
Iana Siomina, Anders Furuskar, and Gabor Fodor

Ericsson Research, Sweden, E-mail: {iana.siomina, anders.furuskar, gabor.fodor}@ericsson.com

Abstract—This paper presents a mathematical modeling framework for studying the capacity of multi-cell orthogonal frequency division multiplexing networks in single- and multi-service scenarios with various quality of service (QoS) constraints. The framework is built around a feasible network load concept which relates cell resource utilization and interference-dependent resource demand generated by traffic. The feasible load problem is formulated for a general irregular network with non-uniform traffic distribution. Unlike in earlier works, the proposed model explicitly takes into account interference which is load- and service-dependent and models the service constraints on a time scale longer than the typical scheduling interval. A computationally efficient stochastic model is proposed for a regular network deployment with a user distribution pattern repeated over cells and then enhanced with QoS constraints. As an example application, the capacity region problem is studied and numerical results are presented for a realistic network setup. The framework can further be used, for example, for studying radio resource management algorithms or optimizing QoS parameters with respect to an operator QoS policy.

Index terms – OFDM, feasible load, capacity, mixed traffic.

I. INTRODUCTION

The problem of providing quality of service (QoS) in multi-cell networks in general and in networks using orthogonal frequency division multiplexing (OFDM) in particular has been addressed by research and standardization for a long time. There seems to be a common understanding that admission control (AC) and packet scheduling play a key role in providing QoS in a resource efficient manner [1], [5], [10]. Both radio resource management (RRM) algorithms depend to a large extent on traffic intensity and interference which define the load on the radio interface and have a great impact on the capacity of individual cells and the entire network. Load distribution has therefore been explicitly taken into account in many proposed algorithm realizations (see, e.g. [2], [8] for AC). In fact, load balancing has been found to significantly improve the overall capacity and user satisfaction in multi-cell wideband code division multiple access (WCDMA) systems [7].

The design of multi-cell AC is challenging, not only because of the diverse QoS requirements, but also because the service requirements are typically formulated at a much longer time scale than that of the scheduling of practical systems [6]. For the same reasons, network capacity modeling is also a non-trivial task. There is a need for a model that captures the resource demand during a time window that is (much) greater than a scheduling time interval. Such a model should help to establish the capacity region of multi-cell systems supporting multiple service classes without worrying about the short time random variations of the positions, scheduling decisions or channel variations of the served mobile stations [3]. The model could also be used for developing multi-cell RRM algorithms and designing QoS profiles for a given operator policy.

Although these aspects are not only well understood but also practiced in operating WCDMA networks [9], to our best knowledge there are no multi-cell RRM (AC in particular) or multi-service QoS-aware capacity models for OFDM in the literature that address these requirements. In this paper, we develop a multi-cell OFDM network model in three steps. First, we consider the problem of allocating resources to a user in a given system state and finding feasible cell load levels. Secondly, we formulate the problem of maximizing the number of admissible users and thereby of finding the capacity region of the system. In the third step we propose a stochastic model to capture the important case of users dynamically entering and departing the system. Within this framework, we also propose two solution approaches for solving the feasible load problem and the capacity region problem.

The paper is organized as follows. In the next section we introduce the system model. The subsequent sections describe the feasible load concept and formulate the capacity region problem, respectively. Finally, we present numerical results illustrating the practical use of the modeling framework in a realistic scenario.

II. SYSTEM MODEL

We consider an OFDM network with a set of cells $\mathcal{I}$ and a set of services $\mathcal{S}$. The network is in general irregular and the user distribution is non-uniform, unless otherwise stated. The load, or cell resource utilization, in cell $i$ is denoted by $\rho_i$, where $\rho_i \in (0, \rho_i^{\text{max}}]$, $i \in \mathcal{I}$, and $\rho_i^{\text{max}} \leq 1$ is the maximum allowed load. The latter restriction could be, for example, a radio network design parameter decided by an operator. The entire network load is given by a vector $\rho = (\rho_1, \ldots, \rho_{\mathcal{I}})$. For each user location $j \in \mathcal{J}$, $g_{ij}$ is the total power gain between $j$ in cell $i$. Then, for user $j$ served by cell $i$ the average SINR is defined as

$$\gamma_{ij}(\rho) = \frac{P_i \cdot g_{ij}}{\sum_{b \in \mathcal{I}, b \neq i} P_b \cdot g_{bj} \cdot \rho_b + \sigma^2},$$

where $P_i$ is the power spectral density per resource block (RB) and $\sigma^2$ is the non-zero noise power. Eq. (1) describes in general a power-controlled system. By one possible interpretation (more typical for wideband radio technologies), the total transmit power is proportional to the cell load. In OFDM networks, the co-channel interference is not necessarily received in all subcarriers. In this case, the cell load can be viewed as a transmission probability on the measured subcarrier, and the total co-channel received power can be thought of as statistical interference.

The effective bitrate $b_{ij}(\rho)$ for user $j$ per RB, which is the minimum time-frequency scheduling unit, is a function of the average SINR which is assumed to be known, e.g.

$$b_{ij}(\rho) = a_1 \cdot \log_2 (1 + a_2 \cdot \gamma_{ij}(\rho)),$$

where $a_1$ and $a_2$ are the given model parameters. Given a set of services $\mathcal{S}$ and an observation period of $N_T$ time transmission intervals (TTIs), the RB demand for transmitting in cell $i$ the data of service $s \in \mathcal{S}$ in the amount of $V_s$ to a user $j$ is

$$n_{j,s}(\rho) = \frac{V_s}{b_{ij}(\rho)}.$$
Let $N_{RB}$ be the maximum number of available RBs in frequency space, and $N$ be the maximum number of RBs in the studied time-frequency space, i.e.

$$N = N_T \cdot N_{RB}.$$  

(4)

The cell load can be interpreted as an expected fraction of $N$ RBs that are utilized in the cell. In our system model, the load vector is unknown and requires that the entire system is in balance, which is the basic idea in the presented modeling framework. The presented models have been derived for downlink (DL) communication, but could also be adapted for uplink (UL).

III. THE FEASIBLE LOAD CONCEPT

A. A Basic Feasible Load Problem

With the system model presented in Section II, the basic optimization problem would be to find a feasible network load ensuring full coverage for a given user snapshot, known traffic demand and users’ serving cells, i.e. a feasibility problem with respect to $\rho$ needs to be solved. The problem is further referred to as the feasible load problem and can be stated as follows: Find a network load vector $\rho$ such that the resource utilization is in balance with the resource demand in all cells and the cell loads are within given ranges. Below is a mathematical formulation of the problem given as a system of non-linear equations,

\[
\begin{align*}
\text{[FLP]} & \quad \sum_{s \in S} \sum_{j \in J_s} n_{j,s}(\rho) = N \cdot \rho_i, \quad i \in I, \\
& \quad \rho_i \in (0, \rho_{i,\max}], \quad i \in I.
\end{align*}
\]

(5a)

(5b)

In FLP, $J_s$ is the set of users of service $s$ served by cell $i$. The left-hand side (LHS) of (5a) is the total resource demand in cell $i$ based on the other cells load, and the RHS is the resource utilization in cell $i$.

Property FLP-1: For a given user snapshot and user assignment, there always exists at most one vector $\bar{\rho}$ satisfying (5).

Property FLP-1 can be proved based on the definition of $n_{j,i}(\rho)$ and the observation from (5a) that for any cell $i$, $\rho_i$ is a non-decreasing function which is concave in any variable $\rho_i$, $l \neq i$. Fig. 1 sketches a feasible and infeasible solutions to system (5) in a two-cell network scenario. With the feasible solution, the two solid curves intersect within a feasible region constrained by (5b), and the system has thus a unique solution $\bar{\rho}$.

In FLP, an infeasible solution may occur due to tight constraints (5), which may result from minimum service quality requirements and user subscription profile and may limit the maximum cell load, or due to a high resource demand in a cell because of high interference from neighboring cells, low sensitivity, too many users in the cell, and/or high traffic volumes.

Property FLP-2: For any constant $C \in (0, \lbrack 2 \rbrack]$, $\sum_{i \in I} \rho_i = C$ defines a set of feasible solutions with the same total RB demand.

Property FLP-2 follows from the RHS of system (5). The property can be exploited when the user assignment is to be decided (an extended FLP). Observe that the optimal user assignment from the resource consumption point of view is the one which minimizes $C$, or the sum of the individual cell loads.

In a general case, FLP can be solved, for example, by the Newton-Raphson method. The problem significantly simplifies under the assumption of equal load and repeating traffic demand pattern in all cells when the set of equations (5a) reduces to a single equation to be solved with respect to a single variable $\rho$ (assuming $\rho_i = \rho$, $i \in I$), i.e.

\[
\sum_{s \in S} \sum_{j \in J_s} n_{j,s}(\rho) = N \cdot \rho.
\]

(6)

The new feasible load constraint has the following property.

Property FLP-3: $\sum_{s \in S} \sum_{j \in J_s} n_{j,s}(\rho)$ is a monotonically non-decreasing function of $\rho$ and is logarithmically bounded.

If exists, the unique feasible solution to FLP with constraint (6) is still non-trivial in general, but it can be efficiently found numerically exploiting Property FLP-3. An example is illustrated in Fig. 2, where the curve is the normalized RB demand (LHS of (6)) and the line is the normalized RHS of (6). The balance between the RB demand and the resource utilization is achieved in A. At $\rho = 1$, the resource utilization (D) exceeds RB demand (B), and in C the demand at a relatively low load in other cells generates full load in own cell, which contradicts with the uniform cell load assumption. Note also a difference between the feasible load problem for a given state (finding a feasible $\rho$ for a fixed demand curve) and the network capacity problem (the maximum shift up of the curve such that a feasible load exists).

To formulate a stochastic version of FLP, we assume that user positions are random and modeled for each service $s$ as a random variable $X_s$. Given a number of users of service $s$ in a cell, $\eta_s$, the set of random user positions is modeled as a set of independent identically distributed random variables $\{x_{j,s}, j = 1, \ldots, \eta_s\}$. Let $n(x_{j,s}, \rho)$ be the RB demand by user $j$ under load $\rho$. Then, with $\eta_s$ random users, the total resource demand is $\sum_{s \in S} \sum_{j = 1}^{\eta_s} n(x_{j,s}, \rho)$ and the feasible load constraint with a $95\%$ threshold probability can be formulated as follows,

\[
\text{[FLSP]} \quad \text{P}\left(\sum_{s \in S} \sum_{j = 1}^{\eta_s} n(x_{j,s}, \rho) \leq N \cdot \rho\right) \geq 0.95, \quad \rho \in (0, \rho_{i,\max}].
\]

(7a)

(7b)

B. QoS Constraints

In this section, we introduce a set of probabilistic QoS constraints to model service and user requirements. The requirements are given per service, but in a similar way could
also be user- or user-group specific to model operator-defined QoS profiles. User bitrate requirements are given as minimum and maximum limits being the model parameters, and user throughput requirements are given as minimum targets. The minimum requirements typically define the service quality, whilst the maximum requirements are often designed to minimize service over-provisioning and maximize capacity. Both the minimum and the maximum requirements are often a part of the service subscription. Probability thresholds in the QoS constraints are the model parameters (5% are further assumed).

1) Constrained user bitrate: User bitrate requirements define a range of instantaneous bitrates of a single transmission. The minimum user bitrate is a hard constraint, i.e. a cell load at which a minimum user bitrate requirement cannot be met. For at least one service makes the load infeasible to FLSP when the QoS requirements are taken into account, i.e.,

\[ P(\bar{b}(X_s, \rho) \leq \beta_s^\text{min}) \leq 0.05 \]  

User bitrates that exceed the maximum bitrate requirement are reduced to the target level, i.e.

\[ \bar{b}(X_s, \rho) = \min[\beta_s^\text{max}, \beta_s^\text{max}, \bar{b}(X_s, \rho)] \]  

which can be viewed as a correction to the bitrate distribution. The maximum bitrate requirement is thus a soft constraint.

2) Constrained average user throughput: In a feasible load solution, the total RB consumption in the cell is \(\rho\) \(\epsilon\) \(\mathcal{N}\). Under the assumption of a round robin scheduler, the minimum throughput requirement can be modeled as a hard constraint as follows,

\[ \rho \leq \min \left\{ 1, \min_{s \in \mathcal{S}} \left\{ \frac{\nu_s}{N \cdot \beta_s^\text{min}} \right\} \right\} \]  

where \(\beta_s^\text{min}\) is the minimum user throughput for service \(s\) calculated per utilized RB in frequency. In (10), the RHS depends only on parameters and gives a simple upper bound.

3) Constrained instantaneous user throughput: Short-time, or instantaneous, user throughput can be modeled when the number of simultaneous users (average or its distribution) of each service is known. Given the expected number \(Y_s\) of users to service \(s\), the minimum instantaneous throughput can be modeled as

\[ P\left( \frac{\alpha_s \cdot \bar{b}(X_s, \rho)}{\sum_{s \in \mathcal{S}} \alpha_s \cdot Y_s(\rho)} \leq \varphi_s^\text{min} \right) \leq 0.05 \]  

where \(\varphi_s^\text{min}\) is the minimum user throughput for service \(s\), \(\alpha_s, s \in \mathcal{S}\) are the service weights that control how the resources are shared among multiple services and which, for the sake of simplicity, are assumed to be a convex combination.

C. Stochastic Feasible Load Problem with QoS Constraints

With the QoS constraints introduced in Section III-B, the stochastic feasible load model can further be extended as follows. [FLSP-QoS]

\[ P\left( \sum_{s \in \mathcal{S}} \sum_{j=1, \ldots, \eta_s} \tilde{n}(x_{js}, \rho) \leq N \cdot \rho \right) = 0.95 \]  

\[ \tilde{n}(x_{js}, \rho) \leq \beta_s^\text{min} \leq 0.05 \]  

\[ \frac{\alpha_s \cdot \bar{b}(X_s, \rho)}{\sum_{s' \in \mathcal{S}} \alpha_{s'} \cdot Y_{s'}(\rho)} \leq \varphi_s^\text{min} \leq 0.05 \]  

\[ \sum_{s \in \mathcal{S}} \alpha_s = 1 \]  

\[ \alpha_s \in [0, 1] \]  

\[ \rho \leq \min \left\{ 1, \min_{s \in \mathcal{S}} \left\{ \frac{V_s}{N \cdot \beta_s^\text{min}} \right\} \right\} \]  

In FLSP-QoS, \(\tilde{n}(\cdot)\) is the RB demand calculated for modified bitrates \(\bar{b}(X_s, \rho)\). An existing feasible solution to FLSP-QoS indicates that the traffic demand generated by all services can be satisfied with the available radio network resources such that the resource demand and the cell resource utilization are in balance and the specified QoS requirements are met.

Constraints (12c)-(12e) can be viewed as a subproblem to FLSP-QoS formulated as a linear program with no objective function (a feasibility problem):

[FLSP-QoS-sharing]

\[ F_{5\%}(\bar{b}(X_s, \rho)) \geq \frac{\varphi_s^\text{min}}{\alpha_s} \cdot \sum_{s \in \mathcal{S}} \alpha_s \cdot Y_s(\rho), \quad s \in \mathcal{S}, \quad (13a) \]

\[ \sum_{s \in \mathcal{S}} \alpha_s = 1, \quad \alpha_s \in [0, 1], \quad s \in \mathcal{S}, \quad (13c) \]

where \(F_{5\%}(\cdot)\) is the 5th percentile of the CDF of the random variable in brackets. FLSP-QoS-sharing has typically a small number of variables and is easy to tackle. A feasible solution indicates an existing service sharing scheme with which the instantaneous throughput requirement is met for each service.

To solve FLSP-QoS, we adapt an interpolation search algorithm which incorporates validity check for the QoS constraints (see Algorithm 1). The algorithm accuracy is controlled by a parameter \(c\). \(F(z)\) is used to denote the cumulative distribution function (CDF) of random \(z\), and \(F_{q\%}(z)\) denotes the \(q\)-th percentile of the CDF. To evaluate the total cell RB demand (in the LHS of (12a)), we construct a CDF as a convolution of total per-service RB demands (line 14), where the total RB demand of service \(s\) is found as an \(\eta_s\)-way autoconvolution of the RB demand CDF for an individual user of service \(s\) (line 12). A fast way to calculate convolution of multiple distributions is by means of the fast Fourier transform operation and its inverse.

In the next section, we present an application of the FLSP-QoS model in the context of radio network capacity planning.

IV. THE CAPACITY REGION PROBLEM

A. The Maximal Vector Formulation

Let \(\psi = (\eta_1, \eta_2, \ldots, \eta_S)\) denote a cell state which may or may not have a corresponding feasible load solution to FLSP-QoS. The state is said to be feasible when a feasible load \(\rho\) exists and infeasible otherwise. A network state can be defined as a set of \([2]\) elements where the \(i\)-th element is the state of cell \(i\). A network state is feasible if and only if the states of all cells are feasible. In the special network scenario considered in the paper, i.e., when all cells have the same load and are in the same state, the network state is always feasible when a cell state is feasible.

In a real life, there is no network that stays in a single state during the entire life cycle, during a day or even a shorter time period. With varying radio environment and dynamic traffic variations, a network can in principle be in any state, although the probability of each state is different and depends on many factors, e.g., time, environment conditions, users’ behavior, etc. A capacity region [3] describes a network capacity limit and consists of a set of all feasible states which cannot be further improved. For an operator, knowing the network limits is extremely important for being able to optimally configure the radio network, efficiently utilize the available resources and control traffic demand while increasing the network profitability.

The network planning task of defining the capacity region can be formulated as an optimization problem, e.g. the maximal...
vector problem [4] which aims at finding a subset of vectors such that each vector is not dominated by any other from the set. A vectors is a cell state vector. A cell state is said to be dominated by another one if the second state vector is greater than the first one. The maximal vectors are Pareto optimal and form a Pareto set which is also the set of cell states that compose the capacity region. Mathematically, the capacity region problem can be formulated as a non-linear optimization problem:

\[
\text{[CRP]} \quad \Psi^* - \text{max } \psi, \quad \text{s.t. } \psi = \{(\eta_1, \eta_2, \ldots, \eta_\mathcal{S}|), \eta_s \in \mathbb{Z}, s \in \mathcal{S} : \exists \rho | (12)\},
\]

\[
\psi \in \Psi. \quad (14b)
\]

\[
\psi \in \Psi. \quad (14c)
\]

B. A Solution Approach

To solve CRP, one could adopt an algorithm developed for the maximal vector problem or skyline design queries [4]. The problem specifically, however, is very important in developing a computationally efficient algorithm, e.g.,

- Load increases with the number of users, i.e. a dominating state has a higher load than the dominated;
- The capacity region function is not necessarily linear in the per-service number of users and is not always convex as it has been shown in [3];
- Evaluating a cell state requires solving FLSP-QoS, which may be relatively computationally expensive when (12a) is tackled directly.

Observe that the states from the Pareto set of CRP may differ in the feasible load values, but the differences are likely to be small and are only due to integrality of \(\eta_t\)-values. Otherwise, all the states in the solution would have the same load \(\rho\) which would be the maximum over all feasible states, and the CRP objective would reduce then to maximizing \(\rho\). This observation has been exploited in Algorithm 2, an approximation algorithm to CRP.

The main idea of the algorithm is to find feasible loads \(\rho\) for a basic set \(\Lambda = \{\lambda^{(i)}, l = 1, \ldots, L\}\) of \(L\) service mixes, where a service mix \([3]\) is a vector of non-negative elements

\[
\lambda^{(i)} = (\lambda_{1}^{(i)}, \ldots, \lambda_{\mathcal{S}}^{(i)}, \ldots, \lambda_{|\mathcal{S}|}^{(i)}),
\]

such that

\[
\lambda_{s}^{(i)} = -\left(\frac{\eta_{s}}{\sum_{t \in \mathcal{S}} \eta_{t}}\right), \quad s = 1, \ldots, |\mathcal{S}| - 1,
\]

and \(\sum_{t \in \mathcal{S}} \lambda_{s}^{(i)} = 1\). The minimum proposed basic set \(\Lambda^*\) contains \(|\mathcal{S}|\) standard basis vectors of the \(|\mathcal{S}|\)-dimensional Euclidean space and its normalized sum \(\frac{1}{\mathcal{S}}(1, 1, \ldots, 1)\). The maximum among the found \(\rho\)-values is then assumed to be valid for all states of the capacity region and is used for validating a larger set of service mixes \(\Lambda^*(\mathcal{N} \cap \Lambda - \emptyset)\). With a given \(\rho\), the most computationally demanding constraint (12a) can be handled in an efficient way for service mixes in \(\Lambda^*\).

Algorithm 2 uses as input the basic and the extended sets of service mixes, \(\Lambda\) and \(\Lambda^*\), and an accuracy parameter \(\varepsilon\). For each service mix in \(\Lambda\), the algorithm calls function FLSPQoS which uses Algorithm 1 to find a feasible load \(\rho\) and the corresponding state \(\psi\). Additionally, the function saves CDFs for \(1, \ldots, \eta^\text{max}_t\) users of service \(l\) for the service mixes that are the standard basis vectors. The CDFs are used in the state search to quickly validate the feasible load constraint (12a) for a given load \(\rho^*\) and a given state \(\psi\) (line 16). The other constraints of FLSP-QoS hold by default. The search over states for each service mix in \(\Lambda^*\) aims at finding a non-dominated state for which \(\rho^*\) is feasible (lines 14-21). All found non-dominated states for service mixes in \(\Lambda \cup \Lambda^*\) are stored in the capacity region set \(\Psi^*\).

V. NUMERICAL RESULTS

An illustrative simulation study has been conducted for a 3GPP LTE-like network with a regular hexagonal layout adopting a wrap-around technique. A realistic network setup, including three-sector tilted directional antennas, has been configured in a static Monte-Carlo simulator. Some of the parameters are listed in Table I. The simulator has been used to generate realistic bitrate distributions for a set of traffic loads. Two service types have been considered: voice over IP (VoIP) and web traffic. For VoIP, 12.2 kbps codec was assumed with one frame generated...
The obtained capacity region is shown in Fig. 3 where a filled contour plot of feasible loads $\rho$ for various cell states is shown. The maximum feasible load is 0.48 and it is achieved in most of the states within the capacity region. Non-normalized user bitrates (before applying the soft maximum bitrate constraints) for various loads are shown in Fig. 4. The dotted line pointed by an arrow is at the bitrate where the soft constraint for web users "corrects" the CDFs. The 5th percentile bitrates are high enough to not limit the feasible load $\rho$. Fig. 5 explains the impact of the other active QoS constraints. The normalized LHS and RHS of the load balancing equation are shown for three cell states. The last state has no feasible load. For the other two states, the resource demand and utilization are balanced within $[0,1]$ range of $\rho$ (A and B) and below $\rho - 0.5$, i.e. the maximum that satisfies the average throughput constraints. Point A, however, is behind the maximum $\rho$ that the sharing constraint admits and is therefore infeasible in FLSP-QoS. The state with 100 VoIP and 5 data users is thus the only feasible state among the three.

Note that the presented capacity figures are on a relative scale, and accurate modeling of the radio access has not been in the main scope of the study, e.g. the control channel limitation that may have an impact on VoIP capacity has not been considered.

VI. DISCUSSION

The feasible load problem and the problem of capacity planning in mixed-traffic scenarios with various QoS constraints have been investigated in the paper. With its simple and computationally efficient models, but yet reasonable and flexible, the presented framework has a great potential and can further be extended, for example, for studying admission control schemes or designing efficient operator QoS policies. However, it is important to note that not all dynamic effects can be captured by mathematical models and such studies need to be complemented by more detailed dynamic simulations, measurements analysis, etc. The models proposed in this paper are a reasonable trade-off between the amount of computations, solution space coverage, and solutions accuracy.

REFERENCES


