this assumption in our present work.

The *handoff blocking* probability in [111, 89] is defined as the number of handoffs which are blocked to the total number of handoff attempts. Although the handoff blocking probability in [111, 89] captures the dynamic aspect (and in particular the effect of the customer speeds) it is different from what we call cut probability.

10.4.3 Associated SMQ process

We extend the model described in §10.3 to account for mobility of users during their calls. We make the same assumptions and use the same notation of §10.3 for call arrivals and durations.

We consider a network of hexagonal cells where mobility may occur either within each cell or between cells as described in §6.5.4. We consider the FC admission criterion. Due to user mobility between cells, we may not study each cell individually. We use the same notation of §6.5.4 for the network description and user mobility. In particular, the location space is denoted $\mathbb{D} = \mathbf{U} \times \mathbf{J}$ where $\mathbf{U}$ is the set of cells and $\mathbf{J}$ is the set of rings which each cell. If there are multiple service classes, then $\mathbb{D}$ is replaced by $\mathbb{D}\times$ "Set of classes". Hence a value $x \in \mathbb{D}$ corresponds to a given *service class* and a given *geographic location*. For brevity, we will say SERVICE CLASS x for the service class associated to x and GEGRAPHIC LOCATION x for the geographic location associated to x.

As usual, the state of the system is described by a process $N = \{N_t\}_{t\geq 0}$ with state space $\mathbb{M}$ the space of all finite counting measures on $\mathbb{D}$. We consider three types of networks as described below.

- Free: In this case the capacity of each cell is supposed to be infinite. We use the notation $N = \{N_t\}_{t\geq 0}$ for the stochastic process representing the state of this network.
- Constrained: we consider two types of blocking as described in §8.2

- Transition bocking model. We use the notation $N^{\text{tb}} = \left\{N_t^{\text{tb}}\right\}_{t\geq 0}$ for the stochastic process representing the state of this network. (This model is called *Repetitive Service Blocking* with *Random Destination* in [30].)
- Forced termination model: We use the notation $N^{\text{ft}} = \{N_t^{\text{ft}}\}_{t\geq 0}$ for the stochastic process representing the state of this network. (This model is called *Rejection Blocking* in [30].)

Free

The generator q of the Markov process N is given by

$$q(\nu,\Gamma) = \int_{\mathbb{D}} 1_\Gamma(T_{ox}\nu)\,\lambda_x dx + \int_{\mathbb{D}} 1_\Gamma(T_{xo}\nu)\,\mu_x\nu(dx) + \int_{\mathbb{D}\times\mathbb{D}} 1_\Gamma(T_{xy}\nu)\,\lambda'_x p'(x,dy)\,\nu(dx)$$

where

- λ_x designates the intensity (per unit of surface) of the arrivals at the gegraphic location x;
- μ_x designates the inverse of call duration for the service class x;
- λ'_x designates the sojourn duration parameter of users at the gegraphic location x;
- $p'(x,dy)$ is the probability Kernel of motions, i.e. a user finishing its sojourn at node x, is routed to dy according to $p'(x,dy)$.

We have a MPL process with service rates $r_{xy}(\nu) = 1$.and routing rates

$$\lambda(o,dx) = \lambda_x dx, \quad \lambda(x,\{o\}) = \mu_x, \quad \lambda(x,dy) = \lambda'_x p'(x,dy)$$

Regular. The generator q is stable iff Condition (6.12) is satisfied, which is equivalent to

$$\lambda(o,\mathbb{D}) < \infty, \quad \text{and } \forall x\in\mathbb{D}, \lambda(x,\mathbb{D}) = \int \lambda'_x p'(x,dy) < \infty$$

what will be assumed from now on. Consider the birth rates $\{b_n\}$ and death rates $\{d_n\}$ defined by (6.16), i.e.

$$b_n = \sup_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o,\mathbb{D})$$

$$d_n = \inf_{\nu\in\mathbb{M}:\nu(\mathbb{D})=n} q(\nu, T_{\mathbb{D}o}\nu)$$

Observe that $q(\nu, T_{\mathbb{D}\mathbb{D}}\nu) = \sum_{x\in\nu}\nu_x\lambda(x,\mathbb{D})$ hence Condition (6.17) is satisfied if

$$\sup_{x\in\mathbb{D}}\lambda(x,\mathbb{D}) < \infty$$

what will be assumed from now on. Moreover we have $\sup_{\nu\in\mathbb{M}} q(\nu, T_{o\mathbb{D}}\nu) = \lambda(o,\mathbb{D}) < \infty$, hence by Lemma 4, q is regular. Let $\{N_t; t\geq 0\}$ be a Markov jump process associated to the generator q.

Limiting distribution and invariant probability measure. We have exactly the same observations as those made in §10.3.3 for the case without mobility.

Transition blocking model

Let $\mathbb{M}^{\mathrm{f}}$ be the feasibility set corresponding to FC

$$\mathbb{M}^{\mathrm{f}}=\left\{\nu\in\mathbb{M}:\int_{u}\varphi\left(y\right)\nu\left(dy\right)<C,u\in\mathbf{U}\right\} \tag{10.33}$$

Consider the truncated generator q^{tb} on $\mathbb{M}^{\mathrm{f}}$ given by (8.3). The truncated generator q^{tb} is also a SMQ process with the service rates deduced from Equation (8.4)

$$r_{xy}^{\mathrm{tb}}\left(\nu\right)=1\{\nu\in\mathbb{M}^{\mathrm{f}},T_{xy}\nu\in\mathbb{M}^{\mathrm{f}}\}+1\left\{\nu\notin\mathbb{M}^{\mathrm{f}}\right\}$$

Regular. Note that $q^{\mathrm{tb}}\left(\nu\right)\leq q\left(\nu\right)$; then the generator q^{tb} is stable. Observe that $q^{\mathrm{tb}}\left(\nu,T_{\mathbb{DD}}\nu\right)\leq q\left(\nu,T_{\mathbb{DD}}\nu\right)$, hence Condition (6.17) is satisfied. Moreover we have $\sup_{\nu\in\mathbb{M}}q^{\mathrm{tb}}\left(\nu,T_{o\mathbb{D}}\nu\right)\leq\lambda\left(o,\mathbb{D}\right)<\infty$, hence by Lemma 4, q^{tb} is regular. Let $\left\{N_t^{\mathrm{tb}};t\geq0\right\}$ be a Markov jump process associated to the generator q^{tb}.

Limiting distribution and invariant probability measure. We have exactly the same observations as those made in §10.3.3 for the case without mobility. Let Π_ρ^{tb} be the unique invariant with respect to q^{tb}.

Blocking probability. The following lemma proves a condition needed to apply Lemma 15 when calculating the blocking probability.

Lemma 18 *If*

$$\int_{\mathbb{D}}\lambda\left(x,\mathbb{D}\right)\rho\left(dx\right)<\infty$$

what will be assumed from now on*, then*

$$\int_{\mathbb{M}}q\left(\nu\right)\Pi_\rho\left(d\nu\right)<\infty\ \text{and}\ \int_{\mathbb{M}^{\mathrm{f}}}q^{\mathrm{tb}}\left(\nu\right)\Pi_\rho^{\mathrm{tb}}\left(d\nu\right)<\infty \tag{10.34}$$

Proof. Note first that $q^{\mathrm{tb}}\left(\nu\right)\leq q\left(\nu\right)$ then, by the truncation property, it is enough to show the first inequality. We have

$$\begin{aligned}\int qd\Pi_\rho&=\mathbf{E}_{\Pi_\rho}\left[q\left(\nu,T_{o\mathbb{D}}\nu\right)\right]+\mathbf{E}_{\Pi_\rho}\left[q\left(\nu,T_{\mathbb{D}o}\nu\right)\right]+\mathbf{E}_{\Pi_\rho}\left[q\left(\nu,T_{\mathbb{DD}}\nu\right)\right]\\&=\lambda\left(o,\mathbb{D}\right)+\mathbf{E}_{\Pi_\rho}\left[\int_{\mathbb{D}}\lambda\left(x,\mathbb{D}\right)\nu\left(dx\right)\right]\\&=\lambda\left(o,\mathbb{D}\right)+\int_{\mathbb{D}}\lambda\left(x,\mathbb{D}\right)\rho\left(dx\right)\end{aligned}$$

which is finite if the condition of the Lemma holds. ■

From the above lemma and Corollary 10, we deduce that the blocking rate is given by

$$p^{\mathrm{tb}}(o,x) = 1 - \frac{\Pi_\rho\left(\mathbb{M}^{\mathrm{f}}_x\right)}{\Pi_\rho\left(\mathbb{M}^{\mathrm{f}}\right)} \tag{10.35}$$

where

$$\begin{aligned} &\mathbb{M}^{\mathrm{f}}_x \\ &= \left\{\nu \in \mathbb{M}^{\mathrm{f}} : T_{ox}\nu \in \mathbb{M}^{\mathrm{f}}\right\} \\ &= \left\{\nu \in \mathbb{M} : \int_u \varphi d\nu < C - \varphi(x) \text{ for cell } u \text{ containing } x \text{ and } \int_w \varphi d\nu < C \text{ for all } w \neq u\right\} \end{aligned} \tag{10.36}$$

The product form of the stationary distribution leads to

$$p^{\mathrm{tb}}(o,x) = 1 - \frac{\Pi_\rho\left(\int_u \varphi d\nu < C - \varphi(x)\right)}{\Pi_\rho\left(\int_u \varphi d\nu < C\right)}, \quad x \in u \in \mathbf{U}$$

If the traffic ρ is independent from the speed of the users (as in the cases enumerated in Proposition 32), then the blocking probability in the transition blocking model is identical to the case without mobility.

Cut probability. We defined in § 8.2.3 the (*fictitious*) cut probability. Unfortunately, we don't have explicit formula for the term $\Pi^{\mathrm{tb}}\left(N + \delta_x \in \mathbb{M}^{\mathrm{f}}, N + \delta_y \notin \mathbb{M}^{\mathrm{f}}\right)$ comprised in the expression of this cut probability. In order to get a first crude approximation, we average over the geometry inside each cell by replacing FC with a condition on the number M of users in the cell

$$M \leq \Gamma$$

where Γ is the pole capacity defined by (4.4) for the downlink and by (4.10) for the uplink. (This approximation lead to the erlang approximation of the blocking probability presented in § 10.3.4.) Hence we get the approximation

$$c^{\mathrm{tb}} \simeq \frac{d^{\mathrm{tb}}}{1 + d^{\mathrm{tb}}} \tag{10.37}$$

where d^{tb} is given by (8.28). We may attempt to approximate the cut probability, i.e. the term $\Pi^{\mathrm{tb}}\left(N + \delta_x \in \mathbb{M}^{\mathrm{f}}, N + \delta_y \notin \mathbb{M}^{\mathrm{f}}\right)$, more accuretly in future work.

Forced termination model

Regularity and limiting distribution. The arguments are identical to the transition blocking model.

Invariant probability measure. Let Π^{ft}_ρ be an invariant probability measure with respect to q^{ft}. By Proposition 25, there is a unique such probability measure which is the limiting distribution.

Blocking and cut probabilities. In order to apply Lemma 15 when calculating the blocking and cut probabilities, we need that

$$\int_{\mathbb{M}^{\mathrm{f}}} q^{\mathrm{ft}}(\nu)\,\Pi_{\rho}^{\mathrm{ft}}(d\nu) < \infty$$

what will be assumed from now on.

10.4.4 Numerical results

In the simulations we take the call-duration average $\mu^{-1} = 1\,\mathrm{s}$. For other values of the call-duration average μ, we can always retreive this particular case, since we may replace the generator q by q/μ without altering the ergodicity and invariant distribution. In doing so, we should replace the speed v by v/μ. In other words we may view our speed as expressed in kilometers per $\mu\times$ seconds.

Figure 10.8 shows the blocking probability for the models tb and ft for different speeds $v = 0.1, 1, 10$. This figure shows, as expected, that the tb blocking probability is independent of speed. If shows also that the ft blocking probability decreases with speed.

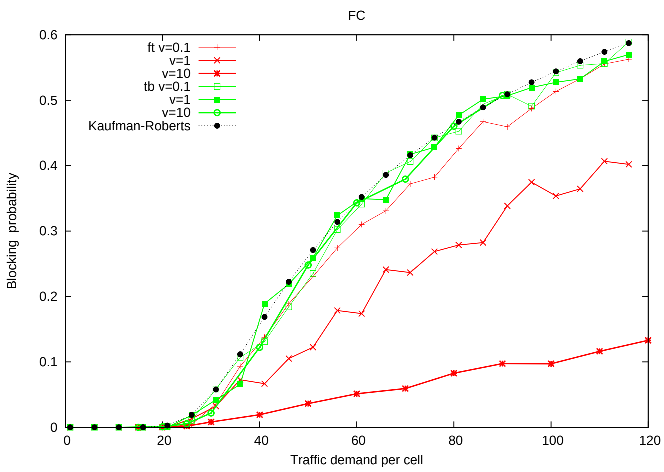

Figure 10.8: Blocking probability for the models tb and ft for different speeds $v = 0.1, 1, 10$.

Figure 10.9 shows the cut probability for the models tb and ft for different speeds $v = 0.1, 1, 10$. (Recall that for the model tb this cut probability is fictitious.) This figure shows that the ft cut probability increases with speed and is well approximated by the tb cut probability as long as it remains small, typically less than 0.05.

Figure 10.10 shows the tb cut probability and its analytical approximation (10.37) for different speeds $v = 0.1, 1, 10$. This figure shows that the approximation is accurate when the cut probability is less than 0.05.

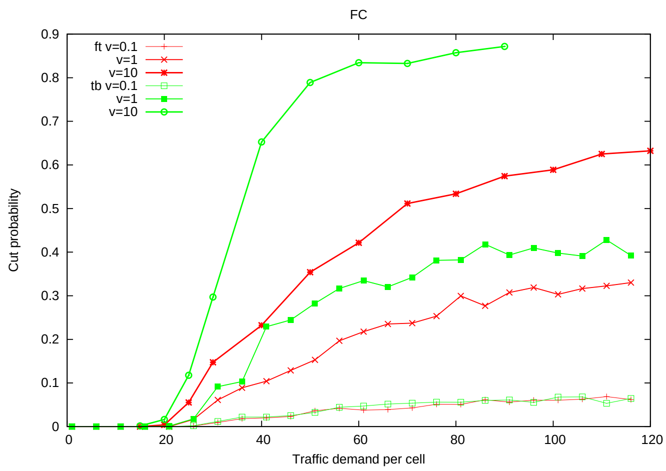

Figure 10.9: Cut probability for the models tb and ft for different speeds $v = 0.1, 1, 10$.

In real networks the cut probability should be maintained an order of magnitude less than 0.05, hence the approximation (10.37) may be used accurately in this case to estimated the ft cut probability.

It remains to calculate the ft blocking probability. This may be carried using the following observation deduced from Figure 10.11: the sum of the ft blocking and cut probabilities is nearly independent of the speed. This is coherent with our previous observation that the ft blocking decreases wheras the ft cut increases with speed. Since for a null speed the cut probability is null and the blocking probability may be calculated by the methods presented in § 10.3.4, we may easily deduce the ft blocking probability for an arbitrary speed.

10.5 Integration of elastic and streaming traffic

We consider in the present section a UMTS Release 99 network carrying both streaming and elastic traffic *on the same bandwidth*. So interference between streaming and elastic users has to be taken into account. Each streaming user is served by a specific DCH whereas the elastic users in a given cell are served simultaneously by the DSCH.

We aim to establish analytical formulae (or bounds) for the quality of service indicators for each type of service in this mixed scenario.

The notations are the same as those of the previous two chapters. Moreover, in order to distinguish the streaming and elastic traffic characteristics, we denote:

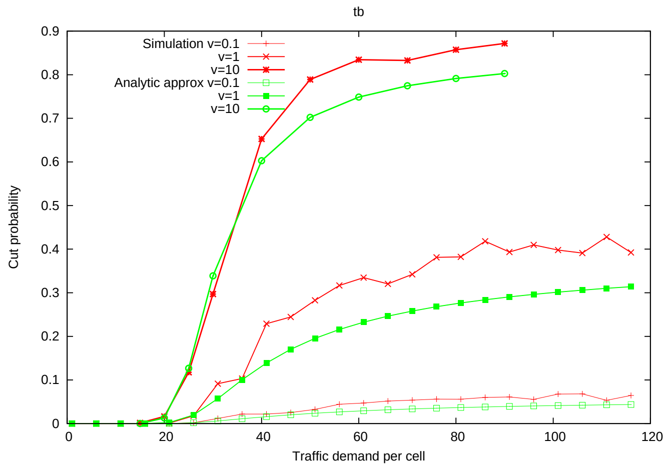

Figure 10.10: The tb cut probability and its analytical approximation (10.37) for different speeds $v = 0.1, 1, 10$.

- λ^{s} intensity of arrivals of streaming users per surface unit
- λ intensity of arrivals of elastic users per surface unit
- $1/\mu^{\mathrm{s}}$ mean duration of streaming calls
- $1/\mu$ mean volume of elastic calls
- N_t^{s} number of streaming users in a cell
- N_t number of elastic users in a cell

10.5.1 Related works

We will always assume that streaming traffic has preemptive priority over elastic traffic, which has two important consequences. Firstly, the evolution of the streaming users is independent of the elastic ones[2]. Secondly, the elastic users are served with the capacity left free by the streaming users. Hence the novelty when we consider the integration is that elastic traffic observes a random environment.

This problem is studied in [4], [95], and [42] for wired communication networks. In this case, the ergodicity condition is, roughly speaking, that the traffic demand of the elastic traffic should be less than the average of the capacity left free by the streaming users.

In [4] and [95] the stationary distribution of the Markov process describing the number of elastic users is calculated numerically using the *matrix-geometric*

[2] This is obvious for feasibility based load control algorithms.

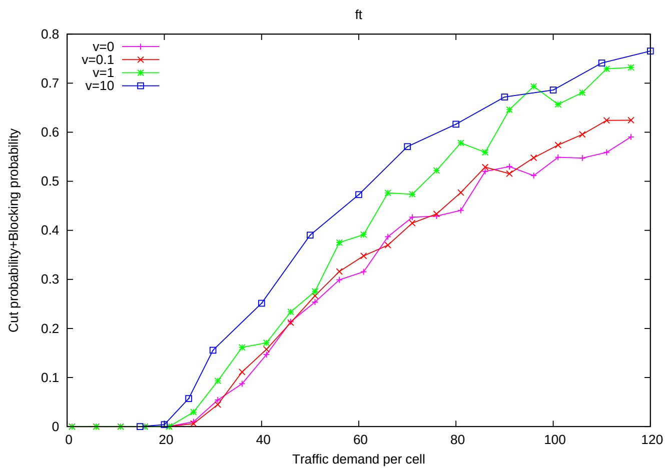

Figure 10.11: Sum of the ft blocking and cut probabilities for different speeds $v = 0, 0.1, 1, 10$.

approach of [93]. This approach considers the two-dimensional Markov process (N_t^{s}, N_t) counting the number of streaming and elastic users respectively which is a *quasi-birth-death process* whose infinitesimal generator can be written in a block matrix form. (Note that the parameters concerning the streaming traffic are denoted with the subscript "s". If there is no risk of confusion, this subscript "s" is omitted.) The stationary distribution can be calculated numerically by solving iteratively a matrix fix-point problem $R = f(R)$ where R is the unknown matrix and f is some non-decreasing function. The convergence is assured by the ergodicity condition. Once R is known, one calculates the distribution of N_t and in particular its average $\mathbf{E}[N_t]$. The expected delay for elastic traffic can then be obtained through the Little's formula. The throughput is obtained as the ratio of the mean volume and the delay. In [4] the trade-off between the performance of streaming and elastic traffic (through the streaming admission threshold) is studied numerically.

In [42] the performance of the streaming traffic are bounded using the so-called *fluid* (fl) and *quasi-stationary* (qs) regimes. The FLUID REGIME corresponds to the case where the elastic traffic is served by a constant capacity equal to the average of the capacity left free by the streaming users. The QUASI-STATIONARY REGIME corresponds to the assumption that, for each given streaming user number, the fluid users attain their stationary regime. The basic inequalities concern the workload W_t (which denotes the volume remaining to be transmitted at time t). In fact, for the GI/GI/1 queue with variable service rate and for any work-conserving service discipline, we have the following inequalities between the stationary workloads

$$W^{\mathrm{fl}} \le_{\mathrm{icx}} W \le_{\mathrm{icx}} W^{\mathrm{qs}}$$

where icx designates the increasing convex ordering [17, p.272] (the fluid bound is first established in [94], [18]). In the particular case of GI/M/1 (i.e. exponential volume distribution) the authors of [42] prove the following inequalities between the average number of elastic users

$$Q^{\mathrm{fl}} \leq_{\mathrm{icx}} Q \leq_{\mathrm{icx}} Q^{\mathrm{qs}}$$

and applying Little's formula they get inequalities between the expected delay for elastic traffic

$$\mathbf{E}[T^{\mathrm{fl}}] \leq \mathbf{E}\,[T] \leq \mathbf{E}[T^{\mathrm{qs}}]$$

As the throughput is the ratio of the volume average and the delay, one gets the inequalities between the average throughputs

$$\mathbf{E}[r^{\mathrm{qs}}] \leq \mathbf{E}\,[r] \leq \mathbf{E}[r^{\mathrm{fl}}]$$

These bounds are used in [42] to investigate some questions such as: the admission threshold for the streaming traffic which assures uniform stability (i.e. stability for each streaming user number) of elastic traffic; the tightness of the bounds; the integration gain, etc. The performance of the systems giving the bounds in [42] are insensitive to the distribution of the streaming call durations and elastic call volumes; whereas the matrix-geometric approach in [4] and [95] rely on the Markovian assumption (i.e. streaming call durations and elastic call volumes are assumed to have exponential distributions).

The above models are well adapted to wired communication networks. In wireless communication models, the relative location of radio channels, determines their joint feasibility. Hence we have to take into account the *spatial geometry of the network*, absent in the above models. An attempt to extend the approach of [4] to the uplink of a CDMA networks is made in [62] with the assumption that the inter-cell interference is proportional to intra-cell interference.

[27] extends to the spatial context the ergodicity condition for elastic traffic served with a randomly varying capacity. This extension is made for generally distributed arrivals and call volumes.

10.5.2 Our contribution

In the integration of streaming and elastic traffic, we take into account the exact representation of the geometry of inter-cell and intra-cell interferences with the help of our FC.

In the mixed scenario, the streaming traffic are supposed to have priority on elastic ones. Thus the streaming users are admitted as long as the power allocation problem for these users (supposed alone) is feasible. The elastic users use the remaining capacity according to a (spatially pondered) processor-sharing policy.

The ergodicity condition for such networks is derived from the ergodicity condition of the GI/GI/1 queue in random environment [94], [18] (or from the ergodicity condition of spatial GI/GI/1 queue established in [27]).

The fluid and quasi-stationary bounds in [42] are used in our spatial context to derive expressions of the delay and throughput of the fluid and quasi-stationary regimes for UMTS networks. In the fluid regime, the *spatial Erlang formula* permits to calculate the capacity left free by the streaming traffic.

10.5.3 Traffic model

The traffic models for streaming and elastic are described in §10.3.3 and §10.2.4 respectively. We assume that the arrival and call-volume processes for elastic traffic are independent from arrival and call-duration processes of streaming traffic. Moreover we assume that each user is motionless during his call (semi-dynamic model).

We consider the FC load control algorithm (admission control for streaming traffic and congestion control for elastic traffic). Hence we may study each cell independently from the other cells. From now we consider a given cell denoted $\mathbb{D}$.

10.5.4 Performance analysis

Feasibility condition

Let N_t^{s} and N_t be the measures representing the positions of streaming and elastic users respectively at time t. The feasibility condition FC may be written as follows

$$\sum_{x\in N_t} X_x\xi_x' < C - \sum_{x\in N_t^{\mathrm{s}}} X_x\xi_x'$$

where

- ξ_x' designates the modified SINR $\xi_x' = \xi_x/\left(1+\alpha\xi_x\right)$;
- ξ_x designates the signal-to-interference-and-noise ratio (SINR) threshold;
- α, C are given constants and X_x is some factor characterizing the geometry. (In the case of uplink with power limitation, the parameter C is not constant, since it depends on the bit-rate of elastic traffic. In this case, we don't take into account the power limit. This approximation may be refined in future studies.)

We assume that streaming traffic has preemptive priority over elastic traffic, then FC may be written as

$$\sum_{x\in N_t} X_x\xi_x' < \mathbf{C}\left(t\right)$$

where $\mathbf{C}\left(t\right) = C - \sum_{x\in N_t^{\mathrm{s}}} \xi_x' X_x$ varies randomly over time.

Similar arguments to those in §10.2.3 show that if the bit-rates $\{r_x; x\in N_t\}$ satisfy

$$\sum_{x\in N_t} \gamma\left(x\right) r_x = \mathbf{C}\left(t\right)$$

where the weighting coefficients $\gamma(x)$ are given by (10.3) then FC holds true.

Queueing System

The system at hand may be seen as a spatial M/GI queue in a random environment. The birth rates and death rates are given respectively by

$$\lambda(o, dx) = \lambda_x dx, \quad \lambda(x, \{o\}) = \mu_x$$

and the service rates are given by

$$r_{xy}(N_t) = \begin{cases} 1 & \text{if } x = o, y \in \mathbb{D} \\ \mathbf{C}(t)\, \gamma(x)^{-1} N_t(\mathbb{D})^{-1} & \text{if } y = o, x \in \mathbb{D} : N_t(\{x\}) > 0 \end{cases}$$

As in Lemma 12, it is equivalent to consider the routing rates

$$\lambda'(o, dx) = \lambda_x dx, \quad \lambda'(x, \{o\}) = \mu_x \gamma(x)^{-1}$$

and the service rates

$$r'_{xy}(N_t) = \begin{cases} 1 & \text{if } x = o, y \in \mathbb{D} \\ \mathbf{C}(t)\, N_t(\mathbb{D})^{-1} & \text{if } y = o, x \in \mathbb{D} : N_t(\{x\}) > 0 \end{cases}$$

(where we replaced $\lambda(x, \{o\})$ by $\lambda(x, \{o\})\, \gamma(x)^{-1}$ and $r_{xo}(N_t)$ by $r_{xo}(N_t)\, \gamma(x)$) which should satisfy

$$\sum_{x \in N_t} r'_{xo}(N_t) = \mathbf{C}(t)$$

Stability condition

Similarly to Proposition 37, we may associate to our system a M/GI/1 queue with arrival rate $\lambda' = \lambda'(o, \mathbb{D})$, with mean call-volume

$$\mu'^{-1} = \frac{1}{\lambda'(o, \mathbb{D})} \int_{\mathbb{D}} \frac{\lambda'(o, dx)}{\lambda'(x, \{o\})}$$

and such that, if at time t there are n users in the queue, each one is served at rate $\mathbf{C}(t)/n$. This is called a M/GI/1 queue in RANDOM ENVIRONMENT. From the properties of such queues [94], [18] (or from the ergodicity condition of spatial GI/GI/1 queue established in [27]), we deduce that the ergodicity condition writes $\rho'(\mathbb{D}) < \bar{\mathbf{C}} = \mathbf{E}_{\Pi^{\text{tb}}}[\mathbf{C}(0)]$ where the traffic demand $\rho'(\mathbb{D})$ is given by

$$\begin{aligned} \rho'(\mathbb{D}) &= \lambda' \mu'^{-1} \\ &= \int_{\mathbb{D}} \frac{\lambda'(o, dx)}{\lambda'(x, \{o\})} \\ &= \int_{\mathbb{D}} \frac{\gamma(x)\, \lambda(o, dx)}{\lambda(x, \{o\})} \\ &= \int_{\mathbb{D}} \frac{\gamma(x)\, \lambda_x}{\mu_x} dx = \frac{E_{\text{b}}/N_0}{W'} \int_{\mathbb{D}} X_x \frac{\lambda_x}{\mu_x} dx \end{aligned}$$

and Π^{tb} is the stationary distribution of the streaming users.

Note that

$$\begin{aligned}\bar{\mathbf{C}} &= C - \int_{\mathbb{D}} X_x \xi'_x \mathbf{E}_{\Pi^{\text{tb}}} \left[N_0^{\text{s}} (dx) \right] \\ &= C - \int_{\mathbb{D}} X_x \xi'_x \left[1 - p^{\text{tb}} (o, x) \right] \frac{\lambda_x^{\text{s}}}{\mu_x^{\text{s}}} dx\end{aligned}$$

where for the second equality we use Lemma 16 and $p^{\text{tb}} (o, x)$ is given by the spatial Erlang formula (10.31).

Assume that the parameters $\lambda_x^{\text{s}}, \mu_x^{\text{s}}, \xi'_x$ don't depend on the geographic position (but there may be several streaming traffic classes $j \in J^{\text{s}}$, each characterized by specific values of the parameters $\lambda_j^{\text{s}}, \mu_j^{\text{s}}, \xi'_j$). In this case

$$\bar{\mathbf{C}} = C - \bar{M}\bar{\varphi} + \int_{\mathbb{D}} X_x \xi'_x p^{\text{tb}} (o, x) \frac{\lambda_x^{\text{s}}}{\mu_x^{\text{s}}} dx$$

where $\bar{M}$ and $\bar{\varphi}$ designate respectively the traffic demand per cell and the average of $\varphi(\cdot)$ for streaming calls.

The approximation (10.32) of the blocking rates may be written in the form

$$p^{\text{tb}} (o, x) \simeq \xi'_x X_x q, \quad \text{where } q = \frac{\exp\left[- \left(\left(C - \bar{S} \right) / \sigma_S \right)^2 / 2 \right]}{\sigma_S \sqrt{2\pi} \left\{ 1 - Q \left[\left(C - \bar{S} \right) / \sigma_S \right] \right\}} \tag{10.38}$$

where $\bar{S} = \bar{\varphi}\bar{M}$ and $\sigma_S^2 = \bar{\varphi^2}\bar{M}$, then

$$\int_{\mathbb{D}} X_x \xi'_x p^{\text{tb}} (o, x) \frac{\lambda_x^{\text{s}}}{\mu_x^{\text{s}}} dx \simeq q \int_{\mathbb{D}} X_x^2 \xi'^2_x \frac{\lambda_x^{\text{s}}}{\mu_x^{\text{s}}} dx = q \bar{\varphi^2} \bar{M}$$

Then the average residual capacity is given by

$$\bar{\mathbf{C}} \simeq C + \left(q\bar{\varphi^2} - \bar{\varphi} \right) \bar{M} \tag{10.39}$$

Gathering the above results we deduce that the ergodicity condition for UMTS networks with mixed services writes

$$\frac{E_{\text{b}}/N_0}{W'} \bar{X} \bar{r}_d < \bar{\mathbf{C}} \simeq C + \left(q\bar{\varphi^2} - \bar{\varphi} \right) \bar{M}$$

where $\bar{r}_d$ is the traffic demand for elastic calls.

Fluid bound

Considering a Markovian context M/M (i.e. exponential inter-arrival durations and service requirements), we may attempt to apply the *matrix-geometric* approach to the spatial case. To this end we may make a discretization of the space but this will lead to a discrete multiclass M/M queue whereas only a single class queue is treated in [4]. Even if one succeeds to extend the approach in [4] to the multiclass case, the numerical calculations would be time consuming. We shall focus on the *fluid* and *Quasi-stationary* bound.

Recall that the service policy is given by

$$r_{xo}(N_t) = \mathbf{C}(t)\,\gamma(x)^{-1}\,N_t(\mathbb{D})^{-1}$$

For the fluid bound, the term $C(t)$ in the right-hand side is replaced by $\bar{\mathbf{C}}$, that is

$$r_{xo}^{\text{fl}}(N_t) = \bar{\mathbf{C}}\gamma(x)^{-1}\,N_t(\mathbb{D})^{-1}$$

The expected delay and throughput for elastic calls in $A \in \mathcal{D}$, denoted $\bar{T}^{\text{fl}}(A)$ and $\bar{r}^{\text{fl}}(A)$ respectively, may be deduced from equations (10.14) and (10.16)

$$\bar{T}^{\text{fl}}(A) = \frac{\rho'(A)/\bar{\mathbf{C}}}{\lambda(o,A)\left(1-\rho'(\mathbb{D})/\bar{\mathbf{C}}\right)} = \frac{\rho'(A)}{\lambda(o,A)\left(\bar{\mathbf{C}}-\rho'(\mathbb{D})\right)} \tag{10.40}$$

$$\bar{r}^{\text{fl}}(A) = \frac{\rho(A)}{\rho'(A)/\bar{\mathbf{C}}}\left[1-\rho'(\mathbb{D})/\bar{\mathbf{C}}\right] = \frac{\rho(A)}{\rho'(A)}\left[\bar{\mathbf{C}}-\rho'(\mathbb{D})\right] \tag{10.41}$$

where $\rho(A)$ is given by (10.15). In particular, the expected throughput for users in all the area $\mathbb{D}$ equals

$$\bar{r}^{\text{fl}}(\mathbb{D}) = \frac{\rho(\mathbb{D})}{\rho'(\mathbb{D})}\left[\bar{\mathbf{C}}-\rho'(\mathbb{D})\right]$$

Multiclass case. Assume that the arrival and call-volume parameters λ_x, μ_x for elastic services don't depend on the geographic position (but there may be several elastic traffic classes $j \in J$, each characterized by specific values of the parameters λ_j, μ_j). Then

$$\rho(\mathbb{D}) = \sum_{j\in J}\rho_j, \quad \rho'(\mathbb{D}) = \sum_{j\in J}\rho'_j = \frac{E_{\text{b}}/N_0}{W'}\bar{X}\rho(\mathbb{D})$$

where

$$\rho_j = \pi R^2\frac{\lambda_j}{\mu_j}, \quad \rho'_j = \frac{E_{\text{b}}/N_0}{W'}\bar{X}\rho_j$$

Hence in the fluid regime, for the elastic traffic class j, the delay equals

$$\bar{T}_j^{\text{fl}} = \frac{\rho'_j}{\pi R^2\lambda_j\left(\bar{\mathbf{C}}-\rho'(\mathbb{D})\right)} = \frac{1}{\mu_j}\frac{\rho'(\mathbb{D})}{\rho(\mathbb{D})}\frac{1}{\bar{\mathbf{C}}-\rho'(\mathbb{D})}$$

and the expected throughput per cell is given by

$$\bar{r}_j^{\text{fl}} = \frac{\mu_j^{-1}}{\bar{T}_j} = \frac{\rho(\mathbb{D})}{\rho'(\mathbb{D})}\left[\bar{\mathbf{C}}-\rho'(\mathbb{D})\right]$$

which is independent of elastic class $j \in J$.

Quasi-stationary bound

In this case $\mathbf{C} = \mathbf{C}(t)$ is considered as a random variable depending on the streaming call positions. For elastic calls in $A \in \mathcal{D}$, the delay is given by

$$\bar{T}^{\mathrm{qs}}(A) = \frac{\rho'(A)}{\lambda(o,A)} \mathbf{E}\left[\frac{1}{\mathbf{C} - \rho'(\mathbb{D})}\right]$$

and the expected throughput per cell equals

$$\bar{r}^{\mathrm{qs}}(A) = \frac{\mu^{-1}}{\bar{T}^{\mathrm{qs}}(\mathbb{D})} = \frac{\rho(A)}{\rho'(A)} \left(\mathbf{E}\left[\frac{1}{\mathbf{C} - \rho'(\mathbb{D})}\right]\right)^{-1} \tag{10.42}$$

Comparison

Rewrite the fluid (upper) bound of the throughput (10.41) as follows

$$\bar{r}^{\mathrm{fl}}(A) = \frac{\rho(A)}{\rho'(A)} \left[C - \rho'(\mathbb{D}) - \mathbf{E}\left[\sum_{x \in N_t^{\mathrm{s}}} \xi'_x X_x\right]\right]$$

and the quasi-stationary (lower) bound of the throughput (10.41) as follows

$$\bar{r}^{\mathrm{qs}}(A) = \frac{\rho(A)}{\rho'(A)} \left(\mathbf{E}\left[\frac{1}{C - \rho'(\mathbb{D}) - \mathbf{E}\left[\sum_{x \in N_t^{\mathrm{s}}} \xi'_x X_x\right]}\right]\right)^{-1}$$

We see that the difference between the fluid and quasi-stationary bound comes from the fact that, for a random variable Z, $\mathbf{E}[Z] \neq \left(\mathbf{E}\left[Z^{-1}\right]\right)^{-1}$ in general. (More precisely, from Jensen inequality we get $\mathbf{E}[Z] \geq \left(\mathbf{E}\left[Z^{-1}\right]\right)^{-1}$ which is coherent with the fact $\bar{r}^{\mathrm{fl}} \geq \bar{r}^{\mathrm{qs}}$.) Hence it is enough to consider the quantities

$$\theta^{\mathrm{fl}} = C - \rho'(\mathbb{D}) - \mathbf{E}\left[\sum_{x \in N_t^{\mathrm{s}}} \xi'_x X_x\right] \tag{10.43}$$

and

$$\theta^{\mathrm{qs}} = \left(\mathbf{E}\left[\frac{1}{C - \rho'(\mathbb{D}) - \sum_{x \in N_t^{\mathrm{s}}} \xi'_x X_x}\right]\right)^{-1} \tag{10.44}$$

which we call AVERAGE RESIDUAL CAPACITIES.

10.5.5 Numerical results

Unless otherwise specified, all the numerical applications are made using the default values specified in Section 2.2.5. Moreover we consider DFC.

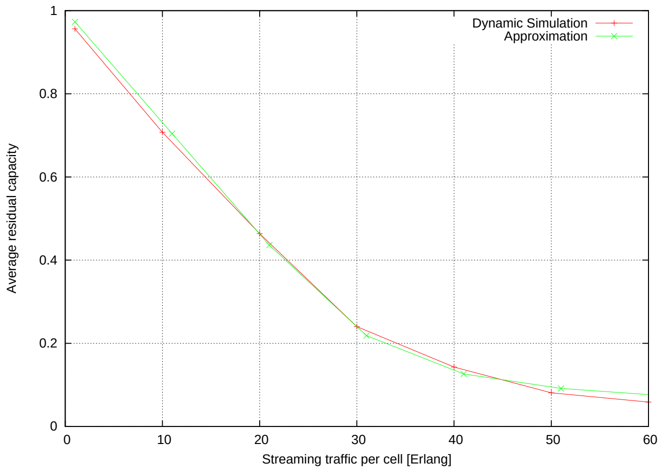

Figure 10.12: Validation of average residual capacity approximation.

Validation of the explicit throughput upper bound

The expression of the fluid (upper) bound of the throughput (10.41) relies on the average residual capacity, denoted $\theta^{\text{fl}} = \bar{\mathbf{C}}$ or $\mathbf{E}[\mathbf{C}]$. The expression (10.39) of the average residual capacity is approximative, relying in particular on the Gaussian approximation (10.38) of the blocking rates of streaming traffic. In order to validate this approximation, we compare the average residual capacity calculated by (10.39) to that estimated by dynamic simulations.

The dynamic simulations are carried as follows. 100 simulations are made, in each one we start from an empty system and simulate 10000 transitions (arrivals or departures of streaming users) in order to attain the stationary regime. The residual capacity at the end of each simulation is recorded, say $\mathbf{C}(n)$. Finally we take the average of the 100 obtained values $\frac{1}{100}\sum_{n=1}^{100}\mathbf{C}(n)$ as an estimation of the residual capacity $\mathbf{E}[\mathbf{C}]$ (this is basically the crude Monte Carlo method).

Figure 10.12 shows that the approximation is good.

Comparison of throughput upper and lower bounds

Since there is no explicit approximate expression available for the quasi-stationary bound, we use the dynamic simulation to estimate both the fluid and quasi-stationary bounds. Figure 10.13 represents the fluid and quasi-stationary average residual capacities θ^{fl} and θ^{qs} as function of the streaming traffic per cell $\bar{M}$. We observe that the variation of the absolute difference between the average residual capacities, $\theta^{\text{fl}} - \theta^{\text{qs}}$, as $\bar{M}$ increases is not monotonic.

But it is interesting to consider the ratio $\frac{\theta^{\text{fl}}}{\theta^{\text{qs}}}$ which in fact increases when $\bar{M}$ increases or equivalently when the blocking probability $\bar{b}$ of streaming traffic increases. Table 10.1 shows that for $\bar{M} = 10$ ($\bar{b} = 0\%$) the throughput upper

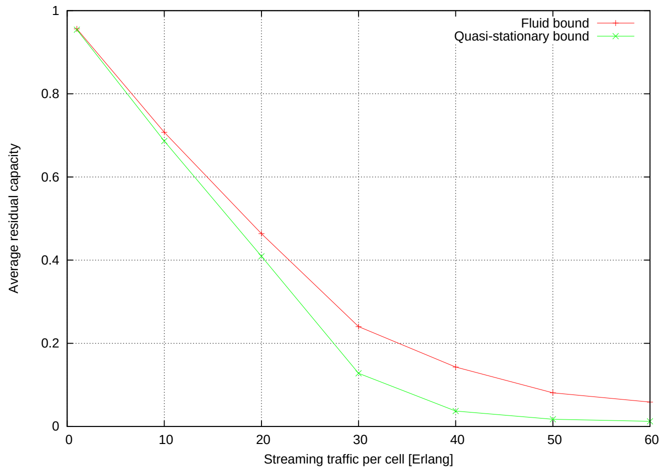

Figure 10.13: Comparison of fluid and quasi-stationary bounds.

$\bar{M}$	**10**	**30**	**60**
$\bar{b}$	0%	2%	30%
$\theta^{\text{fl}}/\theta^{\text{qs}}$	1.03	2	5

Table 10.1: Comparison of fluid and quasi-stationary bounds.

and lower bounds are very close; for $\bar{M} = 30$ ($\bar{b} = 2\%$) the throughput upper bound is about two times the lower bound; and for $\bar{M} = 60$ ($\bar{b} = 30\%$) the throughput upper bound is about five times the lower bound.

Impact of streaming traffic of elastic ones

In order to estimate the impact of streaming traffic of elastic calls, we consider two configurations:

- $\bar{M} = 0$, in this case elastic traffic uses alone the capacity of the system (corresponding to a bandwidth of 5MHz)
- $\bar{M} = 30$, in this case the elastic traffic uses the capacity left free by streaming traffic (the blocking probability of streaming traffic $\bar{b}$ is about 2%)

We calculate the throughput and delay of elastic calls using the expressions given by the fluid bound. Figures 10.14 and 10.15 represent the fluid throughtput and delay for the above two cases. We observe that the presence of streaming traffic reduces significantly the throughput and increases significantly the delay of elastic calls.

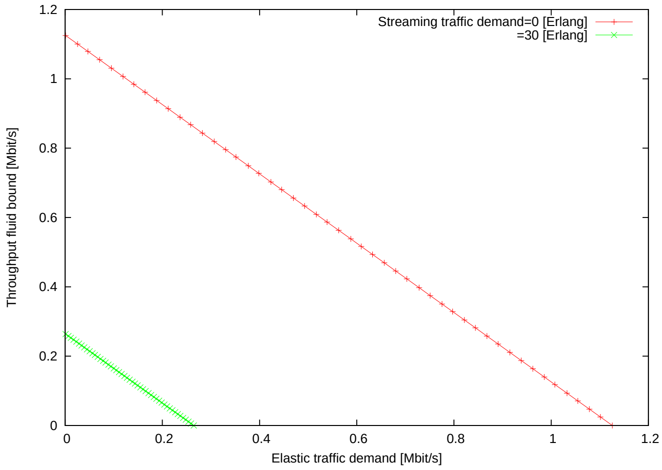

Figure 10.14: Impact of streaming services on throughput of elastic services

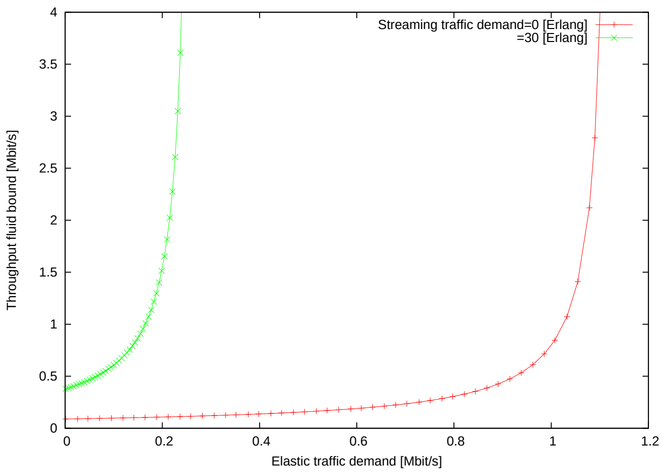

Figure 10.15: Impact of streaming services on delay of elastic services

Summary of numerical results. We make comparison of the fluid (upper) and quasi-stationary (lower) bounds of the throughput. The ratio between these bounds increases when streaming traffic demand increases (or equivalently when the blocking probability of streaming traffic increases). For a blocking probability around 2%, the throughput upper bound is about two times the lower bound.

Finally we show that the presence of streaming traffic reduces significantly the throughput (and equivalently increases significantly the delay) of elastic calls.

Chapter 11

HSDPA

In this section we complement the previous studies by evaluating the performance of the *congestion control policies* for elastic traffic in a UMTS network release HSDPA (analog of HDR) which relies on a channel-aware policy. Assuming Markovian arrivals and departures of customers that transmit some given data-volumes, as well as some temporal channel variability (fading), we study the *mean throughput in different parts of the cell*; i.e., the mean bit-rates that these policies offer in the long-term evolution of the model. Explicit formulas are obtained in the case of *proportional fair policies*, which take advantage of the fading.

We combine known results, in particular concerning the performance of channel aware policies [29], and mobility [22] with our exact representation of the geometry of interferences. More precisely, we use [29] that gives the performance of an opportunistic weight-based rate allocation policy, which is shown in [79] to be a good model for opportunistic schedulers implemented in HSDPA (or HDR). Moreover, following the idea presented in [22] of quasi-stationary and fluid limit, we consider two extreme cases of motionless and infinitely-rapid users.

11.1 From link level to dynamic system level

11.1.1 Link level

We consider here elastic traffic served by a HSDPA network using a specific $W = 5\,\text{MHz}$ bandwidth. We assume that there is no power control, i.e. each base station transmits at the maximal power denoted $\tilde{P}$.

Since there is *no power control* in HSDPA, the power at the input of the receiver, denoted $\mathbf{S}$, is given by

$$\mathbf{S} = \tilde{P}F/L$$

where $\tilde{P}$ designates the transmitted power; L designates the distance pathloss; and F designates the fading effect (cf. [53, §6.3]).

Remark 26 *It is shown in [118, §3.4.3 Eq. (3.124)] that the fading term has the form*

$$F = \sum_{\ell=1}^{L} |h_\ell|^2$$

where $\{h_\ell; \ell = 1, \dots, L\}$ *designate the channel tap gains which may be assumed i.i.d.* $\mathcal{CN}(0, 1/L)$, *i.e. circular symmetric Gaussian random variables [118, Appendix A] with mean* 0 *and variance* $1/L$. *Hence* $\left\{|h_\ell|^2; \ell = 1, \dots, L\right\}$ *are i.i.d. exponential random variables with mean* $1/L$. *Therefore* F *has an Erlang distribution with probability density function*

$$f(x) = \frac{x^{L-1}}{L^L (L-1)!} e^{-x/L}$$

In the particular case $L = 1$, F *is an exponential random variable with mean* 1.

The theoretical maximal bit-rate of the AWGN channel is given by the *Shannon*'s bound

$$r = W \log_2 \left(1 + \frac{\mathbf{S}}{N}\right) \tag{11.1}$$

where W is the bandwidth (5 MHz) and $\frac{\mathbf{S}}{N}$ is the signal to noise power ratio at the input of the receiver.

We assume that the link adaptation (changing the coding and modulation according to the signal to noise conditions) in HSDPA permits to offer the third of the bit-rate Shannon limit, that is

$$r = \frac{1}{3} W \log_2 \left(1 + \frac{\mathbf{S}}{N}\right) \tag{11.2}$$

Moreover we make a linear approximation of the above relation

$$r = \frac{W}{3 \ln(2)} \frac{\mathbf{S}}{N} \tag{11.3}$$

These approximations are compared in Figure 11.1 to the HSDPA link performance obtained by simulations[1] and to the DSCH link performance.

11.1.2 From link level to static system level

By similar arguments to [118, §4.3.1], *interference* may be *approximated by a AWGN* with power averaged over the fast fading. Hence the ratio $\frac{\mathbf{S}}{N}$ in Equations (11.1), (11.2) and (11.3) should be replaced by

$$\frac{\mathbf{S}}{N + I}$$

[1] Curve from [114] treated by A. Saadani and N. Ibrahim (France Telecom R&D).

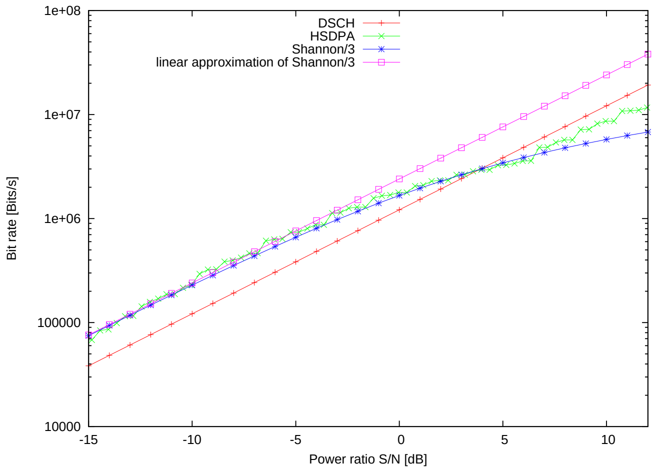

Figure 11.1: HSDPA link performance

where $\mathbf{S}$ comprises the fast fading whereas I is averaged over the fast fading. In particular Equation (11.3) becomes

$$r = \frac{W}{3\ln(2)} \frac{\mathbf{S}}{N+I} \tag{11.4}$$

For a user at position m served by a base station u (which we denote $m \in u$), the signal power is given by

$$\mathbf{S}_m = F_m \tilde{P}_u / L_{u,m}, \quad m \in u$$

and the interference is given by

$$I_m = \sum_{v \in \mathbf{U} \setminus \{u\}} \tilde{P}_v / L_{v,m}, \quad m \in u$$

Note that since the HSDPA is a TDMA system, the HSDPA users in the same cell don't interfere with each other.

The rate for a user at position m, denoted r_m, may be obtained from Equation (11.4)

$$\begin{aligned} r_m &= \frac{W}{3\ln(2)} \frac{\mathbf{S}_m}{N+I_m} \\ &= \frac{W}{3\ln(2)} \frac{F_m \tilde{P}_u / L_{u,m}}{N + \sum_{v \in \mathbf{U} \setminus \{u\}} \tilde{P}_v / L_{v,m}} = \sigma \frac{F_m}{X(m)}, \quad m \in u \end{aligned} \tag{11.5}$$

where

$$X(m) = \frac{N L_{u,m}}{\tilde{P}_u} + \sum_{v \in \mathbf{U} \setminus \{u\}} \frac{\tilde{P}_v}{\tilde{P}_u} \frac{L_{u,m}}{L_{v,m}}, \quad m \in u$$

We assume from now that the powers $\tilde{P}_v$ are the same for all base stations v, then

$$\begin{aligned} X(m) &= NL_{u,m}/\tilde{P} + \sum_{v\in\mathbf{U}\backslash\{u\}} L_{u,m}/L_{v,m} \\ &= NL_{u,m}/\tilde{P} + f(m), \quad m \in u \end{aligned}$$

where $f(m) = \sum_{v\in\mathbf{U}\backslash\{u\}} L_{u,m}/L_{v,m}$ is the so-called f-factor which we studied in depth for the hexagonal base station pattern.

Remark 27 *The* $\{r_m; m \in u\}$ *are called* FEASIBLE RATES *in [29, §II] without explicit information about how to get these feasible rates. Equation (11.5) shows how to relate the feasible rates to the geometry of interference.*

Fading varies and should be considered as a new source of dynamics (with arrivals, mobility, departures); this model is analytically intractable. For a rapidly changing fading, we use the symplifying idea of *separation of the time scales* allows to "inject" the impact of fading to our previous considerations at the static system level.

We consider first a time scale at which users don't move. Hence the number, say M, and positions of the users are fixed. The scheduling over fading of the users in a given cell may be studied independently from the users in the other cells. Hence, at this time scale, we may restrict ourselves to a given base station. Therefore we get a set of feasible rates

$$r_m = \frac{W}{3\ln(2)} \frac{F_m}{X(m)}, \quad m \in \{1, \ldots, M\}$$

At the time scale considered here, the $X(1), \ldots, X(m)$ are constant whereas $F_1, \ldots, F_M$ are i.i.d. exponential random variables. Following [29, §II], we denote $r = (r_1, \ldots, r_M)$ and $p(r)$ its distribution. Note that

$$\mathbf{E}[r_m] = \frac{W}{3\ln(2)} \frac{1}{X(m)}, \quad m \in \{1, \ldots, M\}$$

then

$$\frac{r_m}{\mathbf{E}[r_m]} = F_m, \quad m \in \{1, \ldots, M\}$$

Hence $\{r_m/\mathbf{E}[r_m]; m \in \{1, \ldots, M\}\}$ are i.i.d. (The author of [29, §II] says that we have in this case a symmetric rate distribution.)

In §11.3 we study the HSDPA (proportional fair) scheduler performance with the help of the papers [29, §II] and [79]. In particular the throughput of HSDPA is given by

$$r'_m = \frac{G(M)}{M}\mathbf{E}[r_m], \quad m \in \{1, \ldots, M\}$$

where

$$G(M) = \mathbf{E}\left[\max_{m=1,\ldots,M} F_m\right]$$

If the F_m are exponentially distributed with mean 1, then

$$G(M) = \sum_{m=1}^{M} \frac{1}{m}$$

11.1.3 From static to dynamic system level

Applying the results of Example 9 with

$$\gamma(x) = 1/\mathbf{E}[r(x)] = \frac{3\ln(2)}{W} X(x), \quad \text{and } h(M) = M/G(M)$$

we deduce that the mean number of calls, delay and tthroughput for the no mobility case ($\bar{N}(A)$, $\bar{T}(A)$ and $\bar{r}(A)$ respectively) and infinite mobility case ($\bar{N}$, $\bar{T}$ and $\bar{r}$ respectively) are given by

$$\bar{N}(A) = \rho'(A)\mathcal{H}(\rho'(\mathbb{D})), \quad \bar{N} = \rho'\mathcal{H}(\rho')$$
$$\bar{T}(A) = \frac{\rho'(A)}{\lambda(o,A)}\mathcal{H}(\rho'(\mathbb{D})), \quad \bar{T} = \frac{\rho'}{\lambda(o,\mathbb{D})}\mathcal{H}(\rho')$$
$$\bar{r}(A) = \frac{\rho(A)}{\rho'(A)}\frac{1}{\mathcal{H}(\rho'(\mathbb{D}))}, \quad \bar{r} = \frac{\rho}{\rho'}\frac{1}{\mathcal{H}(\rho')}$$

where $\rho(A), \rho'(A)$ are given by (10.15), (10.13) respectively; ρ, ρ' are given by (10.22), (10.17) respectively and the function $\mathcal{H}(\cdot)$ is given by (7.9).

11.1.4 Comparing DSCH and HSDPA

The differences between HSDPA and DSCH are the following:

(D1) Link performance is better in HSDPA due to link adaptation (changing the coding and modulation according to the signal to noise conditions)

(D2) Proportional fair scheduler permits to take benefit from fast fading

(D3) In HSDPA, due to TDMA there is no interference between users in the same cell

(D4) NRT DSCH shares a 5MHz bandwidth with RT DCH whereas NRT HSDPA uses a specific 5MHz bandwidth

11.2 Numerical results

The numerical applications are made using the default values specified in Section 2.2.5. Moreover we consider a cell radius $R = 1$.

Figure 11.2 shows the expected throughput as function of the traffic demand for different policies in motionless scenario. Figure 11.2 gives an example of the throughput of DSCH (pink curve) and HSDPA (red curve).

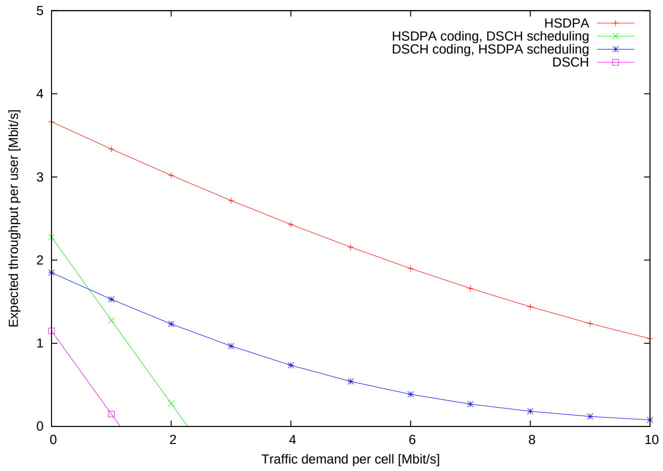

Figure 11.2: Expected throughput for different policies in motionless scenario

Note that the DSCH scheduling gives the throughput that is a linear function of the demanded traffic in the bounded stability region, while the HSDPA one is stable on the whole line, and its gain increases with the traffic.

The green and blue curves consist of mixing DSCH and HSDPA coding and scheduling which are not realistic but permit to understand the benefit in HSDPA coming from coding and that coming from scheduling. In particular we see that the HSDPA coding permits a gain of about 1.5. The HSDPA scheduling gain is particularly important when the traffic demand increases.

11.3 HSDPA scheduler performance

The section 11.3 is simply a reminder (and essentially a transcription, except for the proofs which are more detailed here) of the notations, definitions and results of [29, §II] useful to derive the HSDPA scheduler performance. In the following Section 11.3 we establish the link of Borst's work in [29, §II] with Tse's scheduler [118, §6.7.1] with the help of the Kushner and Whiting work [79].

Reminder of [29, §II]

We assume that the feasible rates for the various users vary over time according to some stationary discrete-time stochastic process $\{R_1(t), \ldots, R_M(t)\}$, with $R_i(t)$ representing the feasible rate for user i in time slot t.

Let $(R_1, \ldots, R_M)$ be a random vector with as distribution the joint stationary distribution of the FEASIBLE rates. We focus on the case where the feasible rates $(R_1, \ldots, R_M)$ have a discrete distribution on some finite set $\mathcal{R} \subseteq \mathbb{R}_+^M$. Let $p(r)$ be the stationary probability that the instantaneous feasible rate vector is

$r \in \mathcal{R}$. With minor modifications, most of the results extend to scenarios with a continuous rate distribution.

An ALLOCATION POLICY is determined by some variables $x_i(r)$ which is the probability that the slot is allocated to user i when the instantaneous rate vector is r. Assuming ergodicity, $x_i(r)$ equals the fraction of time slots allocated to user i in which the instantaneous rate vector is r. Thus, the term $\sum_{r \in \mathcal{R}} p(r) x_i(r) r_i$ represents the THROUGHPUT RECEIVED by user i.

For some policy, let $T_i := \sum_{r \in \mathcal{R}} p(r) x_i(r) r_i$ be the (long-term) throughput received by user i.

Definition 4 *A vector $T \in \mathbb{R}_+^M$ is called* ACHIEVABLE THROUGHPUT *iff there exists an allocation policy determined by variables $x_i(r), i = 1, \ldots, M, r \in \mathcal{R}$ such that*

$$\begin{array}{ll} \sum_{r \in \mathcal{R}} p(r) x_i(r) r_i \geq T_i, & i = 1, \ldots, M \\ \sum_{i=1}^{M} x_i(r) \leq 1, & r \in \mathcal{R} \\ x_i(r) \geq 0, & i = 1, \ldots, M, r \in \mathcal{R} \end{array} \tag{11.6}$$

(i.e. vectors $T \in \mathbb{R}_+^M$ such that there exists a policy such that the throughput received by user i is at least equal to T_i). Let $\mathcal{A} \subseteq \mathbb{R}_+^M$ be the set of ACHIEVABLE *throughput vectors.*

We now consider a scenario where the distribution of the rate vector is symmetric in the sense that the relative fluctuations in the feasible rates for the various users around the respective time-average values are statistically identical. Specifically, we assume that $R_i \stackrel{d}{=} \mathbf{S}_i Y_i Z$, where $\mathbf{S}_i := \mathbf{E}[R_i]$ is the time-average rate of user i, $Y_1, \ldots, Y_M$ are independent and identically distributed copies, and Z represents a possible correlation component with unit mean. Define $G(M) := \mathbf{E}[\max_{j=1,\ldots,M} Y_j]$. We assume that with probability 1, card $\{i : Y_i = \max_{j=1,\ldots,M} Y_j\} = 1$ (this is the case if the Y_j are continuous and i.i.d.).

Lemma 19 *[29, Lemma 2.1] Any achievable throughput vector $T \in \mathcal{A}$ satisfies, for any vector $(\alpha_1, \ldots, \alpha_M) \in \mathbb{R}_+^M$,*

$$\sum_{j=1}^{M} \alpha_j T_j \leq \mathbf{E}\left[\max_{j=1,\ldots,M} \alpha_j R_j\right]$$

with equality for the throughput achieved by the weight-based policy which assigns a weight $B_i = \alpha_i$ to user i.

Proof. By Definition 4, there exists an allocation policy $x_i(r), i = 1, \ldots, M, r \in$

$\mathcal{R}$ satisfying (11.6). Then

$$\begin{aligned}\sum_{j=1}^{M}\alpha_j T_j &\le \sum_{j=1}^{M}\alpha_j \sum_{r\in\mathcal{R}} p(r)x_i\left(r\right)r_i = \sum_{r\in\mathcal{R}} p(r)\sum_{j=1}^{M}\alpha_j r_j x_j\left(r\right) \\ &\le \sum_{r\in\mathcal{R}} p(r)\max_{j=1,\ldots,M}\alpha_j r_j \\ &= \mathbf{E}\left[\max_{j=1,\ldots,M}\alpha_j R_j\right]\end{aligned}$$

For the weight-based policy which assigns a weight $B_i = \alpha_i$ to user i, the inequalities in the above equations may be replaced by equalities. This finishes the proof of the Lemma. ■

Lemma 20 *[29, Lemma 2.2] In the case of a symmetric rate distribution as described above, the weight-based strategy defined by the weights $B_i^* = 1/\mathbf{S}_i$ achieves the throughputs*

$$T_i^* = \frac{G\left(M\right)}{M}\mathbf{S}_i = \frac{G\left(M\right)}{M}\mathbf{E}\left[R_i\right] \tag{11.7}$$

The throughput vector T^* given by (11.7) is optimal for the problem

$$\max_{T\in\mathcal{A}}\sum_{i=1}^{M}\frac{T_i}{\mathbf{E}\left[R_i\right]} \tag{11.8}$$

Proof. The allocation policy is defined by

$$x_i^*\left(R\right) = 1\left\{\frac{R_i}{\mathbf{S}_i} = \max_{j=1,\ldots,M}\frac{R_j}{\mathbf{S}_j}\right\} = 1\left\{Y_i = \max_{j=1,\ldots,M}Y_j\right\} \tag{11.9}$$

where The throughput achieved by this allocation policy is

$$T_i^* = \sum_{r\in\mathcal{R}} p(r)x_i^*\left(r\right)r_i = \mathbf{E}\left[R_i x_i^*\left(R\right)\right] = \mathbf{E}\left[R_i|x_i^*\left(R\right)=1\right]\mathbf{P}\left(x_i^*\left(R\right)=1\right)$$

Note that

$$\mathbf{P}\left(x_i^*\left(R\right)=1\right) = \mathbf{P}\left(Y_i = \max_{j=1,\ldots,M}Y_j\right) = \frac{1}{M}$$

because $Y_1,\ldots,Y_M$ are i.i.d.. On the other hand

$$\begin{aligned}\mathbf{E}\left[R_i|x_i^*\left(R\right)=1\right] &= \mathbf{E}\left[\mathbf{S}_i Y_i Z|Y_i = \max_{j=1,\ldots,M}Y_j\right] = \mathbf{S}_i\mathbf{E}\left[Y_i|Y_i = \max_{j=1,\ldots,M}Y_j\right] \\ &= \mathbf{S}_i\mathbf{E}\left[\max_{j=1,\ldots,M}Y_j\right] = \mathbf{S}_i G(M)\end{aligned}$$

Hence we get (11.7). ■

Our complement

Proposition 53 *The throughput vector T^* given by (11.7) is optimal for the problems*

$$\max_{T \in \mathcal{A}} \min_{i=1,\ldots,M} \frac{T_i}{\mathbf{E}[R_i]} \tag{11.10}$$

and

$$\max_{T \in \mathcal{A}} \sum_{i=1}^{M} \log T_i \tag{11.11}$$

(This last statement justifies the name PROPORTIONAL FAIR *given to the HS-DPA algorithm.)*

Proof. By Lemma 19,

$$\sum_{i=1}^{M} \frac{T_i}{\mathbf{E}[R_i]} \leq \mathbf{E}\left[\max_{j=1,\ldots,M} \frac{R_j}{\mathbf{E}[R_i]}\right] = \mathbf{E}\left[\max_{j=1,\ldots,M} Y_j\right] = G(M) = \sum_{i=1}^{M} \frac{T_i^*}{\mathbf{E}[R_i]}$$

Hence the throughput vector T^* is optimal for Problem (11.8). Since

$$\frac{T_i^*}{\mathbf{E}[R_i]} = \frac{T_j^*}{\mathbf{E}[R_j]}, \quad \forall i, j = 1, ..., M$$

we deduce that T^* is also optimal for Problem (11.10).

We calculate

$$\sum_{i=1}^{M} \log T_i^* = \sum_{i=1}^{M} \log \frac{G(M)}{M} \mathbf{E}[R_i] = \sum_{i=1}^{M} \log \mathbf{E}[R_i] + M \log \frac{G(M)}{M}$$

For any achievable rate

$$\begin{aligned}
\sum_{i=1}^{M} \log T_i &= \sum_{i=1}^{M} \log \mathbf{E}[R_i x_i(R)] = \sum_{i=1}^{M} \log \mathbf{E}[\mathbf{S}_i Y_i Z x_i(R)] \\
&= \sum_{i=1}^{M} \log \mathbf{E}[R_i] + \sum_{i=1}^{M} \log \mathbf{E}[x_i(R) Y_i] \\
&\leq \sum_{i=1}^{M} \log \mathbf{E}[R_i] + \sum_{i=1}^{M} \log \mathbf{E}\left[x_i(R) \max_{j=1,\ldots,M} Y_j\right] \\
&= \sum_{i=1}^{M} \log \mathbf{E}[R_i] + M \log G(M) + \sum_{i=1}^{M} \log \mathbf{E}[x_i(R)]
\end{aligned}$$

Now observe that

$$\sum_{i=1}^{M} \log \mathbf{E}[x_i(R)] = M \frac{1}{M} \sum_{i=1}^{M} \log \mathbf{E}[x_i(R)] \leq M \log \frac{1}{M} \sum_{i=1}^{M} \mathbf{E}[x_i(R)] = M \log \frac{1}{M}$$

Therefore

$$\sum_{i=1}^{M} \log T_i \leq \sum_{i=1}^{M} \log T_i^*$$

■

Remark 28 *[79] shows that the Proportional Fair scheduling algorithm [118, §6.7.1] optimizes the problem*

$$\max_{T \in \mathcal{A}} \sum_{i=1}^{M} \log (T_i + d_i)$$

where the d_i, $i \leq N$ *are positive constants, which can be as small as we wish. This shows that the Proportional Fair scheduling will give a throughput vector* T^* *approximately given by (11.7).*

Chapter 12

GSM

12.1 Introduction

We are interested in FDMA networks such as GSM serving streaming traffic. We aim to extend to such networks the approch developped for CDMA networks such as UMTS.

A key feature is to study whether these so called feasibility-based admission control algorithms may be applied for FDMA networks. In this perspective, the key questions are the following:

- What are the feasibility condition of power allocation?
- What is loss of capacity induced by the sufficient feasibility condition compared to the necessary and sufficient feasibility condition?
- Are there approximate explicit expressions of the blocking probability (in the case of a simple model: regular hexagonal cells, homogeneous traffic demand, no shadowing, etc.)?
- Can we answer to questions specific to FDMA networks:
  - Is the radio constraint or the hard constraint more severe?
  - What is the optimal cluster size?

The objective of the present section is to answer to the above questions. We present here essentially the results for the downlink.

12.2 Model

We make the same assumptions as our previous studies on UMTS: hexagonal pattern of cells, modelisation of the networks on a torus, distance path loss law, no shadowing, no macrodiversity.

12.2.1 GSM versus UMTS

We enumerate in the present section the key differences between the two systems UMTS and GSM.

UMTS. Recall the properties of UMTS:

1. The multiple access scheme is CDMA.
2. The UMTS bandwidth equals $W = 5$MHz.
3. For voice, the bit energy to noise-density ratio threshold equals $\frac{E_b}{N_0} = 9$dB.
4. For voice, the signal to noise power ratio threshold is given by

$$\xi = \frac{r_v}{W'}\frac{E_b}{N_0} = 9\text{dB} - 25\text{dB} = -16\text{dB}$$

   where $W' = 3.84$Mchip/s designates the chip rate and $r_v = 12.2$kbit/s designates the voice bit rate. The noise power is denoted N.
5. The loss of orthogonality of the spreading sequences within a cell due to the multi-path is taken into account through the orthogonality factor α ($\alpha = 0$ for perfectly orthogonal). This multiplies the intra-cell interference. Typically $\alpha = 0.4$ in the downlink and $\alpha = 1$ in the uplink.
6. The transmission powers should not exceed a limit, denoted $\tilde{P}$, fixed by the authority of regulation. The powers of Dedicated CHannels (DCH) are controlled, while Common CHannels (CCH in the downlink) powers are constant. We assume typically that the common channel power is a fraction of the maximal power, say $\epsilon\tilde{P}$, where ϵ is given constant.

GSM. The above properties 1 to 5 in Section 12.2.1 for a UMTS network, are different for a GSM network:

1. The multiple access scheme is FDMA combined with TDMA[1].
2. The GSM bandwidth equals $W = 12.5$MHz. (We consider GSM 900 with bandwidth 25MHz shared equally between two operators.)
3. For voice, the bit energy to noise-density ratio threshold $\frac{E_b}{N_0} = 9$dB.
4. **FDMA**. The GSM bandwidth is divided into 62 FREQUENCY CHANNELS; each one having a bandwidth $W^* = 200$kHz. Among these frequency channels, 12 are used for the BCCH[2] and the remaining $\mathbf{W} = 50$ frequency channels are used for TCH[3]. In the case where there is no frequency hopping, the signal to noise power ratio threshold is given by $\xi^* = \frac{E_b}{N_0} = 9$dB.

[1]Time Division Multiple Access
[2]Broadcast Control CHannel
[3]Traffic CHannel

In the present study, we assume that there is frequency hopping and we will see later (§12.3.1) how to calculate the signal to noise power ratio threshold in this case.

When the distance between frequency channels equals $W^*, 2W^*, 3W^*$ the interference is reduced by $18, 50, 58$dB respectively. We neglect in the present study the interference between the different frequency channels.

TDMA. Time is divided into time slots. A time slot designates an interval of time of duration 0.577ms. There are 8 TDMA time slots in GSM (for more details cf. [92, p.195]).

5. We assume that there is no intra-cell interference.

6. The transmission powers should not exceed a limit, denoted $\tilde{P}$, fixed by the authority of regulation.

12.2.2 FDMA

The principle of FDMA is to increase the distance between the cells using the same frequencies in order to reduce the interferences. To do so, we divide the $\mathbf{W}$ available frequency channels into a given number, say χ, of subsets. This number χ is called the CLUSTER SIZE. Each subset comprises

$$\kappa = \frac{\mathbf{W}}{\chi}$$

frequency channels. We allocate a single subset to each base station and attempt to allocate the same subset to base stations which are as far as possible from each other. We show below that the cluster sizes can not take any integer value.

Hexagonal architecture. Considers a regular cellular network where each cell has the form of a hexagon. The hexagon centers are placed on a regular grid denoted on the complex plane by $\{\Delta_1(p + qe^{i\pi/3}) : (p, q) \in \{0, \pm1, \ldots\}^2\}$ where Δ_1 is the distance between two adjacent hexagons. Then the distance between a hexagon (p, q) and the origin is $\left(p^2 + pq + q^2\right)^{1/2} \Delta_1$. We define the CELL RADIUS, denoted R, as the radius of the virtual disc whose area is equal to that of the hexagon, that is

$$R = \Delta_1 \sqrt{\frac{\sqrt{3}}{2\pi}}$$

If we fix some hexagon center as the origin, the other hexagon center locations constitute successive RINGS which we index by $k = 1, 2, \ldots$. Figure 12.1 shows the first two rings $k = 1, 2$. Let Δ_k be the radius of ring $k = 1, 2, \ldots$. Let

$$\chi_k = (\Delta_k/\Delta_1)^2$$

which equals $p^2 + pq + q^2$ for some $(p, q) \in \{0, \pm1, \ldots\}^2$. Such a number is called RHOMBIC. Rhombic numbers can be characterized in an elegant way by their prime decomposition [83]. From [83, Theorem 1] we get the first rhombic numbers χ_k $(k = 1, 2, \ldots, 13)$.

k	1	2	3	4	5	6	7	8	9	10	11	12	13
$\chi_k = \left(\frac{\Delta_k}{\Delta_1}\right)^2$	1	3	4	7	9	12	13	16	19	21	25	27	28

First rhombic numbers χ_k $(k = 1, 2, \ldots, 13)$

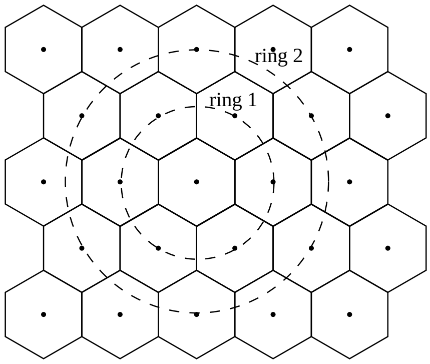

Figure 12.1: The first two rings $k = 1, 2$.

The rhombic numbers are precisely the possible cluster sizes in FDMA networks.

Hard and radio constraints. Given a cluster size χ, the distance between two base stations using the same frequencies, called REUSE DISTANCE, equals

$$\Delta = \Delta_1 \sqrt{\chi}$$

Compared to UMTS, the distance between two base stations using the same frequencies in GSM is multiplied by a factor $\sqrt{\chi}$. This will effectively reduce interference, but at the expense of the apparition of a new constraint: the number of simultaneous (i.e. per time slot) users in a cell, which we denote M, should not exceed the number κ of available frequency channel in the base station i.e.

$$M \leq \kappa = \frac{\mathbf{W}}{\chi} \tag{12.1}$$

We call this constraint HARD CONSTRAINT.

12.2.3 Traffic

We assume that there is no interference between the different TDMA time slots. Moreover we assume that an arriving call chooses a TDMA time slot among the 8 available slots with probability 1/8. So we may consider that each time slot serves 1/8 of the traffic demand and is independent from the other times slots. From now on we consider a given time slot.

New streaming calls arrive as a Poisson processes with intensity $\lambda \times dx$ in any region of surface dx. The call durations (if the call is not blocked) are i.i.d. with mean $1/\mu$ and independent from arrivals. We consider a single service class: voice. We always assume that $\lambda > 0, \mu > 0$. We consider a semi-dynamic model where each user is motionless during his call.

12.2.4 Frequency hopping

Since we neglect the interference between the different frequency channels and we consider motionless users, the sub-networks composed by base stations using the same frequencies may be treated independently from each other. Their evolutions are identically distributed. We consider a given sub-network which we view from now as our network to study. We index the set of frequency channels available at each base station by $0, 1, \ldots, \kappa - 1$ in increasing order. (Recall that κ designates the number of frequency channels available at each base station.)

We describe now the SLOW FREQUENCY HOPPING (SFH) in GSM [92, §4.2.2.3 p.223]. There are 64 hopping sequences indexed by the so called Hopping Sequence Number (HSN), say $H \in \{0, 1 \ldots, 63\}$. We denote by $\left\{F_0^H, F_1^H, \ldots\right\}$ the sequence of frequency channels associated to some hopping sequence number H. (i.e. F_s^H designates the frequency channel at time slot s). For a given $H \neq 0$, $\left\{F_0^H, F_1^H, \ldots\right\}$ is a sequence of independent and identically distributed random variables, having as common distribution the uniform distribution on the set of available frequency channels $\{0, 1, \ldots, \kappa - 1\}$. (The particular case $H = 0$ corresponds to the sequence where the frequencies are used in order, i.e. $F_s^0 = s \bmod \kappa$, for all $s \geq 0$.)

For a given base station, we associated a fixed HSN $H \neq 0$. There are at most κ simultaneous users served by the base station. We associate to each user a specific Mobile Allocation Index Offset (MAIO), say $A \in \{0, 1, \ldots, \kappa - 1\}$. The FREQUENCY HOPPING SEQUENCE of that user is then $\left\{\left(F_0^H + A\right) \bmod \kappa, \left(F_1^H + A\right) \bmod \kappa, \ldots\right\}$. By this way, two users in the same cell never use simultaneously the same frequency.

The frequency hopping sequences associated to two distinct (non null) hopping sequence numbers H, H', are independent, i.e. the sequences $\{F_s^H\}_{s \geq 0}$ and $\{F_s^{H'}\}_{s \geq 0}$ are independent. Then $\{\mathbf{1}\{F_s^H = F_s^{H'}\}\}_{s \geq 0}$ is a sequence of independent and identically distributed random variables, having as common distribution the Bernoulli distribution with parameter

$$\begin{aligned} \mathbf{P}\left(\mathbf{1}\{F_s^H = F_s^{H'}\} = 1\right) &= \mathbf{P}(F_s^H = F_s^{H'}) \\ &= \frac{1}{\kappa} \end{aligned}$$

12.3 Power allocation feasibility

The performance of a wireless cellular network is closely related to the performance of the admission control algorithm implemented in the network.

If the network admits a new user, it must ensure to him a certain quality of service and should preserve the quality of service of other users already present. This quality of service can be expressed in term of signal-to-interference ratio. The network must thus ensure the existence of an allocation of powers to the users respecting the signal-to-interference constraints. Moreover, the powers should not exceed a limit fixed by the authority of regulation. This defines the POWER ALLOCATION PROBLEM. We say that this problem is FEASIBLE if admits a solution. The feasibility condition will be called RADIO CONSTRAINT in order to distinguish it from the hard constraint already described (Inequality (12.1)).

A new streaming user will be admitted if the radio and hard constraints are respected. Such algorithm is called ADMISSION CONTROL algorithm.

The NECESSARY AND SUFFICIENT CONDITION (NSFC) of feasibility induces an admission control algorithm which has optimal performances; but which is unfortunately not practicable in the field, because it requires data coming from the users in all the network. It will be said that it is a CENTRALIZED algorithm.

In [13] we build useful feasibility conditions for the power allocation problem. For UMTS in the downlink, we propose a sufficient feasibility condition which depends on the users in the cell serving (potentially) the new user, but not other users in the network. We say that this condition is DECENTRALIZED and we note it FC. In the uplink we propose also a SUFFICIENT FEASIBILITY CONDITION, that we note SFC, which is unfortunately not decentralized. Therefore we propose a decentralized feasibility condition that we note FC. Hence the term FC designates the decentralized FEASIBILITY CONDITION either for the downlink or for the uplink.

12.3.1 Power allocation problem

We assume that two distinct cells use two different (non null) hopping sequence numbers. This may be easily fulfilled when then number of cells in the networks is not larger than 63. If this is not the case, we reuse the same hopping sequence number for base stations sufficiently far from each other to neglect their mutual interference.

Let $\mathbf{U}$ be the set of base stations using the same frequencies and fix some base station $u \in \mathbf{U}$ and some user $m \in u$ (i.e. m is served by base station u).

Let $P_{u,m}$ designate the power transmitted by base station u to user m and $L_{u,m}$ designate the path-loss between u and m. Let $\{F_s^{H_m}\}_{s \geq 0}$ be the hopping sequence of user m.

The power received by user m from its serving base station u is $P_{u,m}/L_{u,m}$. The interference due to another user n served by a base station $v \neq u$ at time slot s is $\mathbf{1}\{F_s^{H_m} = F_s^{H_n}\} P_{v,n}/L_{v,m}$. By the law of large numbers, the average interference over a time interval comprising a large number of time slots will be $\mathbf{E}\left[\mathbf{1}\{F_s^{H_m} = F_s^{H_n}\}\right] P_{v,n}/L_{v,m} = \frac{1}{\kappa} P_{v,n}/L_{v,m}$. Hence the power allocation problem takes the form

$$\frac{P_{u,m}/L_{u,m}}{N^* + \sum_{v\in\mathbf{U}\setminus\{u\}} \frac{1}{\kappa} \sum_{n\in v} P_{v,n}/L_{v,m}} \geq \xi^*, \quad m \in u \in \mathbf{U}$$

where N^* and ξ_m^* designates the noise power and the signal to noise ratio threshold respectively in a single frequency channel. (The arguments to establish the above result, and in particular the fact that interference may be viewed as noise, are similar to those concerning CDMA in [118, §4.3.1].)

We may rearrange the above inequality as follows

$$\frac{P_{u,m}/L_{u,m}}{\kappa N^* + \sum_{v\in\mathbf{U}\setminus\{u\}} \sum_{n\in v} P_{v,n}/L_{v,m}} \geq \frac{\xi_m^*}{\kappa}, \quad m \in u \in \mathbf{U}$$

Hence we obtain a problem for GSM which has a similar form to the power allocation problem for UMTS. The only differences are in the parameter values. For GSM, the orthogonality factor is null, the noise power is $N = \kappa N^*$ and the signal to noise power ratio threshold is $\xi := \xi^*/\kappa$. Note that the above inequality may be written as follows

$$\frac{P_{u,m}/L_{u,m}}{N + \sum_{v\in\mathbf{U}\setminus\{u\}} \sum_{n\in v} P_{v,n}/L_{v,m}} \geq \xi_m, \quad m \in u \in \mathbf{U}$$

We introduce for future reference $N^{(1)} = \mathbf{W}N^*$ and $\xi^{(1)} = \xi^*/\mathbf{W}$ corresponding to a cluster size $\chi = 1$, and write $N = N^{(1)}/\chi$ and $\xi = \chi\xi^{(1)}$ which shows explicitly the dependence of the noise power and the signal to noise power ratio threshold on the cluster size.

The above analysis shows that the feasibility conditions of the power allocation problem in CDMA networks [13] may be extended to FDMA networks.

12.3.2 FC

In particular, the FC takes the form

$$\sum_{m\in u} \varphi(m) \leq C$$

where, for the downlink,

$$C = 1, \quad \varphi(m) = \left[f(m) + NL_{u,m}/\tilde{P}\right]\xi_m, \quad m \in u$$

and the so called F-FACTOR $f(m)$ is given by

$$f(m) = \sum_{v\in\mathbf{U}\setminus\{u\}} L_{u,m}/L_{v,m}$$

In order to make notations more coherent with UMTS, we denote

$$\alpha = 0, \quad \xi'_m = \frac{\xi_m}{1 + \alpha\xi_m}, \quad \epsilon = 0$$

(We may account for some intra-cell interference due to adjacent channels by taking an orthogonality factor $\alpha \neq 0$ and some adjacent interference from common channels by taking $\epsilon \neq 0$.) Hence

$$C = 1 - \epsilon, \quad \varphi(m) = \left[\alpha + f(m) + NL_{u,m}/\tilde{P}\right] \xi'_m, \quad m \in u$$

A new constraint appears in FDMA networks: the number of users in a cell per time slot should not exceed the number of available frequency channel in the base station. We call this constraint *hard constraint* in order to distinguish it from the power allocation feasibility condition, which we call *radio constraint*. The admission control algorithm includes both the hard and the radio constraint.

12.4 Numerical results

We compare the performances of FC and NSFC in a GSM network. The result depends on cell radius, maximal power and cluster size. The loss of capacity of FC compared to NSFC decreases with the cluster size χ (this shows that the FC is well adapted for FDMA networks with high cluster sizes). For a given cluster size χ, the loss of capacity has the same order of magnitude for different cell radii R. Assuming usual values of maximal power and cell radius, the loss of capacity is about 60% for $\chi = 1$ and about 40% for $\chi = 2$. For $\chi \geq 3$, the loss of capacity is less than 30% if we consider only the radio constraint and negligible if we consider both the radio and the hard constraints.

We build explicit approximate expression of the blocking probability for FC in FDMA networks, and compare it with the blocking probability given by dynamic simulation.

Even though the loss of capacity of the FC compared to the NSFC is sometimes large (especially for the cluster size 1) and the Erlang approximation for the FC is not always accurate, this approximation permits to give an interesting insight on the functioning of a GSM network.

We use the Erlang approximation of the blocking probability to study analytically the following question. Which constraint is more severe: the radio or the hard constraint? We show that the answer depends on the three fundamental parameters: cell radius, maximal power and cluster size.

The investigation of the above question permits to get the following result. For usual values of cell radius and maximal power in GSM networks, the optimal cluster size is between 2 and 3.

12.5 Conclusion

The differences between a CDMA/UMTS and a FDMA/GSM network are identified. We show that sufficient conditions (FC) for the feasibility of the power allocation problem proposed initially for CDMA networks may be extended to

FDMA networks. Our analytical approch permits to investigate questions specific to FDMA networks such as determining whether the radio constraint or the hard one is more severe; or determining the optimal cluster size.

Part IV

Conclusion

We show that the analytic evaluation of the performance of wireless cellular networks is possible, but it requires to use tools and theoretical issues from several disciplines.

Feasibility based load control. The first step is to characterize the single link performance which is fulfilled by looking at the literature analyzing the performance of the multi-path fading by the digital communication techniques. These techniques analyze the performance of the modulation seen as a detection process by expressing the bit error-probability as function of the signal energy-per-bit over noise-density ratio, say E_{b}/N_0. For vehicular speed users (or scatters) the E_{b} is averaged over multi-path fading. For streaming traffic, one fixes the error probability and deduces the corresponding E_{b}/N_0 threshold, and hence the signal to noise power ratio, say $\mathbf{S}/N$. For elastic services, we get a relation between the bit-rate, say r, and the $\mathbf{S}/N$.

The interference between the links is made by formulating the power allocation problem for both the downlink and the uplink. The necessary and sufficient condition for the feasibility (NSFC) of this problem is centralized, and hence difficult to implement. We propose some decentralized conditions (FC) which are related to the feasibility of power allocation. In fact if each base station applies FC, then the global power allocation problem is certainly feasible (with high probability for uplink with power limit).

We introduce a mean model permitting to define precisely the classical notion of pole capacity. This simple notion permits in particular to illustrates the phenomenon called cell breathing.

We evaluate the performance of the FC in terms of the infeasibility probability, defined as the probability that the FC doesn't hold for a given cell when modelling the users as a Poisson process. This notion is closely related to the feasibility of the power allocation which is not the case of the classical outage probability. In certain cases, for example for a Poisson pattern of users in an hexagonal network of base stations, we gives explicit approximate formulae for the infeasibility probability.

We propose admission control schemes for streaming traffic and congestion control schemes for elastic ones based on the FC. These schemes are based on an exact representation of the geometry of both the downlink and the uplink channels and ensure[4] that the associated power allocation problems have solutions under constraints on the maximal power of each station/user. These schemes are decentralized in that they can be implemented in such a way that each base station only has to consider the load brought by its own users to decide on admission. By load we mean here some function of the configuration of the users and of their bit rates that is described in the report. When implemented in each base station, such schemes ensure[5] the global feasibility of the power allocation even in a very large (infinite number of cells) network.

We show that the performance of the FC admission control is close to the

[4]certainly for the downlink and with high probability for the uplink

[5]Cf. previous note

optimum (corresponding to the NSFC admission control). Moreover the constructor algorithms perform better than the FC in some cases (small cells and small blocking) and worst in others (large cells or large blocking). The FC induces a loss of capacity versus the NSFC which not larger than 25%.

From the user's point of view, the performance is more suitably evaluated by the mean of the blocking and cut probabilities of streaming users and the delay and throughput of elastic users in the long run of the network.

Spatial Markov queueing process. We take into account the arrivals, mobility and departures of the users by the mean of a pure-jump Markov process (on the space of point measures on some Polish space). The user locations are represented by a spatial point process that evolves over time as users arrive, move or depart. We call our process spatial Markov queuing process (SMQ) or equivalently spatial birth, mobility and death Markov process. Conditions for regularity of the generator (i.e., uniqueness of the associated Markov process) as well as for its ergodicity are established. In some cases, we show that the stationary distribution is a Gibbs measure.

Analytic performance evaluation. We apply our general results for SMQ to establish the explicit dynamic performance of CDMA wireless cellular networks serving streaming traffic and elastic traffic. In the case of streaming traffic we study two spatial loss models and prove an expression for blocking rates. In particular we show that the infeasibility probability plays an important role in calculating the blocking probability. We apply the spatial Erlang formula and the analytic expression of the infeasibility probability to built an approximate explicit expression of the blocking probability of the FC admission control. In the case of elastic services, we build a model for wireless cellular networks serving elastic traffic and we build explicit analytic expressions for the delay and throughput of the FC congestion control.

Our objectives (Section 0.2) are then attained by the help of tools from different disciplines: digital communication for characterizing the single link performance, linear algebra for building the feasibility based load control schemes, informatics (simulation) for evaluating and comparing numerically the load control schemes, stochastic geometry and Markov queueing processes extended to the spatial case for evaluating analytically the performance of load control in wireless cellular networks.

12.6 Summary of Contributions

Feasibility based load control. We build rapid, scalable and efficient load control schemes by extending the work in [16] from the downlink case without power limit, to both the downlink and the uplink taking into account the maximal power limit for base stations and users respectively. These schemes are the subject of the patents [12] and [11].

Spatial Markov queueing process. Our spatial Markov queueing process might be seen as a generalization of the spatial birth-and-death generator, which allows for mobility of particles. Hence our work on the regularity, ergodicity and invariant measure of the SMQ process extends previous work in [97] by allowing for mobility of particles. Our spatial birth-mobility-and-death process can be seen also as a generalization of the spatial queueing system considered in [106].

Analytic performance evaluation. The combination of the power allocation feasibility condition and our SMQ allows the model to include the exact representation of the geometry of inter-cell and intra-cell interferences, which play an essential role in wireless cellular networks.

We build analytical methods for evaluating the performance of large cellular networks controlled by feasibility based schemes with indicators which are relevant from the user's point of view (blocking, cut delay, and throughput) rather than the classical outage probability. Our formula for the blocking probability might be seen as a spatial extension of the well-known Erlang loss formula.

In the case of elastic services, we build a model for wireless cellular networks serving elastic traffic more suitable than the classical so called Whittle model. In our model, the users move independently form the congestion in the network whereas the users motion is frozen when the network approaches instability in the Whittle model.

The analytic performance evaluation methods permit to build a *new class* of *coherent* methods for the different fundamental problems in wireless cellular systems: quality of service, capacity and dimensioning. The new methods are rapid and accurate, even for large networks. This approach is complementary to simulations. The ease of use of the analytical expressions makes this type of approach more effective than simulations for macroscopic evaluation and optimization. Our methods are implemented in a tool for Orange operator.

12.7 Future Research

For elastic services

- Take into account the *maximal instantaneous bit-rate* per user (by truncation or using the technique in [26, §5])
- Study the conjectures about the stability, monotonicity and performance for *finite non-null speeds*
- Approximate more accuretly the *throughput* for *finite non-null speeds*
- Investigate the effect of a *non linear* relation between bit-rate and signal to noise power ratio
- Compare FC to NSFC and constructor algorithms

For streaming services

- Extend the approach to other cellular systems such as *LTE* wireless networks using *OFDM*
- Study the *Shadowing* effect
- Study the *transmitted power*
- Approximate more accuretly the *cut probability*

Part V

Appendices

Chapter 13

Appendix of Part I

The present chapter comprises appendices of Part I.

Appendix 13.A gives basic results permitting to establish the NSFC and the FC in the case where the size of the network is finite. This Appendix contains well known results which are reproduced here to make the report more self-contained.

Appendix 13.B studies a quantity called f-factor which depends on the geographic position of the user and the base stations in the networks. The f-factor is well known in the previous literature and recognized as a crucial in analyzing the performance of wireless cellular networks. We will show that the f-factor is crucial in analyzing the FC performance. Unfortunately no analytic (even approximate) expressions of the f-factor exist, even in the hexagonal simplified model. We establish such analytic approximation in Appendix 13.B.

In Appendix 13.C we study the f-factor properties and extend the mean model and the infieasibility probability formulae to the case with directional antennas.

Appendix 13.D summarizes the formulae for the analytic approximate evaluation of the FC performance in a semi-static context.

In Appendix 13.E we compare the FC admission control to the optimal scheme as well as to the constructor's schemes in a *semi-dynamic* traffic model.

13.A Power allocation basic results

13.A.1 Power allocation problem

We consider only finite dimension matrices and vectors. For a non-negative matrix A, and a positive vector a, the Power allocation problem is the to find a vector y such that

$$\begin{cases} (\mathbf{1} - A)\, y \geq a \\ y \geq 0 \end{cases} \tag{13.1}$$

13.A.2 Feasibility

Theorem 1 *For a real matrix* $B = [B_{ij}]_{i=1,\ldots m, j=1,\ldots n}$, *and two vectors* $a = (a_1, \ldots, a_n)^T > 0$, $c = (c_1, \ldots, c_m)^T \geq 0$ *exactly one of the following problems*

$$\begin{cases} By \geq a \\ y \geq c \end{cases}$$

and

$$\begin{cases} B^T x \leq 0 \\ x \geq 0, x \neq 0 \end{cases}$$

is FEASIBLE *(i.e. admits a solution).*

Proof. By the so-called *Theorem of alternatives* [119, Theorem 2.1], exactly one of the problems

$$\begin{pmatrix} -B \\ -\mathbf{1}_m \end{pmatrix} y \leq \begin{pmatrix} -a \\ -c \end{pmatrix}$$

and

$$\begin{cases} \left(-B^T, -\mathbf{1}_m\right) \begin{pmatrix} x \\ z \end{pmatrix} = 0 \\ \begin{pmatrix} -a^T & -c^T \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} < 0 \\ x \geq 0, z \geq 0 \end{cases}$$

is feasible. The first problem may be written

$$\begin{cases} By \geq a \\ y \geq c \end{cases}$$

The second problem may be written as follows

$$\begin{cases} B^T x = -z \\ -a^T x - c^T z < 0 \\ x \geq 0, z \geq 0 \end{cases}$$

As $a > 0$ and $c \geq 0$ we have $z \geq 0, x \geq 0, x \neq 0 \Rightarrow -a^T x - c^T z < 0$. On the other hand, $-a^T x - c^T z < 0 \Rightarrow x \neq 0$. The the second problem is equivalent to

$$\begin{cases} B^T x \leq 0 \\ x \geq 0, x \neq 0 \end{cases}$$

This finishes the proof. ■

Theorem 2 *For a real matrix* A, *and two vectors* $a > 0$ *and* $c \geq 0$, *exactly one of the following problems*

$$\begin{cases} (\mathbf{1} - A)\, y \geq a \\ y \geq c \end{cases}$$

and

$$\left(\mathbf{1} - A^T\right) x \leq 0, x \geq 0, x \neq 0$$

is feasible.

Proof. Particular case of the previous theorem. Take $B = \mathbf{1} - A$. ■

Definition 5 *For a square matrix A, the set of* EIGENVALUES *is defined as follows*

$$\sigma(A) = \{\lambda \in \mathbb{C}; \det(A - \lambda \mathbf{1}) = 0\}$$

The number

$$\boldsymbol{\rho}(A) = \max_{\lambda \in \sigma(A)} |\lambda|$$

is called the SPECTRAL RADIUS *of A.*

Theorem 3 *For a non-negative matrix A, the problem*

$$(\mathbf{1} - A)\, x \le 0, x \ge 0, x \ne 0$$

is feasible iff $\boldsymbol{\rho}(A) \ge 1$.

Proof. Denote $\mathcal{N} = \{x | x \ge 0 \text{ with } x \ne 0\}$ and $f(x) = \min_{1 \le m \le M, x_m \ne 0} [Ax]_m / x_m$. From [85, §8.3] we have the following results:

1. The *Collatz–Wielandt formula* for the non-negative matrix A gives $\boldsymbol{\rho}(A) = \max_{x \in \mathcal{N}} f(x)$.
2. There exists some $z \in \mathcal{N}$, such that $Az = \boldsymbol{\rho}(A)\, z$.

If $\boldsymbol{\rho}(A) \ge 1$, then z is a solution to our problem. Inversely, if x is a solution to our problem, then $f(x) \ge 1$, hence $\boldsymbol{\rho}(A) \ge 1$. We conclude that our problem is feasible iff $\boldsymbol{\rho}(A) \ge 1$. ■

Theorem 4 *For a non-negative matrix A, and two vectors $a > 0$ and $c \ge 0$, the problem*

$$\begin{cases} (\mathbf{1} - A)\, x \ge a \\ x \ge c \end{cases}$$

is feasible iff $\boldsymbol{\rho}\left(A^T\right) < 1$.

Proof. The result is deduced from the two previous theorems. ■

The previous result is well known when $c = 0$ (see for example [16] and the references therein). The proof we give here is based on a theorem of alternatives and Collatz–Wielandt formula. It is surprising that the feasibility condition is independent of a and c. The classical literature on power allocation in cellular systems assumes that the matrix A is irreducible (see for example [90] and [19]). This assumption is not necessary as we have seen in the proof.

Theorem 5 *For a non-negative matrix A, and a vector $c \ge 0$, the problem*

$$\begin{cases} (\mathbf{1} - A)\, x > 0 \\ x \ge c \end{cases}$$

is feasible iff $\boldsymbol{\rho}\left(A^T\right) < 1$.

Proof. Suppose that $\boldsymbol{\rho}\left(A^T\right) < 1$, then take an arbitrary $a > 0$. The problem

$$\begin{cases} (\mathbf{1} - A)\, x \geq a \\ x \geq c \end{cases}$$

is feasible, then our problem is feasible. Suppose now that our problem is feasible. Denote y a solution. Denote $a = (\mathbf{1} - A)\, y$ which is positive. Then y is a solution of the problem

$$\begin{cases} (\mathbf{1} - A)\, x \geq a \\ x \geq c \end{cases}$$

This problem is feasible, then $\boldsymbol{\rho}\left(A^T\right) < 1$. ■

13.A.3 Minimal solution

For a non-negative matrix A, and a positive vector a, we will prove that the problem of finding a vector y such that

$$\begin{cases} (\mathbf{1} - A)\, y \geq a \\ y \geq 0 \end{cases}$$

is feasible iff the following problem

$$\begin{cases} (\mathbf{1} - A)\, y = a \\ y \geq 0 \end{cases} \tag{13.2}$$

called LIONTIEF-INPUT-OUTPUT problem is feasible. We will prove then that the solution of Liontief-input-output problem is unique. This solution is called the MINIMAL SOLUTION of the problem at hand.

Theorem 6 *For a matrix $A \in \mathbb{C}^{n\times n}$,*

$$\sigma\left(A^T\right) = \sigma(A)$$

and in particular

$$\boldsymbol{\rho}\left(A^T\right) = \boldsymbol{\rho}(A)$$

Proof. The transpose does not alter the determinant so that

$$\det(A - \lambda\mathbf{1}) = \det\left(A^T - \lambda\mathbf{1}\right)$$

■

Theorem 7 *For a matrix $A \in \mathbb{C}^{n\times n}$, the following statements are equivalent.*

1. *The Newmann series $\sum_{k=0}^{+\infty} A^k$ converges.*
2. $\boldsymbol{\rho}(A) < 1$.

3. $\lim_{k\to+\infty} A^k = 0$.

Moreover, if one of the above statements is true, then $(\mathbf{1}-A)^{-1}$ *exists and* $(\mathbf{1}-A)^{-1} = \sum_{k=0}^{+\infty} A^k$.

Proof. [85, §7.10]. ■

Theorem 8 *For a non-negative matrix* A, $\boldsymbol{\rho}(A) < 1$ *iff* $(\mathbf{1}-A)^{-1}$ *exists and* $(\mathbf{1}-A)^{-1}$ *is non-negative.*

Proof. [85, Example 7.10.3]. ■

Theorem 9 *For a non-negative matrix* A*, and a positive vector* a*, the problem*

$$\begin{cases} (\mathbf{1}-A)\,x = a \\ x \geq 0 \end{cases}$$

is feasible iff $\boldsymbol{\rho}(A) < 1$*. In this case there is a unique solution is* $(\mathbf{1}-A)^{-1} a$.

Proof. If the problem is feasible then, the problem

$$\begin{cases} (\mathbf{1}-A)\,x \geq a \\ x \geq 0 \end{cases}$$

is feasible. Then $\boldsymbol{\rho}(A) < 1$. Suppose now that $\boldsymbol{\rho}(A) < 1$. Then $(\mathbf{1}-A)^{-1}$ exists and $(\mathbf{1}-A)^{-1}$ is non-negative. Then the problem

$$\begin{cases} (\mathbf{1}-A)\,x = a \\ x \geq 0 \end{cases}$$

is feasible. ■

Corollary 12 *For a non-negative matrix* A*, and a positive vector* a*, the problem of finding a vector* y *such that*

$$\begin{cases} (\mathbf{1}-A)\,y \geq a, \\ y \geq 0, \end{cases} \tag{13.3}$$

is feasible iff the following problem

$$\begin{cases} (\mathbf{1}-A)\,y = a \\ y \geq 0 \end{cases}$$

is feasible, which is equivalent to

$$\boldsymbol{\rho}(A) < 1 \tag{13.4}$$

In this case, $y^* = (\mathbf{1}-A)^{-1} a$ *is the minimal solution of the power allocation problem (13.3) (i.e. each other solution* y *satisfies* $y \geq y^*$*). Moreover*

$$y^* = (\mathbf{1}-A)^{-1} a = \sum_k A^k a$$

and the iterates $x(k)$ *defined by* $x(k+1) = Ax(k) + a$ *converge to the minimal solution* $x = (\mathbf{1}-A)^{-1} a$ *for any starting vector* $x(0)$*. This is called Stationary Iterative Method.*

Proof. We have showed that each of these problems is feasible iff $\boldsymbol{\rho}(A) < 1$ which proves that one of these problems is feasible iff the other is also feasible. Suppose now that $\boldsymbol{\rho}(A) < 1$. We have see that in this case $(\mathbf{1} - A)^{-1}$ is non-negative. Consider some solution y of the first problem (with inequalities). We have

$$(\mathbf{1} - A)\, y \geq a$$

As $(\mathbf{1} - A)^{-1}$ is non-negative, we have

$$(\mathbf{1} - A)^{-1} (\mathbf{1} - A)\, y \geq (\mathbf{1} - A)^{-1} a$$

which give

$$y \geq (\mathbf{1} - A)^{-1} a$$

■

Corollary 13 *For a non-negative matrix A, and a positive vector a, the problem of finding a vector y such that*

$$\begin{cases} (\mathbf{1} - A)\, y \geq a \\ 0 \leq y \leq c \end{cases} \tag{13.5}$$

is feasible iff the following problem

$$\begin{cases} (\mathbf{1} - A)\, y = a \\ 0 \leq y \leq c \end{cases} \tag{13.6}$$

is feasible, which is equivalent to

$$\boldsymbol{\rho}(A) < 1 \quad \textit{and} \; (\mathbf{1} - A)^{-1} a \leq c \tag{13.7}$$

In this case, $y^* = (\mathbf{1} - A)^{-1} a$ is the minimal solution of the power allocation problem (13.5).

Proof. Suppose that Problem (13.5) is feasible and let y be a solution. From Theorem 4 we deduce that $\boldsymbol{\rho}(A) < 1$. In this case $(\mathbf{1} - A)^{-1}$ exists and is non-negative. Then

$$(\mathbf{1} - A)^{-1} (\mathbf{1} - A)\, y \geq (\mathbf{1} - A)^{-1} a$$

which give

$$y \geq y^* = (\mathbf{1} - A)^{-1} a$$

Since $y \leq c$ we deduce $y^* \leq c$, then y^* is solution of Problem (13.6). The rest of the proof is evident. ■

Particular cases.

Theorem 10 *Consider some positive real numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$ and*

$$A = \begin{pmatrix} \gamma_1 & \gamma_1 & \cdots & \gamma_1 \\ \gamma_2 & \gamma_2 & \cdots & \gamma_2 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_n & \gamma_n & \cdots & \gamma_n \end{pmatrix}$$

The problem

$$\begin{cases} (\mathbf{1} - A)\, x = a \\ x \geq 0 \end{cases}$$

is feasible iff $\sum_i \gamma_i < 1$. Moreover, in this case

$$x_j = a_j + \gamma_j \sum_i a_i \Big/ \left(1 - \sum_i \gamma_i\right)$$

is the solution.

Proof. The spectral radius of a matrix is between its minimal column sum and maximal column sum. Equivalently,

$$\min_j \sum_i A_{ij} \leq \boldsymbol{\rho}(A) \leq \max_j \sum_i A_{ij}$$

Since the column sums of the matrix A are all equal to $\sum_i \gamma_i$, we deduce that $\boldsymbol{\rho}(A) = \sum_i \gamma_i$. Suppose from now that $\sum_j \gamma_j < 1$. Let's search the solution. Our problem may be written as

$$x_j - \gamma_j \sum_i x_i = a_j$$

Adding over j gives

$$\sum_j x_j - \sum_j \gamma_j \sum_i x_i = \sum_j a_j$$

This may be written

$$\left(1 - \sum_j \gamma_j\right) \sum_j x_j = \sum_j a_j$$

Then $x_j = a_j + \gamma_j \sum_i a_i / \left(1 - \sum_i \gamma_i\right)$ is the solution of our problem. ■

Theorem 11 *Consider some positive real numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$ and*

$$A = \begin{pmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_n \\ \gamma_1 & \gamma_2 & \cdots & \gamma_n \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1 & \gamma_2 & \cdots & \gamma_n \end{pmatrix}$$

The problem

$$\begin{cases} (\mathbf{1} - A)\, x = a \\ x \geq 0 \end{cases}$$

is feasible iff $\sum_i \gamma_i < 1$*. Furthermore, in this case*

$$x_j = a_j + \sum_i \gamma_i a_i / \left(1 - \sum_i \gamma_i\right)$$

is the unique solution.

Proof. The spectral radius of a matrix is between its minimal row sum and maximal row sum. Equivalently,

$$\min_j \sum_i A_{ij} \leq \boldsymbol{\rho}(A) \leq \max_j \sum_i A_{ij}$$

Since the row sums of the matrix A are all equal to $\sum_i \gamma_i$, we deduce that $\boldsymbol{\rho}(A) = \sum_i \gamma_i$. Suppose from now that $\sum_j \gamma_j < 1$. Let's search the solution. Our problem may be written as

$$x_j - \sum_i \gamma_i x_i = a_j$$

Multiplying by γ_j and then adding over j gives

$$\sum_j \gamma_j x_j - \sum_j \gamma_j \sum_i \gamma_i x_i = \sum_j \gamma_j a_j$$

This may be written

$$\left(1 - \sum_j \gamma_j\right) \sum_j \gamma_j x_j = \sum_j \gamma_j a_j$$

Then $x_j = a_j + \sum_i \gamma_i a_i / (1 - \sum_i \gamma_i)$ is the solution of our problem. ■

Theorem 12 *Let A be a $n \times n$ non-negative matrix A and a be a positive vector with constant coordinates. Suppose that the row sums $\sum_{j=1}^n A_{ij}$ are independent of i and less than 1. Then the problem*

$$\begin{cases} (\mathbf{1} - A)\, x = a \\ x \geq 0 \end{cases}$$

has a solution with constant coordinates $x_1 = x_2 = \ldots = x_n = a_1 / \left(1 - \sum_{j=1}^n A_{1j}\right)$.

Proof. The vector x constituted of equal coordinates $a_1 / \left(1 - \sum_{j=1}^n A_{1j}\right)$ is clearly the solution of the problem. ■

13.B The f-factor

In the present section, we are interested in a parameter, called f-factor, which plays an important role in cellular networks. We consider the model presented in Section 2.2 and we will use the notation described in Section 2.3.

Recall in particular that the cell radius R is related to the distance Δ between two adjacent base stations by Equation (2.6). Recall moreover that the set of the base stations is denoted by $\mathbf{U}$.

Definition 6 *We call* F-FACTOR *the following parampeter*

$$f(m) = \sum_{v \in \mathbf{U} \setminus \{u\}} L_{u,m}/L_{v,m}, \quad m \in u$$

Observe that the f-factor may be written as follows

$$f(m) = \left(\sum_{v \in \mathbf{U} \setminus \{u\}} L_{v,m}^{-1} \right) L_{u,m}$$

The first term in the right-hand side of the above display is the sum of the propagation-gains (inverse of propagation-loss) of interferers whereas the second term is the propagation-loss of the serving base station. Hence the f-factor is the signal to interference ratio, if all the base station transmit at the same power.

We aim to find simple closed form expressions and study the properties of the f-factor for the hexagonal cell model. We consider first omni antennas.

No simple closed form expressions nor approximations are known for the f-factor. Typical values are generally calculated from simulations [88, 96].

The constant K in the propagation-loss expression $L(\mathbf{r}) = (K\mathbf{r})^\eta$ has no effect on the f-factor. Hence we take in the present section $K = 1$.

13.B.1 Gain-sum

We focus our attention on the propagation-GAIN-SUM

$$I_{\eta,\Delta}(m) = \sum_{v \in \mathbf{U} \setminus \{u\}} L_{v,m}^{-1} \tag{13.8}$$

This depends on the user position m, the propagation exponent η and the distance Δ between two adjacent base stations. Let the location of the serving base station u be the origin of the coordinate system. We can associate to each base station a complex number $v(p,q) = \Delta\left(p + qe^{i\pi/3}\right)$. Hence

$$I_{\eta,\Delta}(m) = \sum_{p,q} |v(p,q) - m|^{-\eta} \tag{13.9}$$

where the summation is over $p \in \mathbb{Z}$ and $q \in \mathbb{Z}$ with $p \neq 0$ or $q \neq 0$. It is easy to see that

$$I_{\eta,\Delta}(m) = \Delta^{-\eta} I_{\eta,1}(m/\Delta) \tag{13.10}$$

Hence we may limit our study to the case $\Delta = 1$km. For simplicity we denote $I_{\eta,1}$ simply by I_η.

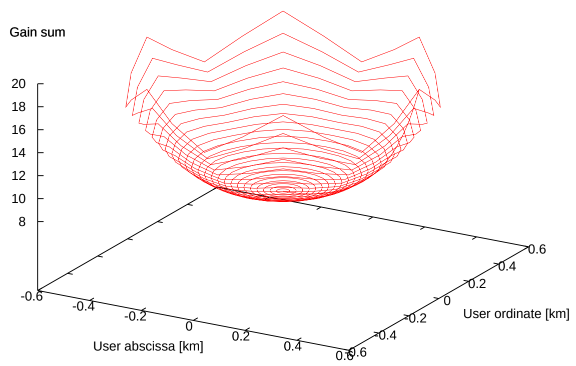

Figure 13.1: Gain-sum versus user location.

Gain-sum versus user location

Gain-sum versus angle.

Approximation 2 *The gain-sum depends only on the distance between the user and its serving base station. This approximation is more accurate when the propagation exponent η is small.*

Proof. Figure 13.1 represents the gain-sum I_η as a function of the user location $(x, y) \in \mathbb{R}^2$. In order to interpret this figure, let $(\mathbf{r}, \theta)$ be the user polar coordinates. First note that, due to the hexagonal model properties, the gain-sum is periodic in the angle θ with a period equal to $\pi/3$. Moreover, Figure 13.1 shows that for a given user-to-base-station distance $\mathbf{r}$, the gain-sum is maximum for $\theta = 0$ and minimum for $\theta = \pi/6$.

To gain more insight on this behavior, we plot in Figure 13.2 the gain-sum I_η for these two values of θ. This figure shows that the gain-sum is more sensitive to the angle θ when the user is far from its serving base station.

Figure 13.3 represents the ratio $I_\eta(R)/I_\eta\left(Re^{i\pi/6}\right)$ as function of the propagation exponent η. This figure shows that the gain-sum is more sensitive to the angle θ

The above results justify the approximation that the gain-sum is independent of the angle θ. We may use the values of the gain-sum for $\theta = 0$ in the calculation. This leads to a slight overestimation of the gain-sum (and consequently a slight underestimation of the capacity). ■

Gain-sum versus distance. We shall now study the gain-sum as function of the user-to-base-station distance $\mathbf{r}$. We observe in Figure 13.1 that the gain-sum is minimal at $\mathbf{r} = 0$ and maximal at $\mathbf{r} = R$. In order to study the dynamics of

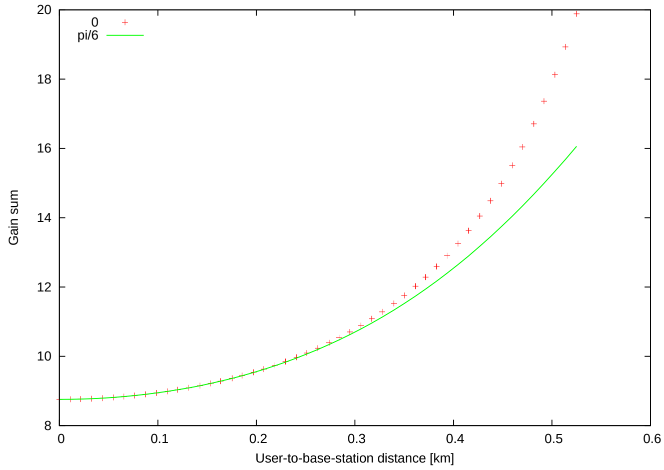

Figure 13.2: Gain-sum versus user-to-base-station distance for $\theta = 0$ and $\pi/6$.

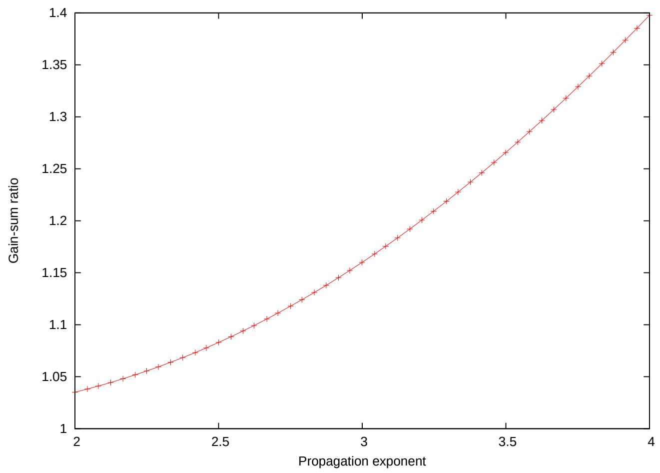

Figure 13.3: Gain-sum ratio $I_\eta\left(R\right)/I_\eta\left(Re^{i\pi/6}\right)$ versus the propagation exponent η.

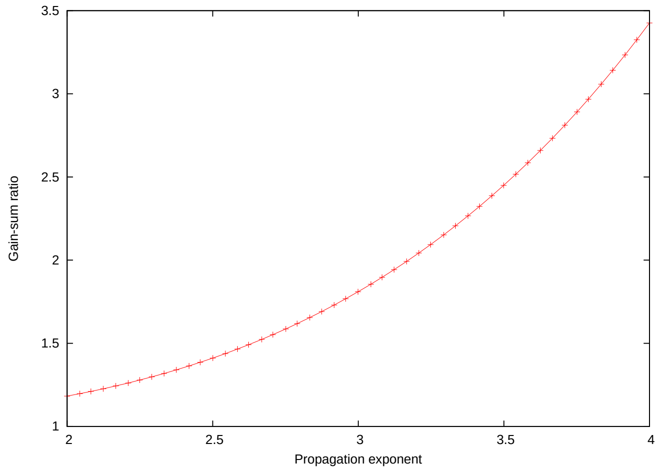

Figure 13.4: Gain-sum ratio $I_\eta(R)/I_\eta(0)$ versus the propagation exponent η.

the gain-sum, we represent in Figure 13.4 the ratio between its maximum and minimum values, i.e. $I_\eta(R)/I_\eta(0)$, as function of the propagation exponent $\eta \in [2,5]$. Observe that the ratio $I_\eta(R)/I_\eta(0)$ is within the interval $[1,4]$ which is too small compared to the propagation-loss ratio $L(R)/L(0)$ which equals ∞ in our model and typically 10^9 in real networks. Hence the dynamics of the gain-sum is too small compared to the dynamics of the propagation-loss, that is $I_\eta(R)/I_\eta(0) \ll L(R)/L(0)$.

In order to gain more insight into the variations of the gain-sum as function of the user-to-base-station distance we make the following approximation.

Approximation 3 *The gain sum may be approximated by*

$$I_{\eta,\Delta}(m) \simeq \zeta(\eta-1)\left[1/L(\Delta-\mathbf{r})+1/L(\Delta+\mathbf{r})+4/L\left(\sqrt{\Delta^2+\mathbf{r}^2}\right)\right], \quad m \in u \tag{13.11}$$

where ζ is the RIEMANN ZETA FUNCTION *given by $\zeta(z)=\sum_{i=1}^{\infty}\frac{1}{i^z}$ and $\mathbf{r}$ is the distance between the user m and its serving base station u.*

Proof. If we fix some hexagon center as the origin O, then the other hexagon centers are located on successive hexagons having O as the center and having increasing radii. These hexagons are called LEVELS and indexed by $j=1,2,\ldots$. Figure 13.5 shows the first two levels $j=1,2$. We decompose the gain-sum $I_{\eta,\Delta}(m)$ over the different levels

$$\begin{aligned} I_{\eta,\Delta}(m) &= \sum_{v\in\mathbf{U}\setminus\{u\}} L_{m,v}^{-1} \\ &= \sum_{j=1}^{\infty}\sum_{v\in\text{Level } j} L_{m,v}^{-1} = \sum_{j=1}^{\infty} A_j \end{aligned}$$

where $A_j = \sum_{v \in \text{Level } j} L_{m,v}^{-1}$ is the contribution of level j. At level j there are exactly $6j$ base stations, six of which (denoted $v_0^{(j)}, v_1^{(j)}, \ldots, v_5^{(j)}$) are at distance $j\Delta$ from the center and the other $6\,(j-1)$ base stations are at distances slightly less than $j\Delta$.Then we get

$$\begin{aligned} A_j &\geq j \sum_{k=0}^{5} L_{m,v_k^{(j)}}^{-1} \\ &= j\left(j^{-\eta} \sum_{k=0}^{5} L_{m,v_k^{(1)}}^{-1}\right) = j^{-(\eta-1)} A_1 \end{aligned}$$

Hence

$$I_{\eta,\Delta}\,(m) \geq \sum_{j=1}^{\infty} j^{-(\eta-1)} A_1 = \zeta\,(\eta-1)\,A_1 \tag{13.12}$$

The first step to get the desired approximation is to assume as in [82] that all the base stations of level j are at the distance $j\Delta$, which gives $I_{\eta,\Delta}\,(m) \simeq \zeta\,(\eta-1)\,A_1$.

We shall now establish an approximation of A_1 which is the contribution of level 1 to the gain sum. The base stations of level 1 are denoted $v_0, v_1, \ldots, v_5$. Figure 13.6 represents a user m at distance $\mathbf{r}$ from its serving base station u. Denote n the point located on the line (u, v_0) at distance $\mathbf{r}$ from u. We make the approximation

$$\begin{cases} L_{v_q,m} \simeq L_{v_q,n} & q \in \{0,3\} \\ L_{v_q,m} \simeq L_{v,n} & q \in \{1,2\} \\ L_{v_q,m} \simeq L_{v',n} & q \in \{4,5\} \end{cases}$$

where v and v' are defined on Figure 13.6. Hence

$$A_1 = \sum_{q=0}^{5} L_{m,v_q}^{-1} \simeq 1/L\,(\Delta - \mathbf{r}) + 1/L\,(\Delta + \mathbf{r}) + 4/L\left(\sqrt{\Delta^2 + \mathbf{r}^2}\right)$$

This finishes the establishment of the desired aprroximation.

Figure 13.7 represents the gain-sum I_η as a function of the user-to-base-station distance calculated with the exact expression and the approximation (13.11) for different values of propagation exponent $\eta = 3, 4, 5$. This figure shows that the approximation is good. ■

Gain-sum at cell center. Using the bound (13.12), we get in particular that the gain sum at the cell center is bounded as follows

$$I_\eta\,(0) \geq 6\zeta\,(\eta - 1)$$

Figure 13.8 represents the gain-sum at cell center $I_\eta\,(0)$ and the lower bound $6\zeta\,(\eta-1)$ as functions of the propagation exponent η. This figure shows that the bound $6\zeta\,(\eta-1)$ is a good approximation of $I_\eta\,(0)$.

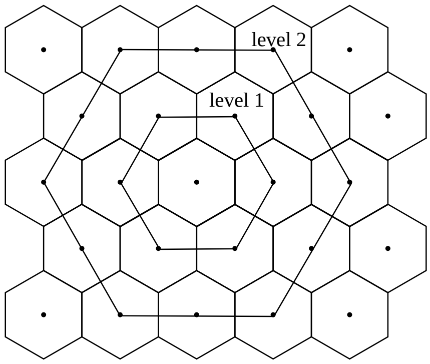

Figure 13.5: The first two levels $j = 1, 2$

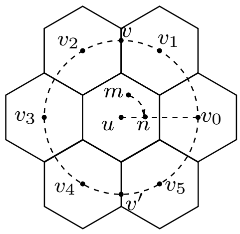

Figure 13.6: Level 1

Proposition 54 *Denote*

$$\varphi(\eta) = I_\eta(0) = \sum_{p,q} \left(p^2 + pq + q^2\right)^{-\eta/2}$$

where the summation is over $p \in \mathbb{Z}$ and $q \in \mathbb{Z}$ with $p \neq 0$ or $q \neq 0$. The function $\varphi :]2, +\infty[\to \mathbb{R}$ is well defined, continuous, convex and decreasing. Moreover

$$\lim_{\eta \to 2} \varphi(\eta) = +\infty \text{ and } \lim_{\eta \to +\infty} \varphi(\eta) = 6$$

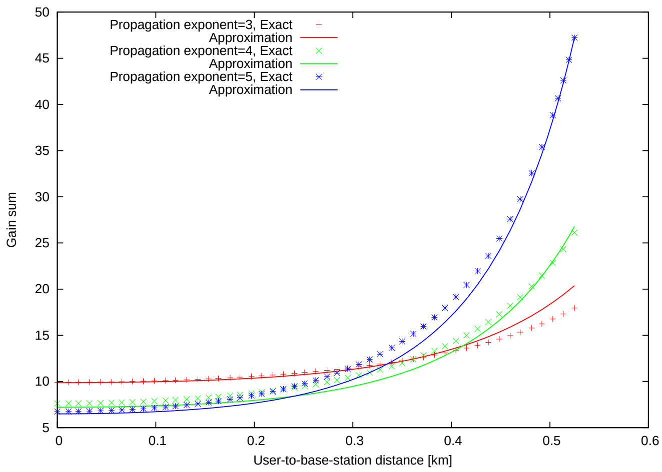

Figure 13.7: Gain-sum versus user-to-base-station distance calculated with the exact expression and the approximation (13.11) for different values of the propagation exponent $\eta = 3, 4, 5$.

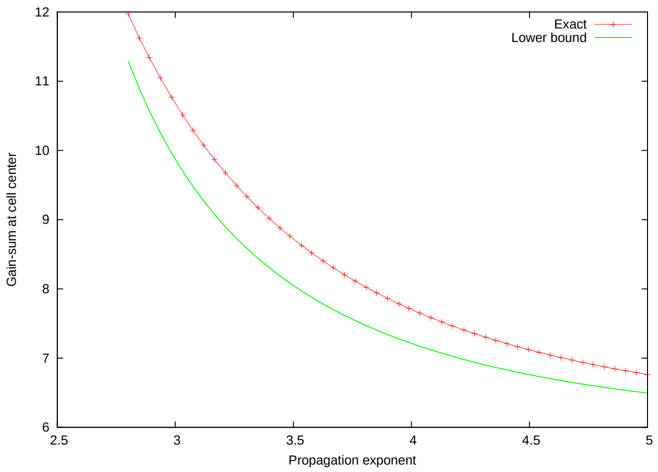

Figure 13.8: Gain-sum at cell center $I_\eta(0)$ and the bound $6\zeta(\eta - 1)$ as functions of the propagation exponent η.

Proof. We have

$$\begin{aligned}
\varphi(\eta) &= \sum_{q\in\mathbb{Z}^*} |q|^{-\eta} + \sum_{p\in\mathbb{Z}^*}\sum_{q\in\mathbb{Z}} \left(p^2+pq+q^2\right)^{-\eta/2} \\
&= \sum_{p\in\mathbb{Z}^*}\left[|p|^{-\eta} + \sum_{q\in\mathbb{Z}} \left(p^2+pq+q^2\right)^{-\eta/2}\right] \\
&= 2\sum_{p=1}^{\infty}\left[2p^{-\eta} + \sum_{s\in\{-1,+1\}}\sum_{q=1}^{\infty} \left(p^2+spq+q^2\right)^{-\eta/2}\right] \\
&= 4\sum_{p=1}^{\infty} p^{-\eta} + 2\sum_{s\in\{-1,+1\}}\sum_{p=1}^{\infty}\sum_{q=1}^{\infty} \left(p^2+spq+q^2\right)^{-\eta/2} \qquad (13.13)
\end{aligned}$$

1. Suppose that $\eta > 2$. Fix some $p > 0$. Using the fact that $p^2+q^2 \geq 2pq$, we deduce that $p^2+spq+q^2 \geq (2+s)\,pq$. Then $\left(p^2+spq+q^2\right)^{-\eta/2} \leq (2+s)^{-\eta/2}\,p^{-\eta/2}q^{-\eta/2}$ which implies $\sum_{q=1}^{\infty}\left(p^2+spq+q^2\right)^{-\eta/2}$ is convergent and

$$\sum_{q=1}^{\infty}\left(p^2+spq+q^2\right)^{-\eta/2} \leq (2+s)^{-\eta/2}\,p^{-\eta/2}\sum_{q=1}^{\infty} q^{-\eta/2}.$$

   Then for $\eta > 2$, the double series in the right-hand side of (13.13) is convergent and we have the upper bound

$$\varphi(\eta) \leq 4\sum_{p=1}^{\infty} p^{-\eta} + 2\left(3^{-\eta/2}+1\right)\left(\sum_{p=1}^{\infty} p^{-\eta/2}\right)^2$$

   Hence the function $\varphi :]2,+\infty[\to \mathbb{R}$ is well defined.

2. The series $\sum_{p,q}\left(p^2+pq+q^2\right)^{-\eta/2}$ of functions of η converges uniformly over each compact in $]2,+\infty[$, then $\varphi(\eta)$ is continuous over $]2,+\infty[$. It is clearly a decreasing function of η. As a sum of convex function, $\varphi(\eta)$ is convex.

3. Using (13.12), we deduce that

$$\lim_{\eta\to 2}\varphi(\eta) = +\infty.$$

   Form the monotone convergence theorem for series, we get

$$\begin{aligned}
\lim_{\eta\to+\infty}\varphi(\eta) &= \lim_{\eta\to+\infty}\sum_{p,q}\left(p^2+pq+q^2\right)^{-\eta/2} \\
&= \sum_{p,q}\lim_{\eta\to+\infty}\left(p^2+pq+q^2\right)^{-\eta/2} = 6
\end{aligned}$$

■

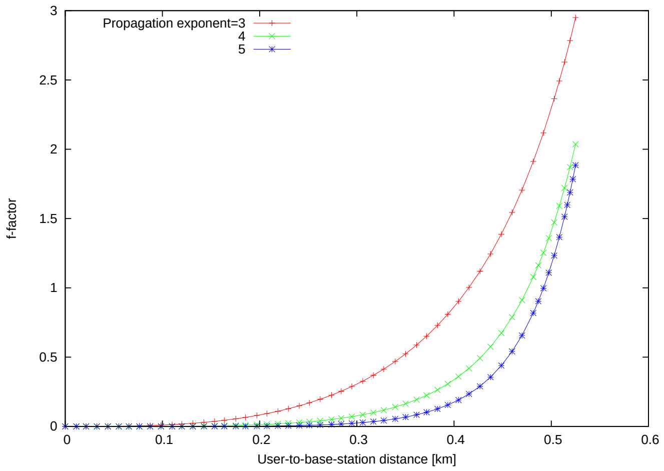

Figure 13.9: f-factor versus user-to-base-station distance for different values of the propagation exponent $\eta = 3, 4, 5$.

13.B.2 f-factor and lf-factor

The f-factor depends on the user position m, the propagation exponent η and the distance Δ between two adjacent base stations. So we denote it by $f_{\eta,\Delta}(m)$. It is related to the gain sum $I_{\eta,\Delta}(m)$ by the relation

$$f_{\eta,\Delta}(m) = I_{\eta,\Delta}(m) L_{u,m}, \quad m \in u$$

The gain-sum approximation given in Equation (13.11), gives an approximation of the f-factor which we denote, with a slight *abuse of notation*, by $f_{\eta,\Delta}(\mathbf{r})$, thus

$$f_{\eta,\Delta}(m) \simeq f_{\eta,\Delta}(\mathbf{r}) = \zeta(\eta-1) L(\mathbf{r}) \left[1/L(\Delta-\mathbf{r}) + 1/L(\Delta+\mathbf{r}) + 4/L\left(\sqrt{\Delta^2+\mathbf{r}^2}\right)\right]$$

and in particular, for $\Delta = 1$,

$$f_\eta(\mathbf{r}) = \zeta(\eta-1)\left[\mathbf{r}^\eta(1-\mathbf{r})^{-\eta} + \mathbf{r}^\eta(1+\mathbf{r})^{-\eta} + 4\mathbf{r}^\eta\left(1+\mathbf{r}^2\right)^{-\eta/2}\right]$$

From now we call, with a slight *abuse of terminology*, the approximation $f_{\eta,\Delta}(\mathbf{r})$ also f-fa*ctor*.

Figure 13.9 represents the f-factor f_η as a function of the user-to-base-station distance $\mathbf{r}$ for different values of propagation exponent $\eta = 3, 4, 5$.

We define the LF-FACTOR by

$$h_{\eta,\Delta}(\mathbf{r}) = f_{\eta,\Delta}(\mathbf{r}) L(\mathbf{r}) / L(R)$$

We shall now calculate the mean of the f-factor and the lf-factor over a given cell.

Proposition 55 *The f-factor mean over a cell equals*

$$\mathbf{E}\left[f_{\eta,\Delta}\right] = \frac{2}{R^2}\int_0^R \mathbf{r} f_{\eta,\Delta}\left(\mathbf{r}\right) d\mathbf{r} = \zeta\left(\eta-1\right)\frac{\left(R/\Delta\right)^{\eta}}{1+\eta/2}\{\ {}_2F_1\left([\eta,2+\eta],[3+\eta]\,,R/\Delta\right)$$
$$+\ {}_2F_1\left([\eta,2+\eta],[3+\eta]\,;-R/\Delta\right)$$
$$+4\times\ {}_2F_1\left(\left[\eta/2,1+\eta/2\right],\left[2+\eta/2\right],-\left(R/\Delta\right)^2\right)\} \tag{13.14}$$

(where ${}_2F_1$ designates the hypergeometric function) and the lf-factor mean over a cell equals

$$\mathbf{E}\left[h_{\eta,\Delta}\right] = \frac{2}{R^2}\int_0^R \mathbf{r} h_{\eta,\Delta}\left(\mathbf{r}\right) d\mathbf{r} = \zeta\left(\eta-1\right)\frac{\left(R/\Delta\right)^{\eta}}{1+\eta}\{\ {}_2F_1\left([\eta,2+2\eta],[3+2\eta]\,,R/\Delta\right)$$
$$+\ {}_2F_1\left([\eta,2+2\eta],[3+2\eta]\,;-R/\Delta\right)$$
$$+4\times\ {}_2F_1\left(\left[\eta/2,1+\eta\right],\left[2+\eta\right],-\left(R/\Delta\right)^2\right)\} \tag{13.15}$$

Proof. We calculate

$$\mathbf{E}\left[f_{\eta,\Delta}\right] = \frac{2}{R^2}\int_0^R \mathbf{r} f_{\eta,\Delta}\left(\mathbf{r}\right) d\mathbf{r}$$
$$= \frac{2}{R^2}\zeta\left(\eta-1\right)\{\int_0^R \mathbf{r}^{1+\eta}\left(\Delta-\mathbf{r}\right)^{-\eta} d\mathbf{r} + \int_0^R \mathbf{r}^{1+\eta}\left(\Delta+\mathbf{r}\right)^{-\eta} d\mathbf{r}$$
$$+4\int_0^R \mathbf{r}^{1+\eta}\left(\Delta^2+\mathbf{r}^2\right)^{-\eta/2} d\mathbf{r}\}$$

We calculate

$$\int_0^R \mathbf{r}^{1+\eta}\left(\Delta-\mathbf{r}\right)^{-\eta} d\mathbf{r} = R^2\frac{\left(R/\Delta\right)^{\eta}}{2+\eta}\ {}_2F_1\left([\eta,2+\eta],[3+\eta]\,,R/\Delta\right)$$

$$\int_0^R \mathbf{r}^{1+\eta}\left(\Delta+\mathbf{r}\right)^{-\eta} d\mathbf{r} = R^2\frac{\left(R/\Delta\right)^{\eta}}{2+\eta}\ {}_2F_1\left([\eta,2+\eta],[3+\eta]\,,-R/\Delta\right)$$

$$\int_0^R \mathbf{r}^{1+\eta}\left(\Delta^2+\mathbf{r}^2\right)^{-\eta/2} d\mathbf{r} = R^2\frac{\left(R/\Delta\right)^{\eta}}{2+\eta}\ {}_2F_1\left(\left[\eta/2,1+\eta/2\right],\left[2+\eta/2\right],-\left(R/\Delta\right)^2\right)$$

This finishes the proof of Equation (13.14).

We have

$$\mathbf{E}\left[h_{\eta,\Delta}\right] = \frac{2}{R^2}\int_0^R \mathbf{r} h_{\eta,\Delta}\left(\mathbf{r}\right) d\mathbf{r}$$
$$= 2R^{-2-\eta}\zeta\left(\eta-1\right)\{\int_0^R \mathbf{r}^{1+2\eta}\left(\Delta-\mathbf{r}\right)^{-\eta} d\mathbf{r} + \int_0^R \mathbf{r}^{1+2\eta}\left(\Delta+\mathbf{r}\right)^{-\eta} d\mathbf{r}$$
$$+4\int_0^R \mathbf{r}^{1+2\eta}\left(\Delta+\mathbf{r}^2\right)^{-\eta/2} d\mathbf{r}\}$$

We calculate

$$\int_0^R \mathbf{r}^{1+2\eta} (\Delta - \mathbf{r})^{-\eta} d\mathbf{r} = \frac{R^{2+\eta}}{2} \frac{(R/\Delta)^\eta}{1+\eta} {}_2F_1 ([\eta, 2+2\eta], [3+2\eta], (R/\Delta))$$

$$\int_0^R \mathbf{r}^{1+\eta} (\Delta + \mathbf{r})^{-\eta} d\mathbf{r} = \frac{R^{2+\eta}}{2} \frac{(R/\Delta)^\eta}{1+\eta} {}_2F_1 ([\eta, 2+2\eta], [3+2\eta], -(R/\Delta))$$

$$\int_0^R \mathbf{r}^{1+\eta} \left(\Delta + \mathbf{r}^2\right)^{-\eta/2} d\mathbf{r} = \frac{R^{2+\eta}}{2} \frac{(R/\Delta)^\eta}{1+\eta} {}_2F_1 \left([\eta/2, 1+\eta], [2+\eta], -(R/\Delta)^2\right)$$

This finishes the proof of Equation (13.15). ■

From (13.10) we deduce that $f_{\eta,\Delta}(\mathbf{r}) = f_\eta(\mathbf{r}/\Delta)$, then

$$\begin{aligned}\mathbf{E}\left[f_{\eta,\Delta}^k\right] &= \frac{2}{R^2}\int_0^R \mathbf{r} f_{\eta,\Delta}^k(\mathbf{r})\, d\mathbf{r} = \frac{2}{R^2}\int_0^R \mathbf{r} f_\eta^k(\mathbf{r}/\Delta)\, d\mathbf{r} \\ &= \frac{2}{(R/\Delta)^2}\int_0^{R/\Delta} u f_\eta^k(u)\, du = \mathbf{E}\left[f_\eta^k\right], \quad k \in \mathbb{N}\end{aligned}$$

Since the ratio R/Δ is constant (see Equation (2.6)), then the moments of $f_{\eta,\Delta}$ and f_η over the correspondent cells are equal. We have also similar result for the lf-fcator. Hence we may limit ourselves to the case $\Delta = 1$ and cell radius equal to

$$R_1 = \sqrt{\sqrt{3}/(2\pi)} \simeq 0.525$$

For the Poisson-Voronoi model, we have from [16] $\mathbf{E}[f_\eta] = 2/(\eta-2)$. This inspires us to make the following approximations.

Approximation 4 *For the hexagonal model, we have the approximations*

$$\mathbf{E}[f_\eta] \simeq 0.936/(\eta-2) \tag{13.16}$$

$$\mathbf{E}[h_\eta] \simeq 0.632/(\eta-2) \tag{13.17}$$

$$\mathbf{E}\left[f_\eta^2\right] \simeq 0.234/(\eta-2) + 1.29/(\eta-2)^2 \tag{13.18}$$

Proof. A least square fit between $\mathbf{E}[f_\eta]$ and $\mathbf{E}[h_\eta]$ versus $1/(\eta-2)$ gives the right hand side of Equations (13.16) and (13.17) respectively. Recall that, by definition,

$$\mathbf{E}\left[f_\eta^2\right] = \frac{2}{R_1^2}\int_0^{R_1} \mathbf{r} f_\eta(\mathbf{r})^2\, d\mathbf{r}$$

A least square fit of $\mathbf{E}\left[f_\eta^2\right]$ versus $1/(\eta-2)$ and $1/(\eta-2)^2$ gives the right hand side of Equation (13.18).

Figure 13.10 represent $\mathbf{E}[f_\eta]$, $\mathbf{E}[h_\eta]$ and $\mathbf{E}\left[f_\eta^2\right]$ obtained by exact calculation and by the approximations (13.16), (13.17) and (13.18) respectively. Visual inspection of this figure show that the approximation is good when the propagation exponent $\eta \in [2.2, 5]$.

■

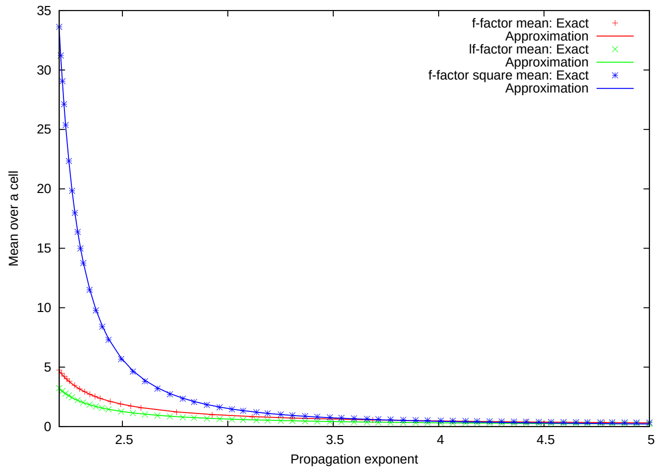

Figure 13.10: Validation of Approximation 4.

13.C Directional antennas

Till now we have considered omni antennas. In this appendix we shall study the f-factor properties and extend the mean model and the infieasibility probability formulae to the case with directional antennas.

In the directional case, the geographic positions of the base stations are called SITES. Each base station has an antenna and its CELL is defined as the set of geographic position where its is the best server.

We use the following notation:

- We assume that there are D antennas (or equivalenly base stations) located at each site.
- These antennas have *azimuths* $a2\pi/D$ where $a \in \mathbf{D} = \{-(D-1)/2, \ldots, (D-1)/2\}$. We use $\mathbf{D}$ as an index set of the antennas located at the same site.
- The set of sites is denoted by $\mathbf{U}$.
- A base station is a pair (u, a) where $u \in \mathbf{U}$ and $a \in \mathbf{D}$.
- The set of base stations is the Cartesian product $\mathbf{U} \times \mathbf{D}$.
- m designates a user position.
- The notation $m \in (u, a)$ means that the user m is in the cell of the base station (u, a).

13.C.1 The f-factor

In this section we shall extend the results about the f-factor presented in §13.B to the case with directional antennas.

The present section is based on [20] for the model and on [117] for the f-factor moments in the hexagonal case.

Model. The propagation-loss has the form

$$L_{(u,a),m} = L_{u,m}/G_{(u,a),m}$$

where:

- $L_{u,m} = L(|m-u|)$ is the distance propagation loss.
- $G_{(u,a),m}$ is the antenna gain given by

  $$G_{(u,a),m} = G(\arg(m-u) - a2\pi/D)$$

  where $G(\cdot)$ is the normalized *radiation pattern*.
- $|z|, \arg(z)$ are, respectively, the Euclidian norm and the argument of the vector $z \in \mathbb{R}^2$ $(-\pi < \arg(z) \leq \pi)$.

We consider two models for the normalized radiation pattern: a perfect one where $G(\theta) = 1\{|\theta| \leq \pi/D\}$ and a realistic one. Each antenna (u,a) covers the sector situated at angle less than π/D around its azimuth. This means that the antenna (u,a) servers the users m satisfying $|\arg(m-u) - a2\pi/D| < \pi/D$.

Lemma 21 *The f-factor in the directional case equals*

$$f^{(D)}(m) = \sum_{b\in\mathbf{D}\setminus\{a\}} \frac{G_{(u,b),m}}{G_{(u,a),m}} + \sum_{v\in\mathbf{U}\setminus\{u\}} \frac{L_{u,m}}{L_{v,m}} \frac{G_{v,m}}{G_{(u,a),m}}, \quad m \in (u,a)$$

where $G_{v,m}$ designates the TOTAL ANTENNA GAIN *from a given site v which is defined by*

$$G_{v,m} = \sum_{b\in\mathbf{D}} G_{(v,b),m}$$

Proof. For a user $m \in (u,a)$, the f-factor equals by definition

$$\begin{aligned} f^{(D)}(m) &= \sum_{(v,b)\in\mathbf{U}\times\mathbf{D}\setminus\{(u,a)\}} \frac{L_{(u,a),m}}{L_{(v,b),m}} \\ &= \sum_{(v,b)\in\mathbf{U}\times\mathbf{D}\setminus\{(u,a)\}} \frac{L_{u,m}}{L_{v,m}} \frac{G_{(v,b),m}}{G_{(u,a),m}} \end{aligned}$$

By distinguishing the case $v = u$ and $v \neq u$ in the above sumn, we get the desired result. ■

Perfect radiation pattern. In the case of perfect radiation pattern, the f-factor equals

$$f^{(D)}(m) = \sum_{v \in \mathbf{U} \setminus \{u\}} \frac{L_{u,m}}{L_{v,m}}, \quad m \in (u,a)$$

which is independent of the number of sectors D, and hence equals the f-factor in the omni case $f_m^1 = f_m^1$ which we have already studied.

Realistic radiation pattern.

Approximation 5 *We approximate the total antenna gain by its mean*

$$G_{v,m} \simeq \bar{G} = \frac{D}{2\pi} \int_{-\pi}^{\pi} G(\theta)\, d\theta \tag{13.19}$$

Proof. Consider the antenna (u,a) serving the user m. Without loss of generality we may take u as the origin of the coordinate system and $m-u$ as the reference for angles. We have

$$\begin{aligned} G_{v,m} &= \sum_{b \in \mathbf{D}} G_{(v,b),m} \\ &= \sum_{b \in \mathbf{D}} G(\arg(m-v) - b2\pi/D) \\ &= \sum_{b \in \mathbf{D}} G(\arg(-v) - b2\pi/D) \end{aligned}$$

Let

$$\mathbb{G}(\theta) = \sum_{b \in \mathbf{D}} G(\theta - b2\pi/D) \tag{13.20}$$

where $\theta = \arg(-v)$. Then

$$\begin{aligned} \bar{G} &= \mathbf{E}[\mathbb{G}(\theta)] \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbb{G}(\theta)\, d\theta \\ &= \sum_{b \in \mathbf{D}} \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\theta - b2\pi/D)\, d\theta = \frac{D}{2\pi} \int_{-\pi}^{\pi} G(\theta)\, d\theta \end{aligned}$$

■

Using the approximation in Equation (13.19) we get

$$f^{(D)}(m) = \frac{\sum_{b \in \mathbf{D} \setminus \{a\}} G(\theta - b2\pi/D)}{G(\theta)} + \frac{\bar{G}}{G(\theta)} f(m), \quad m \in (u,a)$$

where $\theta = \arg(m-u)$ and

$$f(m) = \sum_{v \in \mathbf{U} \setminus \{u\}} \frac{L_{u,m}}{L_{v,m}}, \quad m \in (u,a)$$

is the f-factor in the omni case. Using the approximation in Equation (13.11), $f(m)$ depends only on the distance $|m-u|$. Using the fact that $\theta = \arg(m-u)$ is uniformly distributed in $(-\pi/D, \pi/D)$, we may calculate the f-factor moments

$$\bar{f}^{(D)} = \bar{G}G_2\bar{f} + G_1 \quad (13.21)$$

$$\bar{f^2}^{(D)} = \bar{G}^2G_4\bar{f^2} + 2\bar{G}G_5\bar{f} + G_3 \quad (13.22)$$

where the ANTENNA PARAMETERS $\bar{G}, G_1, \ldots, G_5$ are given by

$$\begin{aligned}
\bar{G} &= \tfrac{D}{2\pi}\textstyle\int_{-\pi}^{\pi} G(\theta)\, d\theta \\
G_1 &= \tfrac{D}{2\pi}\textstyle\int_{-\pi/D}^{\pi/D} \frac{\sum_{b\neq 0} G(\theta - b2\pi/D)}{G(\theta)} d\theta \\
G_2 &= \tfrac{D}{2\pi}\textstyle\int_{-\pi/D}^{\pi/D} \frac{1}{G(\theta)} d\theta \\
G_3 &= \tfrac{D}{2\pi}\textstyle\int_{-\pi/D}^{\pi/D} \left[\frac{\sum_{b\neq 0} G(\theta - b2\pi/D)}{G(\theta)}\right]^2 d\theta \\
G_4 &= \tfrac{D}{2\pi}\textstyle\int_{-\pi/D}^{\pi/D} \frac{1}{G^2(\theta)} d\theta \\
G_5 &= \tfrac{D}{2\pi}\textstyle\int_{-\pi/D}^{\pi/D} \frac{\sum_{b\neq 0} G(\theta - b2\pi/D)}{G^2(\theta)} d\theta
\end{aligned} \quad (13.23)$$

Denote the normalized propagation-loss in the directional case by

$$l^{(D)}(\mathbf{r}, \theta) = \frac{l(\mathbf{r})}{G(\theta)} = \frac{L(\mathbf{r})}{L(R)\, G(\theta)}$$

Its moments are given by

$$\bar{l}^{(D)} = \mathbf{E}\left[G^{-1}l\right] = G_2\bar{l}$$

$$\bar{l^2}^{(D)} = \mathbf{E}\left[G^{-2}l^2\right] = G_4\bar{l^2}$$

$$\begin{aligned}
\bar{fl}^{(D)} &= \mathbf{E}\left[f^{(D)}G^{-1}l\right] \\
&= \mathbf{E}\left[\frac{\sum_{b\neq 0} G(\theta - b2\pi/D)}{G(\theta)^2} l + \frac{\bar{G}}{G(\theta)^2} fl\right] = \bar{G}G_4\bar{fl} + G_5\bar{l}
\end{aligned} \quad (13.24)$$

Study of measured patterns. We consider antennas used for UMTS networks. We study several antennas and present in detail the results for three representative antennas with horizontal beamwidths $\phi_0 = 69°$, $90°$ and $112°$. (The antennas are named ALG_7520-00_T2_2200, ALG_7740-00_T0_2200 and K_742-149_T0_2200 respectively.) The radiation measured pattern $G(\cdot)$ is given as pairs $(\theta, G(\theta))$ for $\theta = 1, 2, \ldots 359°$. The BEAMWIDTH ϕ_0 satisfies by definition $G(\phi_0/2) = 1/2 = 3dB$.

The antenna parameter $\bar{G}$ given by (13.23) are calculated numerically

$$\bar{G} = \begin{cases} 0.619 & \text{for } \phi_0 = 69° \\ 0.805 & \text{for } \phi_0 = 90° \\ 0.941 & \text{for } \phi_0 = 112° \end{cases}$$

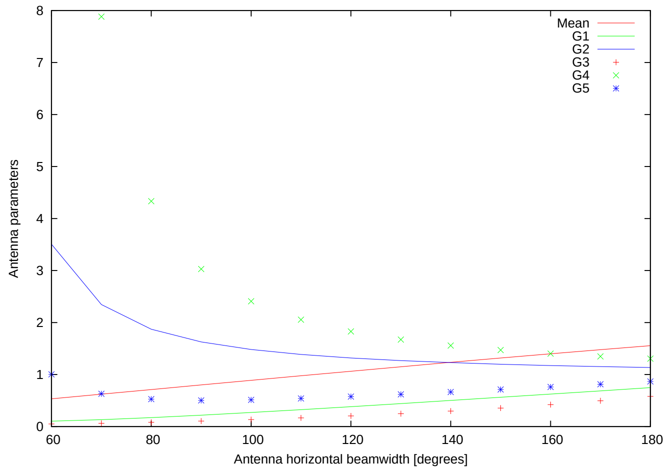

Figure 13.11: Antenna parameters $\bar{G}, G_1, G_2, G_3, G_4$ and G_5 as functions of the horizontal beamwidth ϕ.

The total antenna gain $\mathbb{G}(\theta)$ given by (13.20) varies in the intervals

$$\mathbb{G}(\theta) \in \begin{cases} [0.26, 1.00] & \text{for } \phi_0 = 69^\circ \\ [0.49, 1.01] & \text{for } \phi_0 = 90^\circ \\ [0.80, 1.02] & \text{for } \phi_0 = 112^\circ \end{cases}$$

So the approximation in Equation (13.19) is better when the beamwidth is larger.

The antenna parameters $G_1, \ldots, G_5$ given by (13.23) are calculated numerically

$$(G_1, \ldots, G_5) = \begin{cases} (0.127, 2.43, 0.0572, 8.60, 0.629) & \text{for } \phi_0 = 69^\circ \\ (0.131, 1.58, 0.0483, 2.94, 0.324) & \text{for } \phi_0 = 90^\circ \\ (0.164, 1.24, 0.0798, 1.67, 0.288) & \text{for } \phi_0 = 112^\circ \end{cases}$$

Antenna parameters versus beamwidth. We shall now study the effect of the horizontal beamwidth on the antenna parameters (13.23). We consider the antenna with horizontal beamwidth $\phi_0 = 69^\circ$ as a reference antenna and we construct radiation patterns $G(\cdot)$ with arbitrary horizontal beamwidth ϕ by scaling the original one $G_0(\cdot)$ as follows

$$G(\theta) = G_0(\theta\phi_0/\phi), \quad |\theta| \leq \pi$$

which respects $G(\phi/2) = G_0(\phi_0/2) = 1/2$. Figure 13.11 shows the evolution of the antenna parameters $\bar{G}, G_1, G_2, G_3, G_4$ and G_5 as functions of the horizontal beamwidth ϕ.

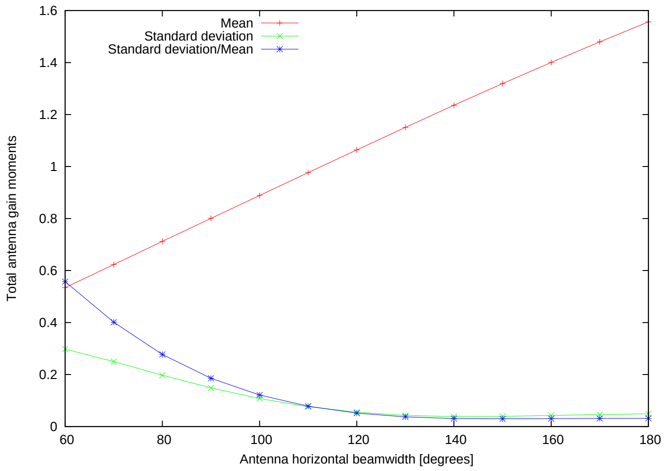

Figure 13.12: The mean $\bar{G} = \mathbf{E}[\mathbf{G}(\theta)]$, the standard deviation $\sigma = (\mathrm{Var}[\mathbf{G}(\theta)])^{1/2}$ of the total antenna gain $\mathbf{G}(\theta)$ and the ratio $\sigma/\bar{G}$ as functions of the beamwidth ϕ.

We study now the precision of the approximation in Equation (13.19) for different values of the beamwidth. Figure 13.12 shows the mean $\bar{G} = \mathbf{E}[\mathbb{G}(\theta)]$, the standard deviation $\sigma = (\mathrm{Var}[\mathbb{G}(\theta)])^{1/2}$ of the total antenna gain $\mathbb{G}(\theta)$ and the ratio $\sigma/\bar{G}$ as functions of the beamwidth ϕ. (The mean and the standard deviation are calculated assuming that θ is uniformly distributed in $]-\pi,\pi]$.) This figure shows that the larger is the beamwidth, the better is the approximation in Equation (13.19). For $\phi \geq 90°$, the approximation is acceptable, and for $\phi \geq 120°$, the approximation is good.

13.C.2 Mean model

In this section we shall extend the mean model presented in §4.1 to the case with directional antennas.

Proposition 56 *The feasibility condition for the mean model in the directional case may be written in the form*

$$\bar{M} \leq \Gamma^{(D)} = \frac{CD}{\bar{\varphi}^{(D)}}$$

where $\bar{M} = \lambda_{\mathrm{M}} \pi R^2$*;*

$$\bar{\varphi}^{(D)} = \begin{cases} \left(\alpha + \bar{f}^{(D)}\right)\bar{\xi}' & \text{for DFC, UFC, EUFC} \\ \left[\alpha + \bar{f}^{(D)} + \bar{l}^{(D)} L(R)\, N/\tilde{P}\right]\bar{\xi}' & \text{for EDFC} \end{cases} \qquad (13.25)$$

and C *is given by Equation (4.12).*

Proof. To get the results of the directional case, it is enough to make the following substitution in the formulae of the omni case presented in §4.1

$$\bar{f} \to \bar{f}^{(D)}$$
$$\bar{l} \to \bar{l}^{(D)}$$
$$\bar{M} \to \bar{M}^{(D)} = \bar{M}/D$$

■

The pole capacity $\Gamma^{(D)}$ of the directional case will be larger than that for the omni case by a factor

$$\frac{\Gamma^{(D)}}{\Gamma} = D \times \frac{\bar{\varphi}}{\bar{\varphi}^{(D)}}$$

The above formula permits to analyze the effect on pole capacity when we replace omni antennas by directional ones. If the antennas have perfect radiation pattern, then $\Gamma^{(D)} = D \times \Gamma$, hence we get an increase of the pole capacity by a factor D due to the increase of the number of base stations on each site. Otherwise, the pole-capacity increase factor is smaller than D by a factor equal to $\bar{\varphi}/\bar{\varphi}^{(D)}$ which characterizes the effect of the radiation-pattern-imperfection.

Figure 13.13 represents the ratio between the directional pole capacity and the omni pole capacity given in the above display as function of the antenna beamwidth for DFC, UFC, EUFC and EDFC for cell radius $R = 2, 3, 4$km.

This figure shows that the pole capacity increases due to directional antennas:

- is optimal for some beamwidth (approximately equal to 90°);
- is independent of the cell radius for DFC, UFC and EUFC;
- increases with cell radius for EDFC.

13.C.3 Infeasibility probability

In this section we shall extend the mean model presented in §4.2 to the case with directional antennas.

Proposition 57 *The feasibility conditions DFC, EDFC, UFC and EUFC may be written in the generic form*

$$S^{(D)} = \sum_{m \in (u,a)} \varphi^{(D)}(m) < C \tag{13.26}$$

where

$$\varphi^{(D)}(m) = \begin{cases} \left[\alpha + f^{(D)}(m)\right] \xi'_m & \text{for DFC, UFC, EUFC} \\ \left[\alpha + f^{(D)}(m) + NL_{(u,a),m}/\tilde{P}\right] \xi'_m & \text{for EDFC} \end{cases}, \quad m \in (u,a)$$

and C is given by Equation (4.12).

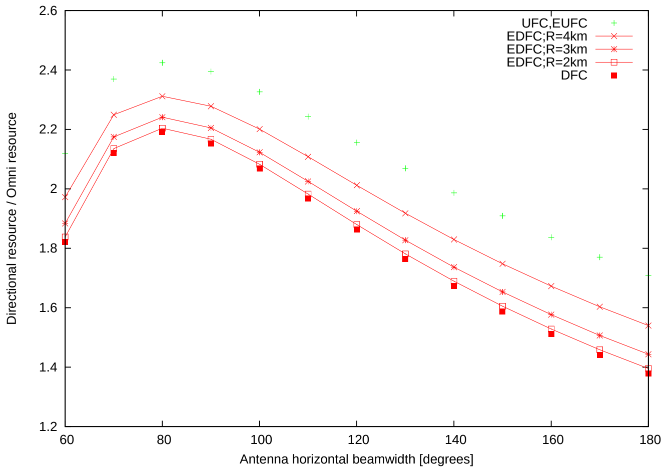

Figure 13.13: Pole capacity increases due to directional antenna.

The INFEASIBILITY PROBABILITY is defined as follows

$$\mathbf{P}\left(S^{(D)} \geq C\right)$$

where $S^{(D)}$ is given by (13.26). Due to Hypothesis 1, the infeasibility probability is the same for all base station $u \in \mathbf{U}$.

Since $S^{(D)} = \sum_{m \in (u,a)} \varphi^{(D)}(m)$ is a shot noise, its mean and standard deviation are respectively given by

$$\bar{S}^{(D)} = D^{-1}\bar{M}\bar{\varphi}^{(D)}, \quad \sigma_S = \left(D^{-1}\bar{M}\bar{\varphi^2}^{(D)}\right)^{1/2}$$

Proposition 58 *The mean of $\varphi^{(D)}(m)$ is given by Equation (13.25). Its second moment is given by*

$$\bar{\varphi^2}^{(D)} = \begin{cases} \left(\bar{f^2}^{(D)} + 2\alpha\bar{f}^{(D)} + \alpha^2\right)\bar{\xi'^2} & \textit{for DFC, UFC, EU} \\ \left[L(R)^2\,\bar{l^2}^{(D)}N^2/\tilde{P}^2 + \bar{f^2}^{(D)} + \alpha^2 + 2\alpha\bar{f}^{(D)} + 2\left\{\alpha\bar{l}^{(D)} + \bar{f}\bar{l}^{(D)}\right\}L(R)\,N/\tilde{P}\right]\bar{\xi'^2} & \textit{for EDFC} \end{cases}$$

13.D Synthesis of formulae

We recall below the essential formulae developed in the present part. Recall that the feasibility conditions DFC, EDFC, UFC, EUFC have the form

$$S = \sum_{m \in u} \varphi(m) \leq C$$

in the case of omni antennas, and

$$S^{(D)} = \sum_{m\in(u,a)} \varphi^{(D)}(m) \le C$$

in the case of directional antennas.

The functions $\varphi(\cdot)$ and $\varphi^{(D)}(\cdot)$ comprise a term, called f-factor and denoted $f(m)$, for which we propose the following approximation

$$f(m) \approx f(|m|) \approx \zeta(\eta-1)L(|m|)\left(\frac{1}{L(\Delta-|m|)}+\frac{1}{L(\Delta+|m|)}+\frac{4}{L\left(\sqrt{\Delta^2+|m|^2}\right)}\right) \quad |m|\le R$$

where $\zeta(x)=\sum_{n=1}^{\infty} 1/n^x$ is the Riemann zeta function. (Recall that Δ is the distance between two adjacent base stations in the hexagonal model and R is the radius of the disc with area equal to that of a hexagon.)

The parameter C is given by

$$C = \begin{cases} 1 & \text{for DFC et UFC} \\ 1-\epsilon & \text{for EDFC} \\ 1-N/\inf_{m\in u}\frac{\tilde{P}_m}{\xi'_m L_{u,m}} & \text{for EUFC} \end{cases}$$

The moments of S can be expressed via some antenna parameters $\bar{G}, G_i$, $i=1,\ldots,5$ and functions $\bar{f}, \bar{f^2}, \bar{l}, \bar{lf}$ of the path loss exponent η, which are given at the end of this section. Concluding, the mean and standard deviation of S are respectively given by

$$\bar{S} = \bar{M}\bar{\varphi}, \quad \sigma_S^2 = \bar{M}\bar{\varphi^2}$$

Those of $S^{(D)}$ have a similar form with the substitution

$$\begin{aligned} \bar{\varphi} &\to \bar{\varphi}^{(D)} \\ \bar{M} &\to \bar{M}^{(D)} = \bar{M}/D \end{aligned}$$

The moments of $\varphi(\cdot)$ are given by

$$\bar{\varphi} = \begin{cases} \left(\alpha+\bar{f}\right)\bar{\xi'} & \text{for DFC, UFC, EUFC} \\ \left[\alpha+\bar{f}+\bar{l}L(R)N/\tilde{P}\right]\bar{\xi'} & \text{for EDFC} \end{cases}$$

$$\bar{\varphi^2} = \begin{cases} \left(\bar{f^2}+2\alpha\bar{f}+\alpha^2\right)\bar{\xi'^2} & \text{for DFC, UFC, EUFC} \\ \left[L(R)^2\bar{l^2}N^2/\tilde{P}^2+\bar{f^2}+\alpha^2+2\alpha\bar{f}+2\left\{\alpha\bar{l}+\bar{fl}\right\}L(R)N/\tilde{P}\right]\bar{\xi'^2} & \text{for EDFC} \end{cases}$$

The moments of $\varphi^{(D)}(\cdot)$ have similar form with the substitution

$$\begin{aligned} \bar{f} &\to \bar{f}^{(D)} = \bar{G}G_2\bar{f}+G_1 \\ \bar{f^2} &\to \bar{f^2}^{(D)} = \bar{G}^2G_4\bar{f^2}+2\bar{G}G_5\bar{f}+G_3 \\ \bar{l} &\to \bar{l}^{(D)} = G_2\bar{l} \\ \bar{l^2} &\to \bar{l^2}^{(D)} = G_4\bar{l^2} \\ \bar{fl} &\to \bar{fl}^{(D)} = \bar{G}G_4\bar{fl}+G_5\bar{l} \end{aligned}$$

The antenna parameters are, for the three-sector case,

$$
\begin{aligned}
\bar{G} &= \frac{3}{2\pi}\int_{\pi}^{\pi} G(\theta)\,d\theta \\
G_1 &= \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} \frac{\sum_{b\neq 0} G\left(\psi - b2\pi/3\right)}{G\left(\psi\right)} d\psi \\
G_2 &= \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} \frac{1}{G\left(\psi\right)} d\psi \\
G_3 &= \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} \left[\frac{\sum_{b\neq 0} G\left(\psi - b2\pi/3\right)}{G\left(\psi\right)}\right]^2 d\psi \\
G_4 &= \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} \frac{1}{G^2\left(\psi\right)} d\psi \\
G_5 &= \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} \frac{\sum_{b\neq 0} G\left(\psi - b2\pi/3\right)}{G^2\left(\psi\right)} d\psi
\end{aligned}
$$

The integral functions of the path-loss exponent η are

$$
\begin{aligned}
\bar{f} = \bar{f}_\eta &= 2\int_0^1 \mathbf{r} f|_{K=1,R=1}(\mathbf{r}) d\mathbf{r} \approx \frac{0.9365}{\eta - 2}, \\
\bar{l} = \bar{l}_\eta &= 2\int_0^1 \mathbf{r}^{\eta+1} d\mathbf{r} = \frac{1}{1+\eta/2}, \\
\bar{f^2} = \bar{f^2}_\eta &= 2\int_0^1 \mathbf{r} f^2|_{K=1,R=1}(\mathbf{r}) d\mathbf{r} \\
&\approx \frac{0.2343}{\eta - 2} + \frac{1.2907}{(\eta-2)^2}, \\
\bar{lf} = \bar{lf}_\eta &= 2\int_0^1 \mathbf{r}^{1+\eta} f|_{K=1,R=1}(\mathbf{r}) d\mathbf{r} \approx \frac{0.6362}{\eta - 2}.
\end{aligned}
$$

Multi-service case Consider the case where there are J classes of service. Each class $j \in \{1, 2, \ldots, J\}$ is characterized by modified SINRs ξ'_j and the mean number of user of this class in a cell $\bar{M}_j$. We have $\bar{M} = \sum_{j=1}^{J} \bar{M}_j$ and

$$
\bar{\xi'} = \sum_{j=1}^{J} p_j \xi'_j, \quad \bar{\xi'^2} = \sum_{j=1}^{J} p_j \xi'^2_j
$$

where $p_j = \bar{M}_j / \bar{M}$.

13.E Comparison of admission control schemes

The objective of the present section is to compare the FC admission control schemes built in Section 3.3 to the optimal scheme (corresponding to the NSFC admission control) as well as to the constructor's schemes. We compare the performance of these admission control schemes in terms of the *blocking* and *cut* probabilities. To this end, the *semi-static* model considered in Chapter 4 is no longer sufficient and we should consider at least a *semi-dynamic model* to define the blocking probability and a *dynamic model* to define the cut probability.

We shall study the evolution of the blocking and cut probabilities as functions of the traffic demand. The TRAFFIC DEMAND is defined as the ratio between the average duration of a call to the average duration between the arrivals of two successive calls. It is expressed in Erlang. (Note that the mean number of users per unit surface denoted λ_{M} in the semi-static model in Chapter 4 is precisely the traffic demand per unit surface.)

We consider the case with power limit, hence the term FC designates EDFC in the downlink and EUFC in the uplink. Unless otherwise specified, all the numerical applications are made using the default values specified in Section 2.2.5. Moreover, we consider in the present Chapter a finite network composed of 36 hexagonal cells on a torus (6 cells in each direction).

13.E.1 Constructor's schemes

The constructors for UMTS infrastructure propose admission algorithms which are decentralized but don't assure power allocation feasibility [64], [81]. Moreover constructor's scheme performance may not be evaluated analytically, to our knowledge.

Downlink. The downlink constructor's admission control is based on the powers transmitted by the base stations just before the arrival of a new user [64], [81]. [54] and [122] show that the power vector $P = (P_u)_{u\in\mathbf{U}}$ is solution of the problem

$$P = \min\left(\tilde{P}, AP + a\right)$$

where A and a are given by (3.7) and (3.8) respectively (where the sum $\sum_{m\in u}$ in the right-hand side of these equations doesn't comprise the new user) and $\tilde{P}$ is the vector of the maximal powers.

The downlink constructor's admission control is now described. Let P be the solution of the above problem. Given some constant $k \in (0,1)$ called DOWNLINK LOAD THRESHOLD by some authors, if

$$P \le k\tilde{P} \tag{13.27}$$

then the new user is admitted, otherwise he is rejected.

Uplink. The uplink constructor's admission control is based on the total interference at the base stations just before the arrival of a new user [64], [81]. Similarly to the downlink, using the arguments in [54] and [122] one may show that the interference vector $I = (I_u)_{u\in\mathbf{U}}$ is solution of the problem

$$I = \min\left(\tilde{I}, BI^{(t)} + b\right)$$

where B, b and $\tilde{I}$ are given by (3.26), (3.27) and (3.35) respectively. (Note that neither the sum $\sum_{m\in u}$ in (3.26) nor the infimum $\inf_{m\in u}$ in (3.35) should comprise the new user.)

The uplink constructor's admission control is now described. Let I be the solution of the above problem. Given some constant $k' \in (0,1)$ called UPLINK LOAD THRESHOLD by some authors, if

$$I \leq k'\tilde{I} \tag{13.28}$$

then the new user is admitted, otherwise he is rejected.

13.E.2 Comparison in a semi-dynamic context

We consider here a *semi-dynamic* traffic model (i.e. users don't move during their calls).

Downlink

Figures 13.14 and 13.15 show the blocking probability as function of traffic demand for the admission algorithms NSFC, FC and constructor and for cell radii $R = 1$ and 5km respectively. We study also cell radii $R = 0.525, 2, 3$ and 4km, but we don't show the curves here because the important phenomena are well illustrated in the figures corresponding to cell radii $R = 1$ and 5km.

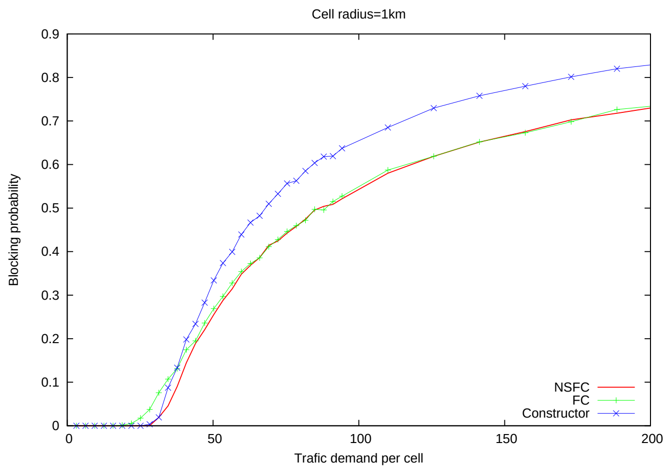

Figure 13.14: Comparison of the admission control algorithms for $R = 1$km

Erroneous admission. In theory, the constructor scheme doesn't assure the power allocation feasibility. We calculate a parameter called ERRONEOUS-ADMISSION representing the proportion of admission decisions which don't assure the power

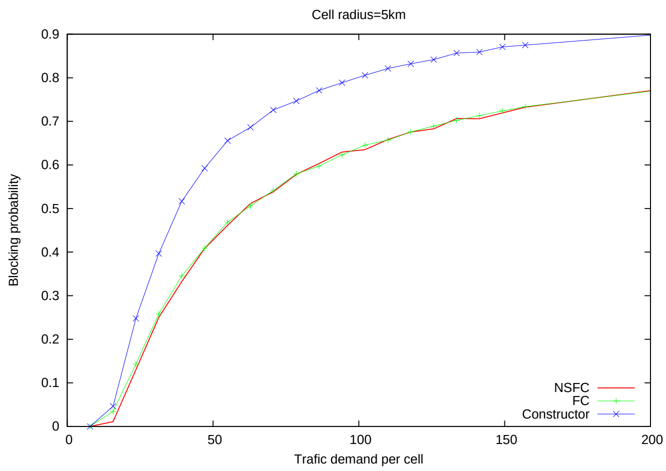

Figure 13.15: Comparison of the admission control algorithms for $R = 5$km

$R\backslash$**Algorithm**	**FC**	**Constructor**
$R = 0.525$km	17.8%	2.7%
$R = 1$	18%	0%
$R = 3$	20%	10%
$R = 5$	23.8%	33%

Table 13.1: Loss of capacity of FC and the constructor algorithm versus NSFC in the downlink.

allocation feasibility. In our simulation for the downlink the erroneous-admission is null.

In fact, the constructor scheme is firstly optimized by adjusting the so called load threshold denoted k in Equation (13.27). The above curves correspond to the best performance (which corresponds to $k = 0.75$). No such calibration is carried (there is no need) for the FC.

Loss of capacity. The capacity is defined as the traffic demand corresponding to a blocking probability of 2%. Table 13.1 gives the loss of capacity of FC and the constructor algorithm versus NSFC.

Uplink

Figures 13.16 and 13.17 show the blocking probability as function of traffic demand for the admission algorithms NSFC, SFC, FC and constructor for cell radii $R = 1$ and 3km respectively.

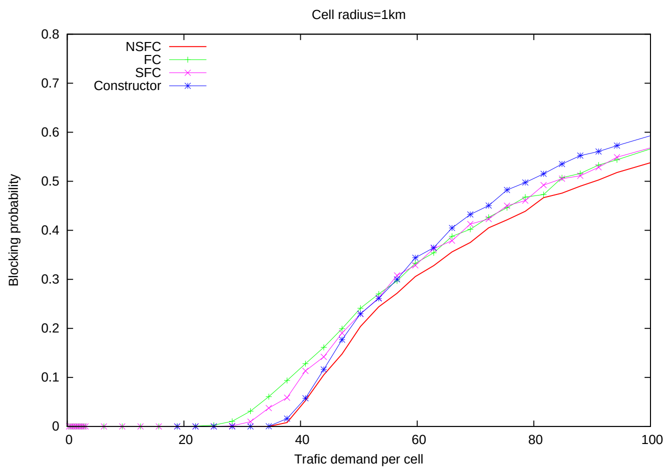

Figure 13.16: Comparison of the admission control algorithms for $R = 1$km

Erroneous admission. In theory, neither FC nor the constructor scheme assure the power allocation feasibility. Hence we have also here the notion of erroneous-admission which is represented in Figures 13.18 and 13.19 for FC and contructor respectively. We deduce that, for assuring the power allocation feasibility:

- FC is better than the constructor scheme for small cells;
- the constructor scheme is better than FC for moderate and large cells.

Note that the constructor scheme performance depends on the so called load threshold denoted k' in Equation (13.28). Adjusting this parameter would probably reduce the erroneous-admission for the constructor scheme, but we use in the above curves the most largely used value for the load threshold in the uplink $k' = 0.5$.

Loss of capacity. Table 13.2 gives the loss of capacity of FC, the SFC and the constructor algorithm versus NSFC.

Synthesis of the comparison

We deduce from the above results that, from the blocking probability point of view:

- constructor scheme performs better than FC for small cells and small blocking

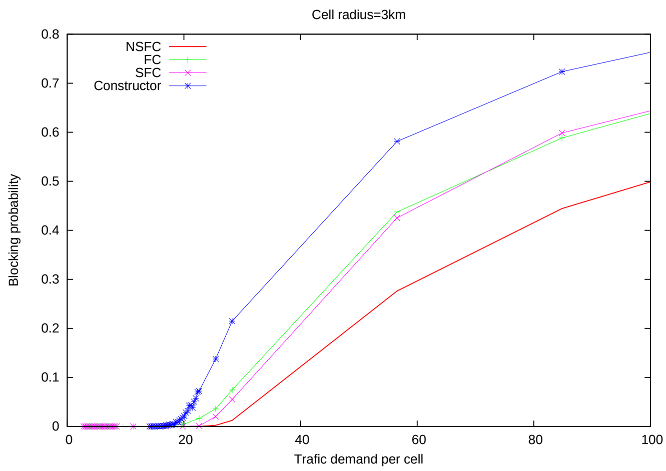

Figure 13.17: Comparison of the admission control algorithms for $R = 5$km

$R\backslash$**Algorithm**	**FC**	**SFC**	**Constructor**
$R = 0.525$km	21.7%	12.6%	0%
$R = 1$	23%	15.5%	0%
$R = 2$	13.8%	11.5%	11%
$R = 3$	21%	12.6%	31%
$R = 4$	24%	13.6%	100%

Table 13.2: Loss of capacity of FC and the constructor algorithm versus NSFC in the downlink.

- FC performs better than constructor scheme for large cells or large blocking
- the loss of capacity of FC versus NSFC is about 25%

From the power allocation feasibility point of view, the constructor scheme may assure the power allocation feasibility with high probability, but this requires to calibrate the so called load threshold. In the downlink, FC and constructor scheme assure the power allocation feasibility respectively with certainty and with high probability. Nevertheless, in the uplink, from the power allocation feasibility point of view, :

- FC performs better than the constructor scheme for small cells
- constructor scheme performs better than FC for moderate and large cells

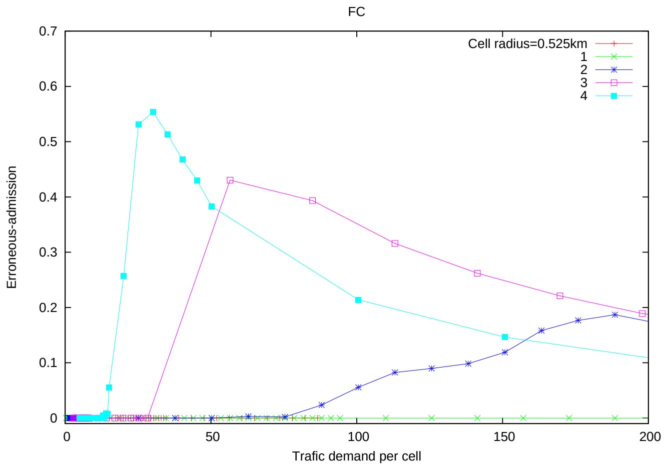

Figure 13.18: FC erroneous-admission

13.E.3 Comparison in a dynamic context

We consider here a *dynamic* traffic model (i.e. users may move during their calls). We consider the mobility model proposed in § 6.5. The average speed per user is denoted v and is expressed in kilometers per $\mu\times$ seconds where μ designates the average call duration. We consider the forced termination loss model presented in § 8.2.2.

The following numerical results concern the downlink, a cell radius $R = 1$ km, speeds $v = 0.1, 1, 10$ and a traffic demand less than 120 Erlang per cell. We observed numerically that the constructor scheme assures the power allocation feasibility in this context.

Figure 13.20 represents the blocking probability as function of the traffic demand for the three admission control schemes NSFC, FC and constructor; and for different speeds $v = 0.1, 1, 10$. We observe that the blocking probability decreases when the speed increases, but the relative loss of capacity (for fixed blocking probability) of the FC and constructor schemes versus the NSFC is approximatively the same for all the speeds.

Figure 13.21 represents the cut probability as functions of the traffic demand for the three admission control schemes NSFC, FC and constructor; and for different speeds $v = 0.1, 1, 10$. We observe that the cut probability increases with speed; and that the relative loss of capacity (for fixed cut probability) of the FC and constructor schemes versus the NSFC is approximatively the same as when we consider the blocking probability.

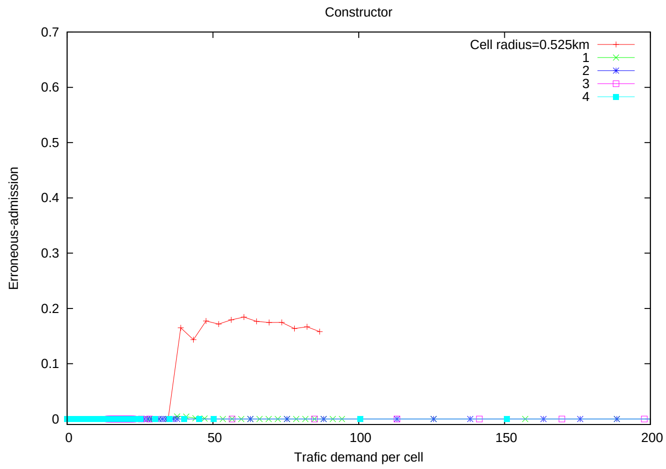

Figure 13.19: Constructor scheme erroneous-admission

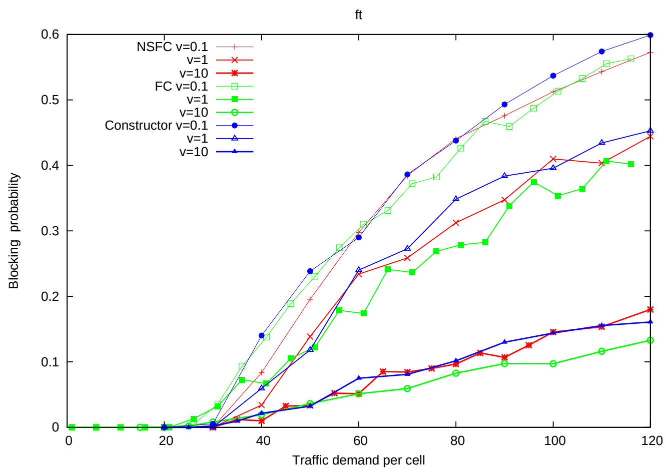

Figure 13.20: Comparison of the blocking probability for different admission control schemes NSFC, FC and constructor; and different speeds $\upsilon = 0.1, 1, 10$.

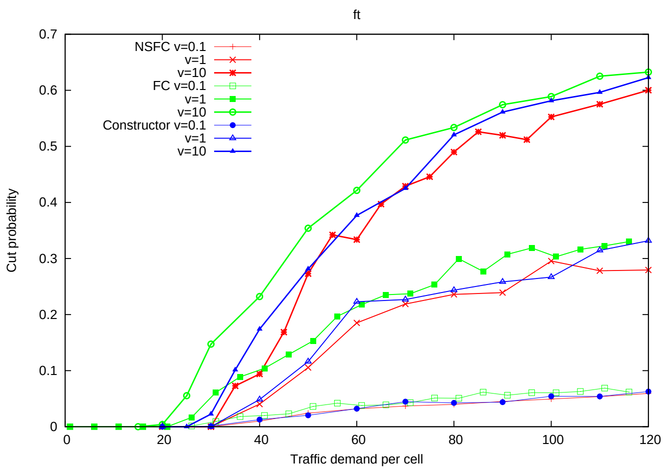

Figure 13.21: Comparison of the cut probability for different admission control schemes NSFC, FC and constructor; and different speeds $\upsilon = 0.1, 1, 10$.

Chapter 14

Appendix of Part II

14.A Classical queues

14.A.1 M/GI/1/PS

First, we recall the definition of a M/GI/1/PS queue. Users arrive at random times and require to transmit some random volume of data. The term M/GI (Kendell's notation) means that

- the interarrivals are i.i.d. with exponential distrubution (M);
- the required volumes are i.i.d. (GI);
- the interarrivals and required volumes are independent;

The term 1/PS, where PS is an abbreviation of PROCESSOR-SHARING, means that there is a single server and when there are n users in the queue each one is served at rate $1/n$.

The bit-rate at which data has to be trasmitted is not fixed, hence such a queue is well adpated for *elastic traffic*.

Lemma 22 *Consider a* M/GI/1/PS *queue with arrival rate* $\lambda \in \mathbb{R}_+^*$ *and mean required volume* μ^{-1} *(where* $\mu \in \mathbb{R}_+^*$*). Assume that the traffic demand* $\rho = \lambda/\mu < 1$*. Then the process* $\{N_t; t \geq 0\}$ *counting the number of users in this queue is ergodic and admits as limiting distribution the* geometric distribution *with parameter* $1-\rho$*, that is*

$$\Pi(\nu) = (1-\rho)\,\rho^{\nu}, \quad \nu \in \mathbb{N} \tag{14.1}$$

Under the limiting distribution, the mean number of users, denoted $\mathbf{E}[N]$*, equals*

$$\mathbf{E}[N] = \frac{\rho}{1-\rho} \tag{14.2}$$

the expected delay, denoted $\bar{T}$, equals

$$\bar{T} = \frac{\rho}{\lambda (1 - \rho)}$$

and the expected throughput, denoted $\bar{r}$, equals

$$\bar{r} = \frac{1}{\mu \bar{T}} = 1 - \rho$$

If the required volumes are exponential, then the process $\{N_t; t \geq 0\}$ is Markov and admits the following generator

$$\begin{cases} q(\nu, \nu + 1) = \lambda, & \nu \in \mathbb{N} \\ q(\nu, \nu - 1) = \mu, & \nu \in \mathbb{N}^* \end{cases}$$

Proof. Cf. [36] and [103, Proposition 7.13 p.194] for ergodicity and the form of the limiting distribution. The expression of the expected delay is deduced from Little's formula. The expected throughput is the average required volume μ^{-1} divided by the expected delay. ■

Note that the TRAFFIC DEMAND ρ equals the ratio between the average volume required by a call to the average duration between the arrivals of two successive calls. This should be distinguished from the expected throughput $\bar{r}$ observed by a given user under the limiting distribution, even though these two quantities have the same unit.

Remark 29 INSENSITIVITY PROPERTY. *Consider a* M/GI/1 *queue with a service discipline which is* WORK CONSERVING *(i.e. the server is active as long as there is some volume in the queue). This includes in particular FIFO (first-in-first-out), LIFO (last-in-first-out) and PS (processor-sharing) disciplines. Define the* WORKLOAD PROCESS *to be the remaining required volume at a given time. It is well kown that the workload process evolution (and in particular its limiting behaviour) doesn't depend on the service discipline (see for example [17, §2.1 p.76]).*

Nevertheless, the CONGESTION PROCESS *counting the number of users in the queueing station depends on the service discipline. Indeed, as noted in [103, §7.5 p.202], the limiting distribution for PS and LIFO is the same. It is given by Equation (14.1). Observe that the limiting distribution depends on the distribution of the required volume only through its mean μ. This is the so-called* INSENSITIVITY PROPERTY. *The* M/GI/1 *PS and LIFO queues are said to be* INSENSITIVE.

It is not the case for the M/GI/1 *FIFO queue. Indeed the limiting distribution for FIFO depends on the distribution of the required volume. In particular, the expectation of the number of users under the limiting distribution for FIFO is given by Pollaczek-Kintchine formula [17, Formula (3.2.10) p.197] which depends on the second moment of the distribution of the required volume.*

Multiclass M/GI/1/PS **queue**

The above results may be generalized to the case where the users belong to different classes. Let $\mathbb{D}$ be the set of classes assumed finite and denote by $x \in \mathbb{D}$ a particular class. The state of the queue is denoted by $\nu = (\nu_x; x \in \mathbb{D})$ where ν_x designates the number of active users of class x. The state space is denoted by $\mathbb{M} = \mathbb{N}^{\mathbb{D}}$. Denote $\nu(\mathbb{D}) = \sum_{x \in \mathbb{D}} \nu_x$.

Lemma 23 *Consider a* M/GI/1/PS *multiclass queue. Assume that users of each class $x \in \mathbb{D}$ arrive with rate $\lambda_x \in \mathbb{R}_+^*$ and require to transmit a volume of mean μ_x^{-1} (where $\mu_x \in \mathbb{R}_+^*$). Let $\rho_x = \lambda_x / \mu_x$ be the traffic demand for class x and assume that the total traffic demand $\rho(\mathbb{D}) = \sum_{x \in \mathbb{D}} \rho_x < 1$. Then the process $\{N_t; t \geq 0\}$ counting the number of users of different classes in this queue is ergodic and admits as limiting distribution*

$$\Pi(\nu) = \Pi(0)\, \nu(\mathbb{D})! \prod_{x \in \mathbb{D}} \frac{\rho_x^{\nu_x}}{\nu_x!}, \quad \nu \in \mathbb{M}$$

where

$$\Pi(0) = 1 - \rho(\mathbb{D})$$

Under the limiting distribution, for a given class $x \in \mathbb{D}$, the mean number of users, denoted $\mathbf{E}[N(x)]$, equals

$$\mathbf{E}[N(x)] = \frac{\rho_x}{1 - \rho(\mathbb{D})}$$

the expected delay, denoted $\bar{T}(x)$, equals

$$\bar{T}(x) = \frac{\rho_x}{\lambda_x (1 - \rho(\mathbb{D}))}$$

and the expected throughput, denoted $\bar{r}(x)$, equals

$$\bar{r}(x) = \frac{1}{\mu_x \bar{T}_x} = 1 - \rho(\mathbb{D})$$

If the required volumes are exponential, then the process $\{N_t; t \geq 0\}$ is Markov and admits the following generator

$$\begin{cases} q(\nu, \nu + \delta_x) = \lambda_x, & \nu \in \mathbb{M} \\ q(\nu, \nu - \delta_x) = \mu_x \frac{\nu_x}{\nu(\mathbb{D})}, & \nu \in \mathbb{M}, \nu_x > 0 \end{cases}$$

where δ_x designates the vector having coordinate 1 *at position x and* 0 *elsewhere.*

Proof. (Cf. [38].) We recall here a simple proof when the required volumes are exponential. The process $\{N(t); t \geq 0\}$ describing the number of users of different classes in the queue is a continuous-time Markov process with discrete state space $\mathbb{M} = \mathbb{N}^{\mathbb{D}}$. It is easy to see that the process $\{N(t); t \geq 0\}$ is regular and irreducible and, if $\rho(\mathbb{D}) < 1$, then it admits Π as invariant distribution; and hence it is ergodic.

The expression of the expected delay of a given class x is deduced from Little's formula. The expected throughput of class x is the average required volume μ_x^{-1} divided by the expected delay. ■

14.A.2 M/GI/∞ and loss queue

First, we recall the definition of a M/GI/∞ queue. Users arrive at random times and require to transmit data for a given duration at rate 1. The term M/GI (Kendell's notation) means that

- the interarrivals are i.i.d. with exponential distrubution (M);
- the required transmission durations are i.i.d. (GI);
- the interarrivals and required volumes are independent;

The term ∞ means that the number of servers is infinite, or, in other terms, when there are n users in the queue each one is served by a specific server at rate 1.

The bit-rate at which data has to be trasmitted is fixed, hence such a queue is well adpated for *streaming traffic*.

Lemma 24 *Consider a* M/GI/∞ *queue with arrival rate* $\lambda \in \mathbb{R}_+^*$ *and mean required duration* μ^{-1} *(where* $\mu \in \mathbb{R}_+^*$*). Then the process* $\{N_t; t \geq 0\}$ *counting the number of users in this queue is ergodic and admits as limiting distribution the* Poisson distribution *with mean* $\rho = \lambda/\mu$*, that is*

$$\Pi(\nu) = e^{-\rho}\frac{\rho^\nu}{\nu!}, \quad \nu \in \mathbb{N}$$

If the required durations are exponential, then the process $\{N_t; t \geq 0\}$ *is Markov and admits the following generator*

$$\begin{cases} q(\nu, \nu+1) = \lambda, & \nu \in \mathbb{N} \\ q(\nu, \nu-1) = \mu\nu, & \nu \in \mathbb{N}^* \end{cases}$$

Proof. Cf. [37, §2.2 p.42,47]. ∎

Note that the TRAFFIC DEMAND ρ equals the ratio between the average duration required by a call to the average duration between the arrivals of two successive calls. It is expressed in ERLANG.

M/GI/C loss queue

Assume now that the number of servers is some positive integer C. Assume also that we loose the users that arrive when there are already C users in the queue.

Lemma 25 *Consider a* M/GI/C *loss queue with arrival rate* $\lambda \in \mathbb{R}_+^*$ *and mean required duration* μ^{-1} *(where* $\mu \in \mathbb{R}_+^*$*). Then the process* $\{N_t; t \geq 0\}$ *counting the number of users in this queue is ergodic and admits as limiting distribution the* Poisson distribution *with mean* $\rho = \lambda/\mu$ truncated *at* C*, that is*

$$\Pi^f(\nu) = \frac{\frac{\rho^\nu}{\nu!}}{\sum_{k=0}^{C}\frac{\rho^k}{k!}}, \quad \nu = 0, 1, \ldots, C$$

The blocking probability is given by

$$b = \frac{\frac{\rho^C}{C!}}{\sum_{k=0}^{C} \frac{\rho^k}{k!}} \tag{14.3}$$

Under the limiting distribution, the mean number of users, denoted $\mathbf{E}[N]$, *equals*

$$\mathbf{E}[N] = \rho(1-b)$$

which satisfies

$$\rho(1-b) \le C, \quad \lim_{\rho\to\infty} \rho(1-b) = C$$

Proof. Cf. [37, §2.2 p.42]. ■

Formula (14.3) is called ERLANG'S FORMULA.

Remark 30 INSENSITIVITY. *The limiting distributions for the number of users in the* M/GI/∞ *and* M/GI/C *loss queues depend on the distribution of the required duration only through its mean* μ. *Then the* M/GI/∞ *and* M/GI/C *loss queues are* insensitive.

Multiclass M/GI/∞

The above results may be generalized to the case where the users belong to different classes. Let $\mathbb{D}$ be the set of classes assumed finite and denote by $x \in \mathbb{D}$ a particular class. The state of the queue is denoted by $\nu = (\nu_x; x \in \mathbb{D})$ where ν_x designates the number of active users of class x. The state space is denoted by $\mathbb{M} = \mathbb{N}^{\mathbb{D}}$. Denote $\nu(\mathbb{D}) = \sum_{x\in\mathbb{D}} \nu_x$.

Lemma 26 *Consider a* M/GI/∞ *multiclass queue. Assume that users of each class* $x \in \mathbb{D}$ *arrive with rate* $\lambda_x \in \mathbb{R}_+^*$ *and require to transmit for a duration of mean* μ_x^{-1} *(where* $\mu_x \in \mathbb{R}_+^*$*). Let* $\rho_x = \lambda_x/\mu_x$ *be the traffic demand for class* x *and let* $\rho(\mathbb{D}) = \sum_{x\in\mathbb{D}} \rho_x$ *be the total traffic demand. The process* $\{N_t; t \ge 0\}$ *counting the number of users of different classes in this queue is ergodic and admits as limiting distribution*

$$\Pi(\nu) = e^{-\rho(\mathbb{D})} \prod_{x\in\mathbb{D}} \frac{\rho_x^{\nu_x}}{\nu_x!}$$

which is the distribution of a Poisson process on $\mathbb{D}$ *with intensity measure* $\{\rho_x; x \in \mathbb{D}\}$.

If the required transmission durations are exponential, then the process $\{N_t; t \ge 0\}$ *is Markov and admits the following generator*

$$\begin{cases} q(\nu, \nu+\delta_x) = \lambda_x, & \nu \in \mathbb{M} \\ q(\nu, \nu-\delta_x) = \mu_x \nu_x, & \nu \in \mathbb{M}, \nu_x > 0 \end{cases}$$

where δ_x *designates the vector having coordinate* 1 *at position* x *and* 0 *elsewhere.*

Multiclass loss queue

Lemma 27 *Consider the setting of Lemma 26. Assume now that the state space is*

$$\mathbb{M}^{\mathrm{f}} = \left\{ \nu \in \mathbb{N}^{\mathbb{D}} : \sum_{x \in \mathbb{D}} \nu_x \varphi(x) \leq C \right\}$$

where C is a given positive constant and $\varphi : \mathbb{D} \to \mathbb{R}_+^*$ is a given function. Assume also that all user whose arrival would result in taking the process to a state outside $\mathbb{M}^{\mathrm{f}}$ is lost. The process $\{N_t; t \geq 0\}$ counting the number of users of different classes in this queue is ergodic and admits as limiting distribution

$$\Pi^f(\nu) = \Pi^f(0) \prod_{x \in \mathbb{D}} \frac{\rho_x^{\nu_x}}{\nu_x!}$$

where

$$\Pi^f(0) = \left(\sum_{\nu \in \mathbb{M}^{\mathrm{f}}} \prod_{x \in \mathbb{D}} \frac{\rho_x^{\nu_x}}{\nu_x!} \right)^{-1}$$

The blocking probability for class x is given by

$$b_x = 1 - \sum_{\nu \in \mathbb{M}_x^{\mathrm{f}}} \Pi^f(\nu), \quad \text{where } \mathbb{M}_x^{\mathrm{f}} = \left\{ \nu \in \mathbb{N}^{\mathbb{D}} : \sum_{y \in \mathbb{D}} \nu_y \varphi(y) \leq C - \varphi(x) \right\}$$

Proof. Cf. [73] and [104]. ■

Assume now that φ takes interger values, that is $\varphi : \mathbb{D} \to \mathbb{N}^*$. The blocking probabilities may be calculate by using the following algorithm.

Algorithm 1 Kaufman-Roberts *algorithm [73, 104]. Let $q(n)$ be the probability that the number of users is n, that is $q(n) = \sum_{\nu \in \mathbb{N}^{\mathbb{D}}:\nu(D)=n} \Pi^f(\nu)$. Then $q(\cdot)$ satisfies the following equations*

$$\sum_{n=0}^{C} q(n) = 1, \quad \text{and } q(n) = \sum_{x \in \mathbb{D}} \rho_x \varphi(x)\, q(c - \varphi(x)), \quad n = 0, \ldots, C$$

and the blocking probabilities are given by

$$b_x = 1 - \sum_{n=0}^{C-\varphi(x)} q(n)$$

14.B Fairness

We present here some general basic results on the ressource allocation problem, and particularly of the fairness issue. There is a rich litterature on this subject, refer for example to [23] and the references therein.

14.B.1 Definitions

We assume that we have N entities which we index with $i = 1, \ldots, N$ to which we have to allocate RESOURCES $x = (x(1), \ldots, x(N))$ respecting some CONSTRAINT in the form $x \in \mathcal{X}$ where $\mathcal{X}$ is some given subset of $\mathbb{R}_+^N$.

Definition 7 FAIRNESS.

[99, Definition 1] A vector $x \in \mathcal{X}$ is called MAX-MIN FAIR *if for each $i \in \{1, \ldots, N\}$ increasing some component $x(i)$ must be at the expense of decreasing some already smaller component. (In other words we may not increase $x(i)$ without decreasing one or more $x(j)$ among those having a lower value: $x(j) < x(i)$.)*

[50] A vector $x \in \mathcal{X}$ is called PROPORTIONALLY FAIR *if for each other vector $y \in \mathcal{X}$, the sum of relative variations is non-positive:*

$$\sum_{i=1}^{N} \frac{y(i) - x(i)}{x(i)} \leq 0$$

with the convention $\frac{a}{0} = +\infty$ for all $a \geq 0$ which will be adopted from now on.

[91] For a given vector $w = (w(1), \ldots, w(N))$ with real positive components and a given $\alpha > 0$, a vector $x \in \mathcal{X}$ is called (α, w)-PROPORTIONALLY FAIR *if for each other vector $y \in \mathcal{X}$, the sum of relative variations pondered by coefficients w is non-positive:*

$$\sum_{i=1}^{N} w(i) \frac{y(i) - x(i)}{x(i)^{\alpha}} \leq 0$$

Let $\mathbf{1}$ the vector with N coordinates equal to 1. When $w = \mathbf{1}$ we say that x is α-PROPORTIONALLY FAIR. *When $\alpha = 1$, we say that x is a* w-WEIGHED PROPORTIONALLY FAIR.

Definition 8 OPTIMALITY.

A vector $x \in \mathcal{X}$ is called (GLOBALLY) OPTIMAL *if it maximizes $\sum_{i=1}^{N} x(i)$. (If users pay proportionally to the allocated resource, the revenue is proportional to $\sum_{i=1}^{N} x(i)$.)*

A vector $x \in \mathcal{X}$ is called (STRICTLY) PARETO OPTIMAL *if there is no solution y dominating it. (We say that y dominate x if $y(i) \geq x(i)$ for each i, with at least one strict inequality.)*

Remark 31 *Note that a α-proportionally fair vector is globally optimal when $\alpha = 0$; and proportionally fair when $\alpha = 1$.*

14.B.2 Basic properties

Lemma 28 *If a max-min fair vector exists on a set $\mathcal{X}$, then it is unique and strictly Pareto optimal.*

Proof. For uniqueness cf. [99, Proposition 1]. For Pareto optimality cf. [99, Proposition 3]. ∎

Lemma 29 *If a proportionally fair vector exists on a set $\mathcal{X}$, then it is unique and strictly Pareto optimal.*

Proof. Uniqueness. Note that a vector $x \in \mathcal{X}$ is proportionally fair iff for each other vector $y \in \mathcal{X}$,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{y(i)}{x(i)} \leq 1$$

That is the arithmetic mean of $\{y(i)/x(i)\}$ is not larger than 1. Since the harmonic mean is not larger than the arithmetic mean, we get

$$\frac{N}{\sum_{i=1}^{N}\frac{x(i)}{y(i)}} \leq \frac{1}{N}\sum_{i=1}^{N}\frac{y(i)}{x(i)} \leq 1$$

with equality if and only if $y = x$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\frac{x(i)}{y(i)} \geq \left(\frac{1}{N}\sum_{i=1}^{N}\frac{y(i)}{x(i)}\right)^{-1} \geq 1$$

If y is also proportionally fair, then we get equalities in the above display, hence $y = x$.

Pareto optimality. Assume that some $x \in \mathcal{X}$ is not Pareto optimal. Then there exists some $y \in \mathcal{X}\setminus\{x\}$ dominating x, i.e. $y(i) \geq x(i)$ for each i, with at least one strict inequality. Hence

$$\sum_{i=1}^{N}\frac{y(i) - x(i)}{x(i)} > 0$$

Therefore x is not proportionally fair.

We deduce that if some $x \in \mathcal{X}$ is proportionally fair, then it is strictly Pareto optimal. ■

Proposition 59 *If a proportionally fair vector exists on a set $\mathcal{X}$, then it is solution of* $\max_{x\in\mathcal{X}}\prod_{i=1}^{N}x(i)$ *which is equivalent to* $\max_{x\in\mathcal{X}}\sum_{i=1}^{N}\log x(i)$ *with the convention* $\log 0 = -\infty$ *which will be adopted from now on.*

More generally, if a w-weighed proportionally fair vector exists on a set $\mathcal{X}$, then it is solution of $\max_{x\in\mathcal{X}}\prod_{i=1}^{N}x(i)^{w(i)}$ *which is equivalent to* $\max_{x\in\mathcal{X}}\sum_{i=1}^{N}w(i)\log x(i)$.

Proof. Let $x \in \mathcal{X}$ be $(1, w)$-weighed proportionally fair. For each other vector $y \in \mathcal{X}$,

$$\frac{1}{N}\sum_{i=1}^{N}w(i)\frac{y(i)}{x(i)} \leq 1$$

Since the geometric mean is not larger than the arithmetic mean, we get

$$\left(\prod_{i=1}^{N}\left(\frac{y(i)}{x(i)}\right)^{w(i)}\right)^{1/N} \leq \frac{1}{N}\sum_{i=1}^{N}w(i)\frac{y(i)}{x(i)} \leq 1$$

Then

$$\prod_{i=1}^{N} \frac{y(i)^{w(i)}}{x(i)^{w(i)}} \leq 1$$

which is the desired result. ■

Theorem 13 Generalized fairness. *Assume that*

$$\mathcal{X} = \left\{ x \in \mathbb{R}_+^N : Ax \leq c \right\}$$

for some given matrix $A \in \mathbb{R}_+^{N \times N}$ *and vector* $c \in \mathbb{R}_+^N$.

(i) A vector is (α, w)*-proportionally fair if and only it it solves the problem*

$$\max_{x \in \mathcal{X}} \sum_{i=1}^{N} w(i) f_\alpha(x(i)), \quad \text{where } f_\alpha(u) = \begin{cases} \log u & \text{if } \alpha = 1 \\ \frac{u^{1-\alpha}}{1-\alpha} & \text{otherwise} \end{cases} \tag{14.4}$$

(ii) The (α, w)*-proportionally fair vector approaches the max-min vector as* $\alpha \to \infty$.

Proof. (Cf. [91]) ■

Example 14 *[3] Consider three connections AB, BC and AC as follows*

$$\frac{A \overset{x(1)}{------} B \overset{x(2)}{------} C}{x(3)}$$

Each link has capacity 1*. Hence the constraint writes*

$$\begin{cases} x(1) + x(2) \leq 1 \\ x(2) + x(3) \leq 1 \end{cases}$$

A max-min fair vector is

$$x_m(1) = x_m(2) = x_m(3) = \frac{1}{2}$$

An optimal vector is

$$x_o(1) = x_o(2) = 1, x_o(3) = 0$$

A proportionally fair vector is

$$x_p(1) = x_p(2) = \frac{2}{3}, x_p(3) = \frac{1}{3}$$

The max-min vector x_m *is not proportionally fair, since*

$$\sum_{i=1}^{N} \frac{x_p(i) - x_m(i)}{x_m(i)} = 2\frac{\frac{2}{3} - \frac{1}{2}}{\frac{1}{2}} + \frac{\frac{1}{3} - \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3} > 0$$

The optimal vector x_o is not proportionally fair, since

$$\sum_{i=1}^{N} \frac{x_m(i) - x_o(i)}{x_o(i)} = 2\frac{\frac{1}{2}-1}{1} + \frac{\frac{1}{2}-0}{0} = \infty > 0$$

Note that $x_o(3) = 0$, i.e. the optimal vector gives 0 resources to connection AC. The proposition 59 shows that, when possible (by the constraints), the proportionally fair vector gives always positive allocations to each connection.

14.B.3 Max-min fairness

Lemma 30 *If a max-min fair vector exists on a set $\mathcal{X}$, then it is solution of*

$$\max_{x\in\mathcal{X}} \min_{i=1,\ldots,N} x(i)$$

Proof. Let $x \in \mathcal{X}$ be max-min fair. Assume that there exists some vector $y \in \mathcal{X}$ such that

$$\min_{j=1,\ldots,N} x(j) < \min_{j=1,\ldots,N} y(j)$$

Let $x(i) = \min_{j=1,\ldots,N} x(j)$. We have

$$y(i) \geq \min_{j=1,\ldots,N} y(j) > \min_{j=1,\ldots,N} x(j) = x(i)$$

Then by replacing x by y we increase $x(i)$ without decreasing any $x(j) < x(i)$ (because no such $x(j)$ exists). This contradicts the fact that x is max-min fair. ■

Lemma 31 *The following optimization problems are equivalent*

$$z = \max_{x\in\mathcal{X}} \min_{i=1,\ldots,N} x(i) \tag{P1}$$

$$\begin{array}{ll} \max & z = \min_{i=1,\ldots,N} x(i) \\ \text{sub} & x \in \mathcal{X} \end{array} \tag{P2}$$

$$\begin{array}{ll} \max & z \\ \text{sub} & z \leq x(i), \quad \forall i = 1,\ldots,N \\ & x \in \mathcal{X} \end{array} \tag{P3}$$

More precisely, for all $k, j \in \{1,2,3\}$, a solution (z^k, x^k) of problem (Pk) is also solution of problem (Pj).

Proof. (P1)⇔(P2)? It is clear that

$$z^1 = \max\left\{\min_{i=1,\ldots,N} x(i), x \in \mathcal{X}\right\} = z^2$$

$$x^1 = \arg\max\left\{\min_{i=1,\ldots,N} x(i), x \in \mathcal{X}\right\} = x^2$$

(P2)⇔(P3)? From

$$\begin{cases} z^2 \leq x^2(i), & \forall i = 1, \dots, N \\ x^2 \in \mathcal{X} \end{cases}$$

we get $z^2 \leq z^3$. Let now $z'^3 = \min_{i=1,\dots,N} x^3(i) \geq z^3$. From

$$\begin{cases} z'^3 = \min_{i=1,\dots,N} x^3(i) \\ x^3 \in \mathcal{X} \end{cases}$$

we deduce that $z'^3 \leq z^2$. Then

$$z'^3 \leq z^2 \leq z^3 \leq z'^3$$

Hence $z^2 = z^3$, therefore (z^2, x^2) is a solution of (P3). Moreover $z'^3 = z^3$, then $z^3 = \min_{i=1,\dots,N} x^3(i)$, therefore (z^3, x^3) is a solution of (P2). ■

For each $x \in \mathbb{R}^N$ denote by $\bar{x}$ the vector defined by $\bar{x}(i) = \min_{j=1,\dots,N} x(j)$.

Lemma 32 *Suppose now that $\mathcal{X}$ satisfies the following property: $x \in \mathcal{X} \Rightarrow \bar{x} \in \mathcal{X}$. Then the following optimization problems are equivalent to (P1)-(P3):*

$$\begin{array}{ll} \max & z \\ \text{sub} & z = x(i), \quad \forall i = 1, \dots, N \\ & x \in \mathcal{X} \end{array} \tag{P4}$$

$$\begin{array}{lll} \max & z = \sum_{i=1}^{N} w(i) x(i) & \\ \text{sub} & x \in \mathcal{X}, & \\ & x(i) = x(j) & \forall i, j = 1, \dots, N \end{array} \tag{P5}$$

where the $w(i)$ are non-negative reals satisfying $\sum_{i=1}^{N} w(i) = 1$. More precisely, for all $k \in \{1, 2, 3\}, j \in \{4, 5\}$, if (z^j, x^j) is a solution of problem (Pj) then it is also a solution of problem (Pk); and if (z^k, x^k) is a solution of problem (Pk) then $(z^k, \bar{x}^k)$ is a solution of problem (Pj).

Proof. (P3)⇔(P4)? From

$$\begin{cases} z^4 = x^4(i), & \forall i = 1, \dots, N \\ x^4 \in \mathcal{X} \end{cases}$$

we get $z^4 \leq z^3$. Let now $z'^3 = \min_{i=1,\dots,N} x^3(i) \geq z^3$. From

$$\begin{cases} z'^3 = \min_{i=1,\dots,N} \bar{x}^3(i) \\ \bar{x}^3 \in \mathcal{X} \end{cases}$$

we deduce that $z'^3 \leq z^4$. Then

$$z'^3 \leq z^4 \leq z^3 \leq z'^3$$

Hence $z^4 = z^3$, therefore (z^4, x^4) is a solution of (P3). Moreover $z'^3 = z^3$, then $z^3 = \min_{i=1,\dots,N} x^3(i)$, therefore $(z^3, \bar{x}^3)$ is a solution of (P4).

(P4)⇔(P5)? Obvious. ■

14.B.4 Case of a linear constraint

Fairness

Proposition 60 *We assume that we have N entities (typically users) which we index with $i = 1, \ldots, N$ for which we have to allocate resources $x = (x(1), \ldots, x(N))$ (typically bit-rate) respecting some constraint (typically power allocation feasibility condition) in the form*

$$\sum_{i=1}^{N} \gamma(i)\, x(i) = 1 \tag{14.5}$$

where $\gamma(i)$ are some given positive constants. The α-proportionally fair vector is

$$x_\alpha(i) = \frac{\gamma(i)^{-1/\alpha}}{\sum_{j=1}^{N} \gamma(j)^{1-1/\alpha}} \tag{14.6}$$

The max-min fair vector is

$$x_\infty(i) = \frac{1}{\sum_{j=1}^{N} \gamma(j)} \tag{14.7}$$

The proportional fair vector is

$$x_1(i) = \frac{1}{N\gamma(i)} \tag{14.8}$$

An optimal vector is

$$x_0(i) = \frac{1\{i \in J\}}{\sum_{j \in J} \gamma(j)} \tag{14.9}$$

where $J \subset \{j : \gamma(j) = \min_i \gamma(i)\}$.

Proof. For $\alpha > 0$, let

$$f_\alpha(u) = \begin{cases} \log u & \text{if } \alpha = 1 \\ \frac{u^{1-\alpha}}{1-\alpha} & \text{otherwise} \end{cases}$$

which is clearly a strictly concave function. Let

$$g_\alpha(x) = \sum_{i=1}^{N} f_\alpha(x(i)), \quad \varphi(x) = \sum_{i=1}^{N} \gamma(i)\, x(i) - 1$$

The function $g_\alpha(x)$ is strictly concave on $\mathbb{R}^N$ since it is the sum of strictly concave functions. Then it admits a unique maximum on $\{x \in \mathbb{R}^N : \varphi(x) = 0\}$.

Suppose that x is a maximum point of g_α on $\{x \in \mathbb{R}^N : \varphi(x) = 0\}$. Then, by the Lagrange multiplier Theorem, there exists λ such that

$$\frac{\partial g_\alpha}{\partial x(i)} = \lambda \frac{\partial \varphi}{\partial x(i)}$$

Hence $x(i)^{-\alpha} = \lambda\gamma(i)$ which combined with the constraint $\varphi(x) = 0$ gives the expression of the α-proportionally fair vector given in the claim. (Here is an alternative proof when $\alpha \neq 1$. Observe that

$$\sum_{i=1}^{N} x(i)^{1-\alpha} = \sum_{i=1}^{N} (\gamma(i)\, x(i))^{1-\alpha}\, \gamma(i)^{\alpha-1}$$

Using Hölder's inequality [60, Equation (2.8.5) p. 25] for $a(i) = (\gamma(i)\, x(i))^{1-\alpha}$, $b(i) = \gamma(i)^{\alpha-1}$, $k = \frac{1}{1-\alpha}$ and $k' = \frac{1}{\alpha}$, we get

$$\left(\sum_{i=1}^{N} x(i)^{1-\alpha}\right)^{kk'} \leq \left(\sum_{i=1}^{N} \gamma(i)\, x(i)\right)^{k'} \left(\sum_{i=1}^{N} \gamma(i)\right)^{k}$$

Then

$$g_\alpha(x) \leq \frac{1}{1-\alpha}\left(\sum_{i=1}^{N} \gamma(i)\, x(i)\right)^{1-\alpha} \left(\sum_{j=1}^{N} \gamma(j)^{1-1/\alpha}\right)^{\alpha}$$

then if $\sum_{i=1}^{N} \gamma(i)\, x(i) = 1$ we get $g_\alpha(x) \leq \frac{1}{1-\alpha}\left(\sum_{j=1}^{N} \gamma(j)^{1-1/\alpha}\right)^{\alpha} = g_\alpha(x_\alpha)$.)

The limit $\lim_{\alpha\uparrow\infty} x_\alpha$ is max-min fair by Theorem 13, hence we get (14.7).

Taking $\alpha = 1$ in the expression of x_α gives the expression (14.8) of the proportional fair vector.

We now look for the optimal vector. Fix $j \in \{1, \ldots, N\}$ such that $\gamma(j) = \min_i \gamma(i)$. Observe that

$$\begin{aligned} g_0(x) &= \sum_{i=1}^{N} x(i) \\ &= \frac{1}{\gamma(j)} \sum_{i=1}^{N} \gamma(j)\, x(i) \\ &\leq \frac{1}{\gamma(j)} \sum_{i=1}^{N} \gamma(i)\, x(i) \\ &= \frac{1}{\gamma(j)} = g_0(x_0) \end{aligned}$$

where x_0 is given by (14.9). This shows that x_0 is optimal. (If we take $J = \{j : \gamma(j) = \min_i \gamma(i)\}$, then $\lim_{\alpha\downarrow 0} x_\alpha = x_0$. Proof: Rearranging x_α as follows

$$x_\alpha(i) = \left[\sum_{j=1}^{N} \gamma(j) \left(\frac{\gamma(i)}{\gamma(j)}\right)^{1/\alpha}\right]^{-1}$$

and observing that

$$\lim_{\alpha\downarrow 0} \left(\frac{\gamma(i)}{\gamma(j)}\right)^{1/\alpha} = \lim_{\alpha\downarrow 0} \exp\left(\frac{1}{\alpha} \ln \frac{\gamma(i)}{\gamma(j)}\right) = \begin{cases} \infty & \gamma(i) > \gamma(j) \\ 1 & \gamma(i) = \gamma(j) \\ 0 & \gamma(i) < \gamma(j) \end{cases}$$

gives the desired result.) ■

Remark 32 *The allocation (14.7) is the scheduling algorithm proposed in [16] for UMTS networks. Unfortunately the associated Markov process is not balanced, hence we have not an explicit expression of the stationary distribution. We propose the allocation (14.8) as a scheduling algorithm for UMTS networks, which permits to get explicit expressions of the throughput of elastic traffic.*

Corollary 14 *In the conditions of Proposition 60, the* (α, w)*-proportionally fair vector is*

$$x_{\alpha,w}(i) = \frac{w(i)^{1/\alpha}\gamma(i)^{-1/\alpha}}{\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}}$$

and the value of the maximum in the optimization problem (14.4) is

$$\frac{1}{1-\alpha}\left(\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}\right)^{\alpha}$$

Proof. Assume first that $\alpha \neq 1$. Let

$$x'(i) = w(i)^{\frac{1}{1-\alpha}} x(i)$$
$$\gamma'(i) = w(i)^{-\frac{1}{1-\alpha}} \gamma(i)$$

Then the constraint $\sum_{i=1}^{N}\gamma(i)x(i) = 1$ writes $\sum_{i=1}^{N}\gamma'(i)x'(i) = 1$. The function to maximize in the optimization problem (14.4) is

$$\sum_{i=1}^{N} w(i) f_\alpha(x(i)) = \sum_{i=1}^{N} f_\alpha(x'(i))$$

Then x is (α, w)-proportionally fair with respect to the constraint $\sum_{i=1}^{N}\gamma(i)x(i) = 1$ iff x' is α-proportionally fair with respect to the constraint $\sum_{i=1}^{N}\gamma'(i)x'(i) = 1$. Hence $x_{\alpha,w}(i) = w(i)^{-\frac{1}{1-\alpha}} x'_\alpha(i)$. Using the expression of x'_α given in Proposition 60 we get

$$\begin{aligned} x_{\alpha,w}(i) &= w(i)^{-\frac{1}{1-\alpha}} \frac{\gamma'(i)^{-1/\alpha}}{\sum_{j=1}^{N}\gamma'(j)^{1-1/\alpha}} \\ &= w(i)^{-\frac{1}{1-\alpha}} \frac{w(i)^{\frac{1}{1-\alpha}\frac{1}{\alpha}}\gamma(i)^{-\frac{1}{\alpha}}}{\sum_{j=1}^{N} w(j)^{-\frac{1}{1-\alpha}\left(1-\frac{1}{\alpha}\right)}\gamma(j)^{1-\frac{1}{\alpha}}} \\ &= \frac{w(i)^{1/\alpha}\gamma(i)^{-1/\alpha}}{\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}} \end{aligned}$$

which is the desired expression of the (α, w)-proportionally fair vector.

We calculate

$$\begin{aligned}
\sum_{i=1}^{N} w(i)\, f_\alpha(x_{\alpha,w}(i)) &= \frac{1}{1-\alpha}\sum_{i=1}^{N} w(i)\,(x_{\alpha,w}(i))^{1-\alpha} \\
&= \frac{1}{1-\alpha}\sum_{i=1}^{N} w(i)\left(\frac{w(i)^{1/\alpha}\gamma(i)^{-1/\alpha}}{\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}}\right)^{1-\alpha} \\
&= \frac{1}{1-\alpha}\frac{\sum_{i=1}^{N} w(i)\, w(i)^{(1-\alpha)/\alpha}\gamma(i)^{-(1-\alpha)/\alpha}}{\left(\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}\right)^{1-\alpha}} \\
&= \frac{1}{1-\alpha}\left(\sum_{j=1}^{N} w(j)^{1/\alpha}\gamma(j)^{1-1/\alpha}\right)^{\alpha}
\end{aligned}$$

which proves the last result in the Corollary. ■

Dynamics: Markovian model

We consider a situation in which a server provides capacity (bit-rate) to N elastic bit-rate service classes. The i^{th} class has a Poisson arrival rate of λ_i users per second, each user having an exponentially distributed length of mean $1/\mu_i$. Denote ν_i the number of current users of class i and $\nu = (\nu_1, \ldots, \nu_N)$.

We assume that the capacity constraint has the form

$$\sum_{i=1}^{N} \gamma_i \nu_i \psi_i(\nu) = 1 \tag{14.10}$$

where γ_i are some given constants and $\psi_i(\nu)$ is the capacity (in bits/s) allocated to a user of class i.

This system may be described by a Markov process on $\mathbb{N}^N$ with generator

$$\begin{cases} q(\nu, \nu + e_i) &= \lambda_i \\ q(\nu, \nu - e_i) &= \mu_i \psi_i(\nu)\,\nu_i \end{cases} \tag{14.11}$$

This is a multiclass brith-and-death process which is a particular case of spatial Markov queueing (SMQ) processes. In order to retrieve the notations used for such processes, let $\mathbb{D} = \{1, \ldots, N\}$ be the set of locations, define the routing rates by

$$\lambda_{oi} = \lambda_i, \quad \lambda_{io} = \mu_i, \quad \lambda_{ij} = 0, \quad \text{for all } i, j \in \mathbb{D}$$

and define the service rates by

$$r(\nu, T_{oi}\nu) = 1, \quad r(\nu, T_{ij}\nu) = \psi_i(\nu), \quad \text{for all } i, j \in \mathbb{D} \tag{14.12}$$

Let $\rho_i = \frac{\lambda_i}{\mu_i}$ and $\rho = (\rho_i)_{i=1,\ldots,N}$ which is the solution of the traffic equations. If the service rates $\psi_i(\nu)$ are balanced by some function $\Psi(\nu)$, then an invariant measure is given by

$$\Pi(\nu) = \Psi(\nu) e^{-\rho(\mathbb{D})} \prod_{i=1}^{N} \frac{\rho_i^{\nu_i}}{\nu_i!}$$

where $\rho(\mathbb{D}) = \sum_{i=1}^{N} \rho_i$.

Performance

We assume now that the Markov process $\{X(t)\}_{t\in\mathbb{R}_+}$ describing our system is positive recurrent and consider its stationary version $\{X(t)\}_{t\in\mathbb{R}}$ (extended to negative time in the classical way).

Let us denote $T_{\mathbb{D}}$ the sojourn duration of a user (duration between the arrival and departure from the system). By Little's theorem [17],

$$\mathbf{E}[\nu(\mathbb{D})] = \lambda_{\mathbb{D}} \times \mathbf{E}_{\mathbb{D}}[T_{\mathbb{D}}]$$

where $\nu(\mathbb{D}) = \sum_{i=1}^{N} \nu_i$, $\mathbf{E}_{\mathbb{D}}$ is the Palm probability with respect to the arrival process of users and $\lambda_{\mathbb{D}} = \sum_{i\in\mathbb{D}} \lambda_i$ is the total arrival rate. Hence

$$\mathbf{E}_{\mathbb{D}}[T_{\mathbb{D}}] = \frac{\mathbf{E}[\nu(\mathbb{D})]}{\lambda_{\mathbb{D}}} \tag{14.13}$$

which we call the USER DELAY.

Fairness among the users. For a given ν, we may enumerate the users by some index m and write the capacity constraint (14.10) as follows

$$\sum_{m\in\nu} \gamma_m \psi_m(\nu) = 1 \tag{14.14}$$

where

- we use ν as the name of the index set of the users m;
- γ_m is the constant corresponding to the class of user m;
- and $\psi_m(\nu)$ is the capacity (in bits/s) allocated to user m.

For a given ν, the capacity constraint (14.14) has a similar form as (14.5) where $\gamma(i)$ are replaced by γ_m. In this context, the allocation (14.6) takes the following form

$$\psi_m(\nu) = \frac{\gamma_m^{-1/\alpha}}{\sum_{n\in\nu} \gamma_n^{1-1/\alpha}}$$

We see that $\psi_m(\nu)$ depends only on the class, say i, of the user m, so we may denote

$$\psi_i(\nu) = \psi_m(\nu)$$

that is

$$\psi_i(\nu) = \frac{\gamma_i^{-1/\alpha}}{\sum_{j=1}^{N} \nu_j \gamma_j^{1-1/\alpha}} \tag{14.15}$$

Proposition 61 *The service rates (14.12) where $\psi_i(\nu)$ are given by (14.15) are balanced iff $\alpha = 1$ which corresponds to the proportional fair allocation. This allocation writes*

$$\psi_i(\nu) = \frac{1}{\nu(\mathbb{D})\gamma_i} \tag{14.16}$$

and is balanced by

$$\Psi(\nu) = \Psi(0)\,\nu(\mathbb{D})!\prod_{i=1}^{N}\gamma_i^{\nu_i} \tag{14.17}$$

where $\nu(\mathbb{D})$ designates the total number of users in progress.

Proof. Let's decompose the service rates (14.12) as follows

$$r(\nu, T_{ij}\nu) = r^1(\nu, T_{ij}\nu)\, r^2(\nu, T_{ij}\nu)$$

where

$$r^1(\nu, T_{oi}\nu) = 1, \quad r^1(\nu, T_{io}\nu) = \gamma_i^{-1/\alpha} \tag{14.18}$$

and

$$r^2(\nu, T_{oi}\nu) = 1, \quad r^2(\nu, T_{io}\nu) = \left(\sum_{j=1}^{N}\nu_j\gamma_j^{1-1/\alpha}\right)^{-1} \tag{14.19}$$

The service rates $r^1(\nu, T_{ij}\nu)$ are balanced. Then $r^2(\nu, T_{ij}\nu)$ are balanced iff so are $r^2(\nu, T_{ij}\nu)$. Let $\beta = 1 - 1/\alpha$ and

$$\begin{aligned} h_j(\nu) &= r^2(\nu, T_{oj}\nu)/r^2(T_{oj}\nu, \nu) = 1/r^2(T_{oj}\nu, \nu) = 1/r^2(T_{oj}\nu, T_{jo}(T_{oj}\nu)) \\ &= \sum_k \nu_k\gamma_k^\beta + \gamma_j^\beta \end{aligned}$$

r^2 is balanced iff h satisfies

$$h_j(\nu)\, h_i(T_{oj}\nu) = h_i(\nu)\, h_j(T_{oi}\nu)$$

which is equivalent to

$$h_j(\nu) = h_i(\nu)$$

(it is easy to see that $h_i(T_{oj}\nu) = h_j(T_{oi}\nu)$) that is

$$\gamma_j^\beta = \gamma_i^\beta$$

which is equivalent to $\beta = 0$ that is $\alpha = 1$.

Equation (14.8) gives the proportional fair allocation

$$\psi_i(\nu) = \frac{1}{N\nu_i\gamma_i}$$

which is in the form

$$\psi_i(\nu) = \frac{\Psi(\nu - e_i)}{\Psi(\nu)}$$

where $\Psi(\nu)$ is given by (14.17). Therefore, the service rates (14.16) are balanced by (14.17). ∎

Remark 33 *In the paticular case of a single linear constraint, the allocation (14.16) corresponds to the allocation called balanced fairness in [23]. (An allocation is said* BALANCED-FAIRNESS *if it is balanced and belongs to the boundary of the capacity set. Such allocation is unique [98, Lemma 3.1], [69, Proposition 1.4.2].) Hence in the present paticular case, proportional fairness and balanced-fairness coincide.*

Proposition 62 *For the service rates (14.16),*

$$\Psi(0) = e^{\rho(\mathbb{D})}\left(1 - \sum_{i=1}^{N}\rho_i\gamma_i\right) \tag{14.20}$$

and

$$\mathbf{E}[\nu_i] = \rho_i\gamma_i\left(1 - \sum_{j=1}^{N}\rho_j\gamma_j\right)^{-1}$$

The user delay is given by

$$T^{\mathrm{j}} = \mathbf{E}_{\mathbb{D}}[T_{\mathbb{D}}] = \frac{1}{\sum_{i=1}^{N}\lambda_i}\frac{\sum_{i=1}^{N}\rho_i\gamma_i}{1-\sum_{i=1}^{N}\rho_i\gamma_i} \tag{14.21}$$

Proof. We calculate $\Psi(0)$ in order to satisfy the normalization condition

$$\sum_{\nu}\Pi(\nu) = 1$$

We have

$$\begin{aligned}
\sum_{\nu}\Pi(\nu) &= e^{-\rho(\mathbb{D})}\sum_{\nu}\Psi(\nu)\prod_{i=1}^{N}\frac{\rho_i^{\nu_i}}{\nu_i!} = \Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{\nu}\nu(\mathbb{D})!\prod_{i=1}^{N}\frac{(\rho_i\gamma_i)^{\nu_i}}{\nu_i!} \\
&= \Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{n=0}^{\infty}\sum_{\nu:\nu(\mathbb{D})=n} n!\prod_{i=1}^{N}\frac{(\rho_i\gamma_i)^{\nu_i}}{\nu_i!} \\
&= \Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{n=0}^{\infty}\left(\sum_{i=1}^{N}\rho_i\gamma_i\right)^{n} = \Psi(0)\,e^{-\rho(\mathbb{D})}\left(1-\sum_{i=1}^{N}\rho_i\gamma_i\right)^{-1}
\end{aligned}$$

which proves (14.20).

We calculate

$$\begin{aligned}
\mathbf{E}[\nu_i] &= \sum_{\nu} \nu_i \Pi(\nu) = \Psi(0) e^{-\rho(\mathbb{D})} \sum_{\nu} \nu_i \nu(\mathbb{D})! \prod_{j=1}^{N} \frac{(\rho_j \gamma_j)^{\nu_j}}{\nu_j!} \\
&= \rho_i \gamma_i \Psi(0) e^{-\rho(\mathbb{D})} \sum_{n=0}^{\infty} \sum_{\nu:\nu(\mathbb{D})=n} (n+1)! \prod_{j=1}^{N} \frac{(\rho_j \gamma_j)^{\nu_j}}{\nu_j!} \\
&= \rho_i \gamma_i \Psi(0) e^{-\rho(\mathbb{D})} \sum_{n=0}^{\infty} (n+1) \left(\sum_{j=1}^{N} \rho_j \gamma_j \right)^n \\
&= \rho_i \gamma_i \Psi(0) e^{-\rho(\mathbb{D})} \left(1 - \sum_{j=1}^{N} \rho_j \gamma_j \right)^{-2} = \rho_i \gamma_i \left(1 - \sum_{j=1}^{N} \rho_j \gamma_j \right)^{-1}
\end{aligned}$$

■

Fairness among the service classes. For a given ν, the capacity constraint (14.10) has a similar form as (14.5) where $\gamma(i)$ are replaced by $\gamma_i \nu_i$. In this context, the allocation (14.6) takes the following form

$$\psi_i(\nu) = \frac{(\nu_i \gamma_i)^{-1/\alpha} 1\{\nu_i \neq 0\}}{\sum_{j=1}^{N} (\nu_i \gamma_i)^{1-1/\alpha}} \tag{14.22}$$

(Note that Proposition 60 assumes that the $\gamma(i)$ are positive, whereas now $\gamma(i) = \gamma_i \nu_i = 0$ when $\nu_i = 0$. We consider the results of Proposition 60 with some modifications: the equality in the constraint (14.5) is replaced by inequality $\leq$; and we multiply the expressions of the fair allocations by the indicator function $1\{\gamma(i) \neq 0\}$. This justifies the term $1\{\nu_i \neq 0\}$ in the above display.)

Proposition 63 *The service rates (14.12) where $\psi_i(\nu)$ are given by (14.22) are balanced iff $\alpha = 1$ which corresponds to the proportional fair allocation. This allocation writes*

$$\psi_i(\nu) = \frac{1\{\nu_i \neq 0\}}{N \nu_i \gamma_i} \tag{14.23}$$

and is balanced by

$$\Psi(\nu) = \Psi(0) N^{\nu(\mathbb{D})} \prod_{i=1}^{N} \nu_i! \gamma_i^{\nu_i} \tag{14.24}$$

where $\nu(\mathbb{D})$ designates the total number of users in progress.

Proof. Let's decompose the service rates (14.12) as follows

$$r(\nu, T_{ij}\nu) = r^1(\nu, T_{ij}\nu) r^2(\nu, T_{ij}\nu)$$

where

$$r^1(\nu, T_{oi}\nu) = 1, \quad r^1(\nu, T_{io}\nu) = (\nu_i\gamma_i)^{-1/\alpha} 1\{\nu_i \neq 0\} \tag{14.25}$$

and

$$r^2(\nu, T_{oi}\nu) = 1, \quad r^2(\nu, T_{io}\nu) = \left(\sum_{j=1}^{N} (\nu_j\gamma_j)^{1-1/\alpha}\right)^{-1} \tag{14.26}$$

The service rates $r^1(\nu, T_{ij}\nu)$ are balanced. Then $r^2(\nu, T_{ij}\nu)$ are balanced iff so are $r^2(\nu, T_{ij}\nu)$. Let $\beta = 1 - 1/\alpha$ and

$$\begin{aligned} h_j(\nu) &= r^2(\nu, T_{oj}\nu) / r^2(T_{oj}\nu, \nu) = 1/r^2(T_{oj}\nu, \nu) = 1/r^2(T_{oj}\nu, T_{jo}(T_{oj}\nu)) \\ &= \sum_{k \neq j} (\nu_k\gamma_k)^\beta + ((\nu_j + 1)\gamma_j)^\beta \end{aligned}$$

r^2 is balanced iff h satisfies

$$h_j(\nu) h_i(T_{oj}\nu) = h_i(\nu) h_j(T_{oi}\nu)$$

which is equivalent to

$$h_j(\nu) = h_i(\nu)$$

(it is easy to see that $h_i(T_{oj}\nu) = h_j(T_{oi}\nu)$) that is

$$((\nu_j + 1)\gamma_j)^\beta = ((\nu_i + 1)\gamma_i)^\beta$$

which is equivalent to $\beta = 0$ that is $\alpha = 1$.

Equation (14.8) gives the proportional fair allocation

$$\psi_i(\nu) = \frac{1\{\nu_i \neq 0\}}{N\nu_i\gamma_i}$$

which is in the form

$$\psi_i(\nu) = \frac{\Psi(\nu - e_i)}{\Psi(\nu)}$$

where $\Psi(\nu)$ is given by (14.24). Therefore, the service rates (14.23) are balanced by (14.24). ■

Proposition 64 *For the service rates (14.16),*

$$\Psi(0) = e^{\rho(\mathbb{D})} \prod_{i=1}^{N} (1 - N\rho_i\gamma_i) \tag{14.27}$$

and

$$\mathbf{E}[\nu_i] = N\rho_i\gamma_i (1 - N\rho_i\gamma_i)^{-1}$$

The user delay is given by

$$T^{\mathrm{c}} = \mathbf{E}_{\mathbb{D}}[T_{\mathbb{D}}] = \frac{1}{\sum_{i=1}^{N} \lambda_i} \sum_{i=1}^{N} \frac{N\rho_i\gamma_i}{1 - N\rho_i\gamma_i} \tag{14.28}$$

Proof. We calculate $\Psi(0)$ in order to satisfy he normalization condition

$$\sum_{\nu}\Pi(\nu)=1$$

We have

$$\begin{aligned}\sum_{\nu}\Pi(\nu) &= e^{-\rho(\mathbb{D})}\sum_{\nu}\Psi(\nu)\prod_{i=1}^{N}\frac{\rho_i^{\nu_i}}{\nu_i!}=\Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{\nu}N^{\nu(\mathbb{D})}\prod_{i=1}^{N}(\rho_i\gamma_i)^{\nu_i}\\ &=\Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{\nu}\prod_{i=1}^{N}(N\rho_i\gamma_i)^{\nu_i}\\ &=\Psi(0)\,e^{-\rho(\mathbb{D})}\prod_{i=1}^{N}\left(\sum_{n=0}^{\infty}(N\rho_i\gamma_i)^n\right)=\Psi(0)\,e^{-\rho(\mathbb{D})}\prod_{i=1}^{N}(1-N\rho_i\gamma_i)^{-1}\end{aligned}$$

which proves (14.20).

We calculate

$$\begin{aligned}\mathbf{E}[\nu_i] &= \sum_{\nu}\nu_i\Pi(\nu)=\Psi(0)\,e^{-\rho(\mathbb{D})}\sum_{\nu}\nu_i\prod_{j=1}^{N}(N\rho_j\gamma_j)^{\nu_j}\\ &=\Psi(0)\,e^{-\rho(\mathbb{D})}\left(\sum_{n=0}^{\infty}n\,(N\rho_i\gamma_i)^n\right)\prod_{j\neq i}\left(\sum_{n=0}^{\infty}(N\rho_j\gamma_j)^n\right)\\ &=\Psi(0)\,e^{-\rho(\mathbb{D})}N\rho_i\gamma_i\,(1-N\rho_i\gamma_i)^{-2}\prod_{j\neq i}(1-N\rho_j\gamma_j)^{-1}\end{aligned}$$

■

Static capacity allocation. [76, § 4.4] studied allocations assigning a fixed fraction of the available capacity to each service class (i.e. $\nu_i\psi_i(\nu)$ is independent of ν). The author obtained the optimal allocation which minimizes the user delay. The minimal user delay is given by [76, Equation (4.22)]

$$T^{\mathrm{k}}=\frac{1}{\sum_{i=1}^{N}\lambda_i}\frac{\left(\sum_{i=1}^{N}\sqrt{\rho_i\gamma_i}\right)^2}{1-\sum_{i=1}^{N}\rho_i\gamma_i} \tag{14.29}$$

Comparison. The following lemma is useful for comparing the user delays of the above allocations.

Lemma 33 *For each* $\alpha_1,\ldots,\alpha_N\in[0,1/N)$,

$$\sum_{i=1}^{N}\frac{\alpha_i}{1-N\alpha_i}\geq\frac{\sum_{i=1}^{N}\alpha_i}{1-\sum_{i=1}^{N}\alpha_i}$$

Proof. We have

$$\sum_{i=1}^{N} \frac{\alpha_i}{1-N\alpha_i} = \sum_{i=1}^{N} \frac{\alpha_i - \frac{1}{N} + \frac{1}{N}}{1-N\alpha_i} = -1 + \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-N\alpha_i}$$

The second term of the right-hand side is the arithmetic mean of the $\frac{1}{1-N\alpha_i}$, which is larger than their harmonic mean, that is

$$\frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-N\alpha_i} \geq \frac{N}{\sum_{i=1}^{N}(1-N\alpha_i)} = \frac{1}{1-\sum_{i=1}^{N}\alpha_i}$$

Then

$$\sum_{i=1}^{N} \frac{\alpha_i}{1-N\alpha_i} \geq -1 + \frac{1}{1-\sum_{i=1}^{N}\alpha_i} = \frac{\sum_{i=1}^{N}\alpha_i}{1-\sum_{i=1}^{N}\alpha_i}$$

■

From Lemma 33 we deduce that

$$T^{\mathrm{j}} = \frac{1}{\sum_{i=1}^{N}\lambda_i} \frac{\sum_{i=1}^{N}\rho_i\gamma_i}{1-\sum_{i=1}^{N}\rho_i\gamma_i} \leq \frac{1}{\sum_{i=1}^{N}\lambda_i} \sum_{i=1}^{N} \frac{\rho_i\gamma_i}{1-N\rho_i\gamma_i} = \frac{1}{N}T^{\mathrm{c}}$$

Since the allocation (14.23) satisfies the property $\nu_i\psi_i(\nu)$ is independent of ν, we deduce that

$$T^{\mathrm{k}} \leq T^{\mathrm{c}}$$

Moreover we see from the Equations (14.21) and (14.29) that

$$T^{\mathrm{j}} \leq T^{\mathrm{k}}$$

Suppose moreover that $\lambda_i = \lambda, \gamma_i = \gamma, \mu_i = \mu$, then

$$T^{\mathrm{k}} = \frac{1}{N\lambda}\frac{\left(N\sqrt{\rho\gamma}\right)^2}{1-N\rho\gamma} = \frac{1}{N\lambda}\frac{N^2\rho\gamma}{1-N\rho\gamma} = NT^{\mathrm{j}}$$

Then we deduce that the allocation assuring proportional fairness among the users gives a delay smaller than the delays of the other two allocations (proportional fairness among the classes, and static allocation). The reduction factor is of order $\frac{1}{N}$.

An interesting optimization problem arises: among the allocations defined by the functions $\psi_i(\nu)$ satisfying (14.10), which one minimizes the delay. Equation (14.13), shows that the optimal allocation should minimize $\mathbf{E}[\nu(\mathbb{D})]$. If we consider only the balanced allocations, then we deduce from [98, Lemma 3.1], [69, Proposition 1.4.2] that the optimal allocation is the so-called balanced fairness (which coincide in our case with the allocation (14.16) which realizes the proportional fairness amoung the users).

Bibliography

[1] H. Alavi and R.W. Nettleton. Downstream power control for a spread spectrum cellular mobile radio system. *Proc. Globecom 82, Miami*, pages 84–88, 1982.

[2] E. Altman. Capacity of Multi-service CDMA Cellular Networks with Best-Effort Applications. In *Proceedings of MOBICOM 02*, September 2002.

[3] E. Altman. Stratégies de contrôle de débit dans l'umts, 2002.

[4] E. Altman, D. Artiges, and K. Traore. On the Integration of Best-Effort and Guaranteed Performance Services. *European Transactions on Telecommunications. Special Issue on Architectures, Protocols and Quality of Service for the Internet of the Future*, 2, February-March 1999.

[5] E. Altman, K. Avrachenkov, and R. Nez-Queija. Perturbation analysis for denumerable markov chains with application to queueing models. *Advances in Applied Probability*, 36, 2004.

[6] S. Anand and A. Chockalingam. Performance analysis of voice/data cellular cdma with sir based admission control. *IEEE JSAC*, 21(10), December 2003.

[7] M. Andersin, Z. Rosberg, and J. Zander. Soft and safe admission control in cellular networks. *IEEE Transaction on Networking*, 5(2):414–418, April 1997.

[8] W.J. Anderson. *Continuous-Time Markov Chains.* Springer-Verlag, 1991.

[9] M. Armony and N. Bambos. Queueing dynamics and maximal throughput scheduling in switched processing systems. *Queueing Systems*, 44, 2003.

[10] S. Asmussen. *Applied probability and queues.* Springer, New York, 1987.

[11] F. Baccelli, B. Błaszczyszyn, and M.K. Karray. Contrôle de charge dans le sens montant pour les systèmes de communication sans fil avec contrôle de puissance, April 2003. N^o de dépôt : FR03 05355. Déposant INRIA/FT - Cas 43FT.

[12] F. Baccelli, B. Błaszczyszyn, and M.K. Karray. Dispositif et procédé de contrôle de charge avec contrôle de puissance, April 2003. N^o de dépôt : FR03 05354. Déposant INRIA/FT - Cas 42FT.

[13] F. Baccelli, B. Błaszczyszyn, and M.K. Karray. Up and Downlink Admission/Congestion Control and Maximal Load in Large Homogeneous CDMA Networks. *MONET*, 9(6), December 2004.

[14] F. Baccelli, B. Błaszczyszyn, and M.K. Karray. Blocking Rates in Large CDMA Networks via Spatial Erlang Formula. In *IEEE INFOCOM 05*, 2005.

[15] F. Baccelli, B. Błaszczyszyn, and M.K. Karray. A note on regularity and ergodicity of spatial migration processes. *Submitted to Electronic Communications in Probability*, 2007.

[16] F. Baccelli, B. Błaszczyszyn, and F. Tournois. Downlink admission/congestion control and maximal load in CDMA networks. *IEEE INFOCOM 03*, March 2003.

[17] F. Baccelli and P. Brémaud. *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences.* Springer, 2003.

[18] F. Baccelli and A.A. Makowski. Stability and bounds for single server queues in random environment. Rapport de recherche 536, INRIA, June 1986.

[19] N. Bambos, S.C. Chen, and G.J. Pottie. Channel access algorithms with active link protection for wireless communication networks with power control. *IEEE/ACM Transactions in Networking*, 8(5):583–596, October 2000.

[20] B. Błaszczyszyn and F. Baccelli. Three-sector-antenna path-loss and cdma downlink capacity, 2002. Private communication.

[21] B. Błaszczyszyn and M.K. Karray. Performance Evaluation of Scalable Congestion Control Schemes for Elastic Traffic in Cellular Networks with Power Control. In *IEEE INFOCOM 07*, 2007.

[22] T. Bonald, S. Borst, and A. Proutière. How mobility impacts the flow-level performance of wireless data systems. In *Proceedings of INFOCOM 04*, 2004.

[23] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo. A queueing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems*, 53:65–84, 2006.

[24] T. Bonald and A. Proutière. Wireless downlink data channels: user performance and cell dimensioning. In *Proceedings of MOBICOM 03*. ACM Press, 2003.

[25] T. Bonald and A. Proutière. A queueing analysis of data networks. In *Queueing Networks: A Fundamental Approach (edited by R. Boucherie, N. Van Dijk)*, 2006.

[26] T. Bonald, A. Proutière, J. Roberts, and J. Virtamo. Computational aspects of balanced fairness. In *Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pages 801 – 810, Berlin, Germany, September 2003.

[27] C. Bordenave. Stability of spatial queueing systems. Rapport de recherche 5305, INRIA, September 2004.

[28] C. Bordenave. Stability properties of data flows on a CDMA network in macrodiversity. Rapport de Recherche 5257, INRIA, February 2004.

[29] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proceedings of INFOCOM 03*, March 2003.

[30] S.K. Bose, D. Manjunath, H. Tahilramani, D.M. Bhaskar, and M.N. Umesh. Summary of algorithms implemented in qnat.

[31] P. Brémaud. *Signal et communication Modulation, codage et théorie de l'information*. Ellipses, 1995.

[32] P. Brémaud. *Markov chains. Gibbs Fields, Monte Carlo Simulation, and queues*. Springer, 1999.

[33] M.F. Chen. Coupling for jump processes. *Acta Mathematica Sinica*, 2(2):123–136, 1986.

[34] M.F. Chen. *From Markov chains to non-equilibrum particle systems*. World Scientific, 2 edition, 2004.

[35] E. Chlebus and W. Lidwin. Is handoff traffic really poissonian? *Proceedings ICUP*, pages 348–353, 1995.

[36] E.G. Coffman Jr., R.R. Muntz, and H. Trotter. Waiting time distribution for processor-sharing systems. *J. ACM*, 17:123–130, 1970.

[37] J.W. Cohen. *On Regenerative Processes in Queueing Theory*, volume 121 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, 1976.

[38] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.

[39] P.J. Courtois. *Decomposability: Queueing and Computer System Applications*. ACM monograph series. Academic Press, New York, 1977.

[40] D.J. Daley and D. VereJones. *An introduction to the theory of point processes*. Springer, New York, 1988.

[41] D.A. Dawson. Measure-valued Markov processes. In Springer-Verlag, editor, *Lecture notes in mathematics. Ecole d'Eté de Probabilités de Saint-Flour, XXI-1991*, volume 1541, pages 1–260, Berlin, 1991.

[42] F. Delcoigne, A. Proutière, and G. Régnié. Modelling integration of streaming and data traffic. In *ITC specialist seminar on IP traffic*, July 2002.

[43] F. Delebecque. A reduction process for perturbed markov chains. *SIAM J. App. Math.*, 43:325–350, 1983.

[44] A.K. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektrotkeknikeren*, 13, 1917.

[45] COperation européenne dans le domaine de la recherche Scientifique et Technique. *COST 231 Final Report Digital mobile radio towards future generation systems*. European Commision, Brussels, Belgium, 1999.

[46] J.S. Evans and D. Everitt. Effective bandwidth-based admission control for multiservice cdma cellular networks. *IEEE Transactions on Vehicular Technology*, 48, January 1999.

[47] J.S. Evans and D. Everitt. On the teletraffic capacity of cdma cellular networks. *IEEE Transactions on Vehicular Technology*, 48, January 1999.

[48] W. Feller. *An introduction to probability theory and its applications II.* John Wiley and sons, New York, 1966.

[49] G.J. Foschini and Z Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Trans. Veh. Technol.*, 40:641–646, 1991.

[50] A.K. Maulloo F.P. Kelly and D.K.H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[51] R.G. Gallager. 6.450 course notes: Principles of Digital Communication, 2005.

[52] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, and C.E. Wheatley. On the capacity of a cellular cdma system. *IEEE Trans. Vehic. Technol.*, pages 303–312, May 1991.

[53] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[54] S. Grandhi, J. Zander, and R. Yates. Constrained power control. TRITA-IT Report 94-06, Dept. of Teleinformation, Radio Commun. Systems Lab, Royal Institute of Technology, Sweden, June 1994.

[55] F. Gunnarsson. *Power Control in Cellular Radio Systems: Analysis, design and estimation*. PhD thesis, Linkpings universitet, 1998.

[56] S.V. Hanly. Capacity in a two cell spread spectrum network. *Thirtieth Annual Allerton Conf. Commun., Control and Computing*, IL:426–435, 1992.

[57] S.V. Hanly. *Information Capacity of Radio Networks*. PhD thesis, University of Cambridge, August 1995.

[58] S.V. Hanly. Capacity and power control in spread spectrum macrodiversity radio networks. *IEEE Transactions on communications*, 44(2), February 1996.

[59] S.V. Hanly. Congestion measures in ds-cdma networks. *IEEE Trans. on Comm.*, 47, March 1999.

[60] G. Hardy, JE. Littlewood, and G. Polya. *Inequalities*. Cambridge university press, Cambridge, 1934.

[61] G. Haring, R. Marie, R. Puigjaner, and K. Trivedi. Loss formulas and their applications to optimization for cellular networks. *IEEE trans. on Vehicular Technology*, 50, 2001.

[62] N. Hegde and E. Altman. Capacity of multiservice WCDMA Networks with variable GoS . In *Proceedings of IEEE WCNC*, March 2003.

[63] K. Hiltunen and R. De Bernardi. Wcdma downlink capacity estimation. *VTC2000*, pages 992–996, 2000.

[64] H. Holma and A. Toskala. *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. John Wiley and Sons, 2 edition, 2002.

[65] D. Hong and S.S. Rappaport. Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on vehicular technology*, 35, August 1986.

[66] X. Huang and R. F. Serfozo. Reversible markov processes on general spaces and spatial birth-death and queueing processes. First draft, School of Industrial and Systems Engineering, Georgia Institute of Technology, December 2002.

[67] X. Huang and R.F. Serfozo. Spatial queueing processes. *Math. Oper. Res.*, 24:865–886, 1999.

[68] D.L. Iglehart. Multivariate competition processes. *Ann. Math. Statist.*, 35:350–361, 1964.

[69] M. Jonckheere. *Partage de charge et allocation de capacite insensibles dans les reseaux de telecommunication (in french).* PhD thesis, Ecole Polytechnique, 2005.

[70] E. Jugl and H. Boche. Dwell time models for wireless communication systems. *VTC*, 1999.

[71] O. Kallenberg. *Foundations of modern probability.* Springer, 1997.

[72] S. Karlin and J.L. McGregor. The classification of birth-and-death processes. *Trans. Amer. Math. Soc.*, 86:366–400, 1957.

[73] J.S. Kaufman. Blocking in a shared resource environment. *IEEE Transactions on Communications*, 29(10):1474–1481, 1981.

[74] F.P. Kelly. Effective bandwidths at multi-class queue. *Queuing Systems*, 9:5–16, 1991.

[75] F.P. Kelly. Loss networks. *The Annals of Applied Probabilities*, 1:319–378, 1991.

[76] L. Kleinrock. *Communication Nets.* McGraw-Hill, 1964.

[77] J. Knutsson, P. Butovitsch, M. Persson, and R.D. Yates. Downlink admission control strategies for cdma systems in a manhattan environment. *Vehicular Technology Conference*, 2:1453 –1457, 1998.

[78] I. Koukoutsidis, E. Altman, and J.M. Kelif. Fair rate sharing models in a CDMA link with multiple classes of elastic traffic. Rapport de Recherche 5596, INRIA, June 2005.

[79] H.J. Kushner and P.A. Whiting. Convergence of Proportional-Fair Sharing Algorithms Under General Conditions. *IEEE Transactions On Wireless Communications*, 3(4), July 2004.

[80] M.S. Kwok and H.S. Wang. Adjacent-cell interference analysis of reverse link in cdma cellular radio system. *IEEE PIMRC 95*, pages 446–450, 1995.

[81] J. Laiho, A. Wacker, and T. Novosad. *Radio Network Planning and Optimisation for UMTS.* John Wiley and sons, 2002.

[82] J.S. Lee and L.E. Miller. *CDMA Systems Engineering Handbook.* Artech House, Boston, 1998.

[83] R. A. Leese. A unified approach to the assignment of radio channels on a regular hexagonal grid. *IEEE Transactions on vehicular technology*, 46(4), November 1997.

[84] Z. Liu and M. El Zarki. Sir-based call admission control for ds-cdma cellular systems. *Selected Areas in Communications, IEEE Journal on*, 12:638 –644, May 1994.

[85] C. Meyer. *Matrix Analysis. See Errata at www.matrixanalysis.com.* SIAM, 2000.

[86] R. Michel. On berry-esseen results for the compound poisson distribution. *Insurance: Mathematics and Economics*, 13:35–37, 1993.

[87] C. Mihailescu, X. Lagrange, and Ph. Godlewski. Radio resource management for packet transmission in umts wcdma systems. *VTC2000*, 1999.

[88] C. Mihailescu, X. Lagrange, and Ph. Godlewski. Soft handover analysis in downlink umts wcdma system. *IEEE International Workshop on Mobile Multimedia Communications*, November 1999.

[89] K. Mitchell and K. Sohraby. An analysis of the effects of mobility on bandwidth allocation strategies in multi-class cellular wireless networks. *IEEE INFOCOM*, 2001.

[90] D. Mitra. An asynchronous distributed algorithm for power control in cellular radio systems. *Fourth Winlab worshop on third generartion wireless information networks*, pages 249–257, 1993.

[91] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, 2000.

[92] M. Mouly and M.B. Pautet. *The GSM System for Mobile Communications.* CellAndSys, Palaiseau, France, 1992.

[93] M.F. Neuts. *Matrix-geometric solutions in stochastic models: an algorithmic approach.* Johns Hopkins series in the mathematical sciences. The Johns Hopkins University press, 1981.

[94] J. Neveu. Construction de files d'attente stationnaires. In *Lecture notes in control and information Sciences*, volume 60, pages 31–41, 1983.

[95] S.R. Nunez-Queija. *Processor-Sharing Models for Integrated-Services Networks.* PhD thesis, Technische Universiteit Eindhoven, 20 January 2000.

[96] J. Pérez-Romero, O. Sallent, and R. Agust. Impact of user location in w-cdma downlink resource allocation. *IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA 2002). Prague, Czech Republic*, 2 September 2002.

[97] C. Preston. Spatial birth-and-death processes. *Bull. Int. Statist. Inst.*, 46(2), 1977.

[98] A. Proutière. *Insensibilité et bornes stochastiques dans les réseaux de files d'attente. Application à la modélisation des réseaux de télécommunications au niveau flot (in french).* PhD thesis, Ecole Polytechnique, 2003.

[99] B. Radunovic and J.Y. Le Boudec. A unified framework for max-min and min-max, 2002.

[100] R. Ramjee, R. Nagarajan, and D. Towsley. On optimal call admission control in cellular networks. *Wireless Networks*, 3:29–41, 1997.

[101] G.E.H. Reuter. Competition processes. In *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, volume 2, pages 421–430, 1961.

[102] G.E.H. Reuter and W. Ledermann. On the differential equations fot the transition probabilities of Markov processes with enumerably many states. *Proc. Camb. Phil. Soc.*, 49:247–262, 1953.

[103] Ph. Robert. *Stochastic networks and queues.* Springer, 2003.

[104] J.W. Roberts. A service system with heterogeneous user requirements. In *Performance of Data Communications Systems and their Applications (edited by G. Pujolle)*, 1981.

[105] P.J. Schweitzer. Perturbation theory and finite markov chains. *J. Appl. Prob.*, 5:401–413, 1968.

[106] R. Serfozo. *Introduction to stochastic networks.* Springer, New York, 1999.

[107] R. Serfozo. Reversible markov processes on general spaces: Spatial birth-death and queueing processes. *Infoma. process.*, 2(2):259–261, 2002.

[108] R. Serfozo. Reversible markov processes on general spaces and spatial migration processes. Institut Mittag-Leffler, The Royal Sweddish Academy of Science. Report No. 05, 2004/2005, 13 December 2004.

[109] R. Serfozo. Reversible markov processes on general spaces and spatial migration processes. *Advances in Applied Probability*, 37, 2005.

[110] S. Shin, K. Lee, and K. Kim. Performance of the packet data transmission using the other-cell-interference factor in ds-cdma downlink. In *IEEE Wireless Communications and Networking Conference*, 2002.

[111] M. Sidi and D. Starobinski. New call blocking versus handoff blocking in cellular networks. *Wireless Networks*, 3, 1997.

[112] K. Sipila, Z.C. Honkasalo, J.L. Steffens, and A. Wacker. Estimation of capacity and required transmission power of wcdma downlink. *VTC2000*, pages 1002–1005, 2000.

[113] D. Stoyan, W. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications.* John Wiley and Sons, 1987.

[114] Third-Generation-Partnership-Project. *HSDPA Revised CQI Proposal.* 3GPP, Motorola, Ericsson, 9-12 April 2002. R1-02-0675.

[115] R. Thomas, H. Gilbert, and G. mazziotto. Influence of the moving of the mobile stations on the performance of a radio mobile cellular network. *Third Nordic Seminar on Digital Land Mobile Radio Communication*, 12-15 September 1988.

[116] F. Tournois. *Modélisation et simulation des réseaux CDMA par la géométrie aléatoire. Thesis.* PhD thesis, Ecoles des Mines de Paris, 21 October 2002.

[117] M.A. Tran. Fonction d'atténuation des antennes tri-sectoriels et capacité de la voie descendante des grands réseaux hexagonaux cdma (in frensh), 2003. Rapport de stage. INRIA.

[118] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[119] S. Vajda. *Theory of linear and non-linear programming.* Longman, Willian Clowers and Sons, Great Britain, 1974.

[120] A.J. Viterbi. *CDMA: Principles of Spread Spectrum Communications.* Addison-Wesley, 1995.

[121] A.J. Viterbi and A.M. Viterbi. Erlang capacity of a power controlled CDMA system. *IEEE Journal on Selected Areas in Communications*, 11(6):892–899, August 1993.

[122] R. Yates. A framework for uplink power control in cellular radio systems. *IEEE J. Sel. Areas Commun.*, 13(7):1341–1348, September 1995.

[123] G. Yin and H.Q. Zhang. Two-time-scale markov chains and applications to quasi-birth-death queues. *SIAM journal on applied mathematics*, 65(2), 2005.

[124] G. Yin and Q. Zhang. *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach.* Springer-Verlag, New York, 1998.

[125] J. Zander. Distributed co-channel interference control in cellular radio systems. *IEEE Trans. Veh. Technol.*, 41:305–311, 1992.

[126] J. Zander. Performance of optimum transmitter power control in cellular radio systems. *IEEE Trans. Vehic. Technol.*, 41(1):57–62, February 1992.

[127] J. Zander and M. Frodigh. Comment on performance of optimum transmitter power control in cellular radio systems. *IEEE Trans. Vehic. Technol.*, 43:636–636, August 1994.

[128] C. Zhu and M.S. Corson. A distributed channel probing for wireless networks. *Proc. of IEEE Infocom 2001*, pages 403–411, 2001.

Index