

Inter-Cell Coordination with Inhomogeneous Traffic Distribution

Shuping Liu and Jorma Virtamo
Networking Laboratory
Helsinki University of Technology
P.O. Box 3000, FI-02015 TKK, Finland

Abstract—We study the optimal division of cells into different service areas served under different coordinated transmission modes, i.e. combinations of base stations simultaneously transmitting, in order to avoid the adverse effect of inter-cell interference and to maximize the system capacity. We focus on the general case where the traffic load is inhomogeneous and give a complete characterization of the optimum for two-cell and three-cell systems in terms of a set of necessary conditions. An example of a three-cell system with a traffic load having a gradient in an arbitrary direction is numerically worked out.

I. INTRODUCTION

In the current trend to all-IP world, data applications are becoming more and more important also in cellular systems. Data traffic consists of so-called elastic flows that typically represent a transfer of data file or an element on a web page etc. Elasticity refers to the fact that data applications have less stringent delay requirements than, e.g., the traditional voice service and the instantaneous rates of individual flows can be adapted to the traffic conditions without a drastic reduction of the value of the service to the end user.

Data traffic in wireless system is most advantageously transmitted in a time shared fashion, as in HSDPA, in order to completely eliminate intracell interference, i.e. a base station sends with full power to one customer at a time. This kind of system may still suffer from the inter-cell interference. In particular, customers close to the cell border are adversely affected by the interference from simultaneous transmissions in neighbouring cells. System capacity can then be improved by a coordinated operation of close-by base stations; when one base station is serving a customer in a distant location the base station or base stations that have the highest interfering signal power can be switched off. There is, of course, a trade-off between the benefit obtained by the customer in question and the fact that the base stations that are switched off cannot during this time serve any customers

in their own cells. The benefits of inter-cell scheduling with fixed set of users have been studied, e.g., in [1], [2], [3].

Data traffic in cellular systems in a dynamic setting where flows come and go has been analyzed in, e.g., [4] and [5]. In particular, inter-cell coordination in this kind of dynamic setting was studied in the recent work by Bonald et al. [6] (a fuller account in report form is given in [7]) in which both interference avoidance and load balancing are addressed. The latter refers to dynamically adjusting cell boundaries in order to divert traffic from an overloaded cell to a neighbouring cell with free capacity. Bonald et al. study the system from the point view of stability region, i.e. they try to find the maximal network capacity. To this end, they consider static operation policies; finding the optimal static policy is equivalent to finding the system capacity. In practice, a static policy is not a good choice because of its sub-optimal performance at low load and because of its sensitivity to the assumed traffic distribution at high load. Fortunately, as discussed in [7] the same system capacity is obtained by more robust adaptive policies [8], [9]. The paper by Bonald et al. covers a two-cell network and a symmetric three-cell network with homogeneous traffic. They also give some interesting results pertaining to an infinite regular grid of base stations.

The main contribution of the present paper is the formulation of the general optimality conditions for the static policy in a three-cell system with an arbitrary traffic distribution. These necessary conditions characterize the optimal intra-cell boundaries between different service areas served using different transmission modes. The results are used to work out a numerical example of a three-cell system with inhomogeneous traffic distribution. Though we focus on the optimal static policy, the ultimate purpose, as in [6], is the calculation of the system capacity bounding whatever policy is used.

The rest of the paper is organized as follows. In

Section II, we define our system and basic concepts. In Section III, a two-cell system is considered as a simple introduction to the optimality conditions. The results of this section are basically contained already in [6], [7]. In Section IV, we consider the three-cell system and give the optimality conditions. A numerical example of the three-cell system is presented in Section V. Finally, Section VI concludes our paper.

II. SYSTEM DESCRIPTION

In the present paper we consider systems where a given area A is served by a finite set of base stations (BS). The area is divided into fixed cells so that each user is served by the nearest base station.

It is assumed that the operation of the base stations is fully coordinated and any number of the base stations can be actively transmitting simultaneously. By a transmission mode τ we refer to a given set of transmitting base stations. With n base stations, there are $2^n - 1$ active transmission modes (excluding the one where all base stations are switched off). The set of all transmission modes is denoted by \mathcal{T} .

The set of points that are served by BS t in transmission mode τ are denoted by $A_t^\tau, t \in \tau \in \mathcal{T}$, and is called a service area. For example, $A_1^{(1,2)}$ stands for the set of points in cell 1, served by BS1, using transmission mode (1, 2). Each cell is thus divided into different service areas served under different transmission modes. An intracell boundary between neighbouring service areas is denoted by $A_{t_1}^{\tau_1} | A_{t_2}^{\tau_2}, t_1 \in \tau_1 \in \mathcal{T}, t_2 \in \tau_2 \in \mathcal{T}$.

The location of a user in the system is denoted by \mathbf{x} . The full service rate the user obtains from a given base station, when the whole service time is devoted to that user, is called the feasible rate. Because of the interference effect from other base stations transmitting simultaneously, the feasible rate depends also on the transmission mode. The feasible rate a user at \mathbf{x} obtains from BS t in transmission mode τ is denoted by $C_t^\tau(\mathbf{x}), t \in \tau \in \mathcal{T}$.

We consider data traffic of an infinite user population (in downlink direction, the traffic destined to that population). Let the total arrival rate of flows to the system be λ and let $f(\mathbf{x})$ describe the probability density that an arriving flow is destined to a customer at \mathbf{x} (i.e., this is also the probability density of the user population). The distribution is normalized such that

$$\int_A f(\mathbf{x}) d^2x = 1. \quad (1)$$

Defining

$$\rho_t^\tau = \int_{A_t^\tau} \frac{f(\mathbf{x})}{C_t^\tau(\mathbf{x})} d^2x \quad (2)$$

and denoting the mean flow size by σ , the proportion of time needed to serve the traffic arriving in A_t^τ by BS t in transmission mode τ is $\lambda\sigma\rho_t^\tau$, i.e. $\lambda\sigma\rho_t^\tau$ is the traffic load of the area A_t^τ . Since all the service areas A_t^τ in different cells $t \in \tau$ are served using the same transmission mode τ , the proportion of time needed for transmission mode τ to be on is $\lambda\sigma \max_t \rho_t^\tau$.

Correspondingly, the total proportion of time the system is active, i.e. some transmission mode is being used, is $\lambda\sigma\rho$, where

$$\rho = \sum_{\tau \in \mathcal{T}} \max_t \rho_t^\tau. \quad (3)$$

The stability condition of the system requires that,

$$\lambda\sigma\rho \leq 1. \quad (4)$$

Thus for a given ρ , the maximum value of λ is $1/(\rho\sigma)$, and the system capacity is

$$C = \lambda_{\max} \cdot \sigma = \frac{1}{\rho}. \quad (5)$$

Our objective is to make the capacity as large as possible. This is achieved by choosing the division of the whole area A into service areas $A_t^\tau, t \in \tau \in \mathcal{T}$, such as to minimize ρ ,

$$\rho^* = \min_{\{A_t^\tau\}} \rho = \min_{\{A_t^\tau\}} \sum_{\tau \in \mathcal{T}} \max_{t \in \tau} \int_{A_t^\tau} \frac{f(\mathbf{x})}{C_t^\tau(\mathbf{x})} d^2x. \quad (6)$$

The optimized system capacity C^* is then,

$$C^* = \frac{1}{\rho^*}. \quad (7)$$

Our task is to find the boundaries between areas $A_t^\tau, t \in \tau \in \mathcal{T}$, that realize the minimum of ρ in equation (6).

III. TWO-CELL LINEAR SYSTEM

We illustrate the principles by first considering a two-cell linear configuration depicted in Figure 1. This system could represent, for instance, a road scenario. This simple system was analyzed also in [6] and the results of this section are, in essence, contained in those of [6]. We concentrate on the general case with inhomogeneous traffic load, i.e. we allow an arbitrary distribution of the user population on the line.

As there are two base stations, there are three possible transmission modes, $\mathcal{T} = \{(1), (2), (1, 2)\}$, where either only one of the base stations is active or both of them are active. Accordingly, there are potentially four service

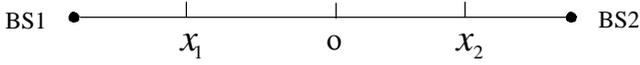


Fig. 1. Linear two-cell system with service area boundaries.

areas: $A_1^{(1)}$, $A_2^{(2)}$, $A_1^{(1,2)}$, and $A_2^{(1,2)}$. Each of the cells is potentially divided into two areas. Users close to a base station are always most advantageously served by letting both base stations be on; as the relative effect of the interference from the other station is negligible, it is a good idea to let the other station simultaneously do useful work. When serving users closer to the cell boundary, however, it may be advantageous to switch the other station off if the gain in the feasible rate in absence of the interference is big enough.

The four areas are delineated by the cell boundary in the middle between the base stations (marked by 0) and by two other points denoted x_1 and x_2 . Without loss of generality we can assume that the base stations are at a unit distance from each other: BS1 is at point 0 and BS2 is at point 1 and the cell boundary is at point 1/2.

According to (3), the total fraction of time needed to serve the whole traffic load is $\lambda\sigma\rho$ with

$$\rho = \max\{\rho_1^{(1,2)}, \rho_2^{(1,2)}\} + \rho_1^{(1)} + \rho_2^{(2)}, \quad (8)$$

where the different load components ρ_i^j defined by (2)¹ depend on the choice of the boundaries x_1 and x_2 . Note x_1 or x_2 may coincide with point 0, so that one or two of the areas may shrink to zero. According to (6), our task is to find the optimal boundaries to minimize ρ . Next we will give the optimality conditions.

A. Characterizing the optimal boundaries

The optimal division is fully characterized by two necessary conditions for optimality. Together, they suffice to uniquely define the optimal boundaries.

Condition 1 The traffic loads of service areas $A_1^{(1,2)}$ and $A_2^{(1,2)}$ must be equal, i.e.

$$\rho_1^{(1,2)} = \rho_2^{(1,2)}. \quad (9)$$

Proof

The proof is by contradiction. So we assume that Condition 1 is not satisfied by the optimal division and reason that ρ can be decreased in contradiction with the optimality. To this end, suppose, e.g. $\rho_1^{(1,2)} > \rho_2^{(1,2)}$. This

¹In the case of a linear system, the distribution $f(x)$ is, of course, one-dimensional and correspondingly (2) is replaced with an integral over a single variable.

means that using transmission mode (1,2) the service of the traffic in area $A_2^{(1,2)}$ completes in a shorter time than that of $A_1^{(1,2)}$. Then it is clear that after completion of the service of $A_2^{(1,2)}$ it is advantageous to switch BS2 off and serve some of the customers in $A_1^{(1,2)}$ with a higher feasible rate in mode (1) instead of mode (1,2). Indeed, because $C_1^{(1)} > C_1^{(1,2)}$, the increase in $\rho_1^{(1)}$ is less than the decrease in $\rho_1^{(1,2)}$, i.e. in the maximum term of (8), with the net effect of decreasing ρ .

Next we establish that if both service areas $A_1^{(1)}$ and $A_2^{(2)}$ do exist, i.e. neither x_1 nor x_2 coincides with the cell boundary, then necessarily the following condition holds.

Condition 2 If the optimal division is such that $x_1 \neq \frac{1}{2}$ and $x_2 \neq \frac{1}{2}$, then one must have

$$\frac{C_1^{(1,2)}(x_1)}{C_1^{(1)}(x_1)} + \frac{C_2^{(1,2)}(x_2)}{C_2^{(2)}(x_2)} = 1. \quad (10)$$

Proof

The assumption $x_1 \neq \frac{1}{2}$ and $x_2 \neq \frac{1}{2}$ means that each of the boundary points can be moved in either direction (by a small amount) without any constraint. This is to say that the presumed minimum of ρ with respect to x_1 and x_2 is obtained in an interior point of the feasible region of these variables. Condition (10) represents then, in essence, the requirement that derivative of the function to be minimized equals zero. In fact, whatever small change we make in the values of x_1 and x_2 , the change of ρ to the first order in terms of these changes must equal zero. In particular, we choose to make a small change in the boundaries in such a way that the traffic load both in area $A_1^{(1,2)}$ and in $A_2^{(1,2)}$ increases by the same amount $d\rho$, cf. Figure 2. Then the maximum term in (8) increases by the amount $d\rho$. Equating this increase to the first order decrease of the sum $\rho_1^{(1)} + \rho_2^{(2)}$, by making use of (2), we have,

$$d\rho = d\rho \times \frac{C_1^{(1,2)}(x_1)}{C_1^{(1)}(x_1)} + d\rho \times \frac{C_2^{(1,2)}(x_2)}{C_2^{(2)}(x_2)},$$

from which (10) follows.

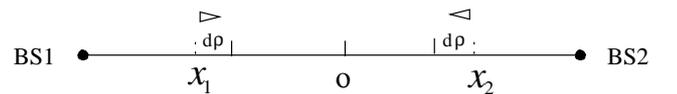


Fig. 2. Changing boundaries $A_1^{(1,2)}|A_1^{(1)}$ and $A_2^{(2)}|A_2^{(1,2)}$.

Conditions 1 and 2 uniquely define the optimal locations x_1 and x_2 . One can find the optimal values

by starting from $x_1 = 0$, whence by Condition 1 $x_2 = 1$. At this point $C_1^{(1,2)}(x_1)/C_1^{(1)}(x_1) = 1$ and $C_2^{(1,2)}(x_2)/C_2^{(2)}(x_2) = 1$ because close to a base station the interference is negligible in relative terms. So the left hand side of (10) equals 2. Moving x_1 and x_2 inwards, so that Condition 1 is always satisfied, the value of the expression on the left hand side of (10) decreases. The optimum is obtained when either this expression attains the value 1 or either x_1 or x_2 reaches the cell boundary in the middle, whichever happens first.

IV. THREE-CELL PLANAR SYSTEM

In this section, we consider a system of three identical base stations at the corners of a triangle. The cell boundaries are assumed fixed and are defined by the mid-planes between pairs of base stations like in a Voronoi tessellation. The considerations are similar as in the linear case, but the 2-dimensional planar geometry requires some extensions of the results.

Now there are 7 different transmission modes, $\mathcal{T} = \{(1), (2), (3), (1, 2), (1, 3), (2, 3), (1, 2, 3)\}$, where either one, two or three base stations are active simultaneously. From the outset we assume that each service area A_t^τ , $t \in \tau \in \mathcal{T}$, is a compact set. Though we do not give a proof that this is the case for the optimal division, it is very plausible that this is true. The whole system could be divided into 12 service areas as illustrated in Figure 3. The cell boundaries, oa , ob and oc are also depicted in the figure. Again, it is possible that some of the service areas shrink to non-existence, though in the following for the sake of simplicity we implicitly assume that they all do exist.

With the aid of (3) we write the total fraction of time

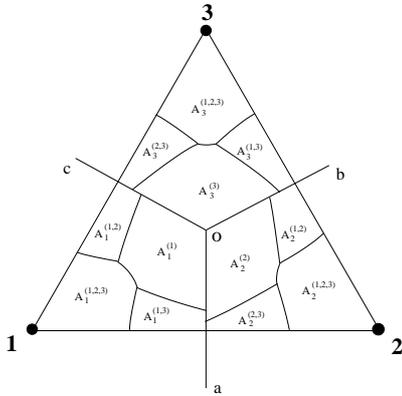


Fig. 3. The division of the three-cell system into service areas.

$$\rho = \max\{\rho_1^{(1,2,3)}, \rho_2^{(1,2,3)}, \rho_3^{(1,2,3)}\} + \max\{\rho_1^{(1,2)}, \rho_2^{(1,2)}\} + \max\{\rho_1^{(1,3)}, \rho_3^{(1,3)}\} + \max\{\rho_2^{(2,3)}, \rho_3^{(2,3)}\} + \rho_1^{(1)} + \rho_2^{(2)} + \rho_3^{(3)}, \quad (11)$$

and our task again is to find the boundaries between the service areas so as to minimize this expression.

A. Characterizing the optimal boundaries

In the following, we characterize the optimal division of the cells into service areas by giving three necessary conditions that have to be satisfied by optimal boundaries.

Throughout the rest of the paper we use the following short hand notation for the feasible rate ratios at a given point \mathbf{x} ,

$$u_t(\mathbf{x}) = \frac{C_t^{(1,2,3)}(\mathbf{x})}{C_t^{(t)}(\mathbf{x})}, \quad \forall t,$$

$$v_{tt'}(\mathbf{x}) = \frac{C_t^{(1,2,3)}(\mathbf{x})}{C_{t'}^{(t,t')}(\mathbf{x})}, \quad \forall t, t', t \neq t',$$

$$w_{tt'}(\mathbf{x}) = \frac{C_{t'}^{(t,t')}(\mathbf{x})}{C_{t'}^{(t)}(\mathbf{x})}, \quad \forall t, t', t \neq t'.$$

As illustrated in Figure 4, at the intracell boundaries between service areas, each of these quantities gives the ratio of feasible rates at different sides of a given type of boundary.

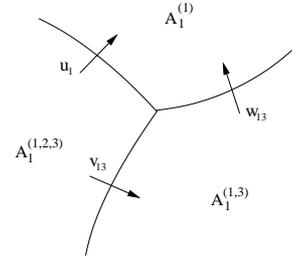


Fig. 4. An example of the definitions of the service rate ratios.

Condition 3 The traffic loads of service areas in different cells served under a given transmission mode are equal, i.e. $\rho_t^{(\tau)} = \rho_{t'}^{(\tau)}$ for all $t \in \tau$, $t' \in \tau$, $t \neq t'$. Explicitly, this gives five equations

$$\begin{aligned} \rho_1^{(1,2,3)} &= \rho_2^{(1,2,3)} = \rho_3^{(1,2,3)}, \\ \rho_1^{(1,2)} &= \rho_2^{(1,2)}, \\ \rho_1^{(1,3)} &= \rho_3^{(1,3)}, \\ \rho_2^{(2,3)} &= \rho_3^{(2,3)}. \end{aligned} \quad (12)$$

Proof

The proof of the above necessary condition is essentially the same as that of Condition 1.

Condition 4 With an optimal division into service areas, the ratio of the feasible rates $C_t^r(\mathbf{x})$ and $C_t^{r'}(\mathbf{x})$ in any cell t is constant on a boundary $A_t^r|A_t^{r'}$ between two service areas. That is, $u_t(\mathbf{x}) = u_t$, $v_{tt'}(\mathbf{x}) = v_{tt'}$, and $w_{tt'}(\mathbf{x}) = w_{tt'}$ are constants on the respective boundaries for all t and t' , $t \neq t'$.

Proof

We already know by Condition 3 that with an optimal division minimizing ρ , the service areas in different cells served under a given transmission mode have equal loads, as stated in (12). Now we wish to prove that along each intracell border between two service areas the ratio of the respective feasible rates is constant. We prove this by contradiction by assuming that this is not the case and showing that this leads to a contradiction.

In any cell t , we have three types of borders to consider: $A_t^t|A_t^{t,t'}$, $A_t^t|A_t^{(1,2,3)}$, and $A_t^{t,t'}|A_t^{(1,2,3)}$, with $t \neq t'$. The proof for the second type of border is identical with that of the first one. So we consider only the first and last types of border.

We start with the boundary of type $A_t^t|A_t^{t,t'}$ and make the counter assumption that the feasible rate ratio $w_{tt'}(\mathbf{x})$ is not constant on the border. In particular, let \mathbf{x}_1 and \mathbf{x}_2 be two points on the border such that $w_{tt'}(\mathbf{x}_1) > w_{tt'}(\mathbf{x}_2)$. Now distort the border about the points \mathbf{x}_1 and \mathbf{x}_2 as shown in Figure 5a, in such a way that the load $\rho_t^{(t,t')}$ does not change (this is always possible by adjusting the sizes of the modified areas appropriately). Then it follows from the assumed inequality and from the definition (2) that $\rho_t^{(t)}$ becomes smaller. As other terms in (11) are not affected, the total effect is that ρ could be made smaller in contradiction with the optimality.

Next, consider the boundary of type $A_t^{t,t'}|A_t^{(1,2,3)}$ and assume that the feasible rate ratio $v_{tt'}(\mathbf{x})$ is not constant

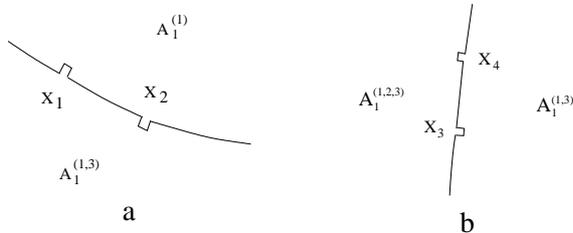


Fig. 5. Distortions of intracell service area boundaries $A_1^{(1)}|A_1^{(1,3)}$ (a) and $A_1^{(1,3)}|A_1^{(1,2,3)}$ (b) needed in the proof of Condition 4.

on the border. By following the above argument, we find two points \mathbf{x}_3 and \mathbf{x}_4 on the border such that $v_{tt'}(\mathbf{x}_3) > v_{tt'}(\mathbf{x}_4)$. In the same way as above, the border can be distorted about the points \mathbf{x}_3 and \mathbf{x}_4 as shown in Figure 5b, in such a way that the load $\rho_t^{(1,2,3)}$ does not change while the load $\rho_t^{(t,t')}$ gets smaller. Because initially, by the assumed optimality and Condition 3, we have the equality $\rho_t^{(t,t')} = \rho_{t'}^{(t,t')}$, the term $\max\{\rho_t^{(t,t')}, \rho_{t'}^{(t,t')}\}$ in (11) does not change though $\rho_t^{(t,t')}$ becomes smaller. Thus we are in a situation where ρ still equals the minimum of an optimal system but $\rho_t^{(t,t')} < \rho_{t'}^{(t,t')}$ in contradiction with Condition 1.

Remark 1 Condition 4 defines the type of any intracell boundary curve: each of them belongs to a known family of curves, viz. equislice contours of the ratio of feasible rates in the respective transmission modes. A curve is uniquely defined as soon as the value of the constant ratio is given. In each cell, there are four service areas with five boundaries. Altogether, all the intracell service area boundaries in the system are defined by 15 unknown constants.

Remark 2 By considering a vertex where three boundaries meet at a single point \mathbf{x} , one immediately sees that the constants satisfy the following so-called vertex conditions (cf. Figure 4):

$$u_t = v_{tt'}w_{tt'}, \quad \forall t, t', t \neq t'. \quad (13)$$

Altogether, there are 6 conditions of this type corresponding to two vertices in each cell.

In analogy with Condition 2, we have the following necessary condition.

Condition 5 The constants defining the optimal boundaries satisfy the relations

$$u_1 + u_2 + u_3 = 1, \quad (14)$$

$$w_{tt'} + w_{t't} = 1, \quad \forall t, t', t \neq t'. \quad (15)$$

Proof

By Conditions 3 and 4 we already know that the optimal boundaries are such that the traffic equalities (12) hold and that each boundary curve is of the type ‘constant feasible rate ratio’. In addition, analogously with the proof of Condition 2, we reason that if the boundaries are optimal, then ρ of (11) must be invariant, to the first order, to any infinitesimal distortion of the boundary curve (‘first derivative equals zero’).

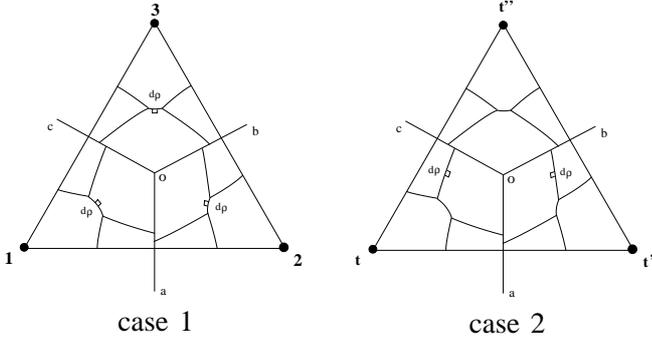


Fig. 6. Distortions of intracell service area boundaries for the proof of Condition 5.

First, we choose to distort the boundaries in the way shown in Figure 6, case 1, i.e. each of the boundaries $A_t^{(t)}|A_t^{(1,2,3)}$, $t = 1, 2, 3$, is slightly distorted so that all the loads $\rho_t^{(1,2,3)}$, $t = 1, 2, 3$, increase by the same amount $d\rho$ (there are infinitely many ways to do this). Then, also the first term of (11) increases by the amount $d\rho$. At the same time, each of the terms $\rho_t^{(t)}$ decreases by an amount which, in view of definition (2) and the constancy of the feasible rate ratio along the boundary, to the first degree equals $d\rho \times u_t$. From equating the positive and negative changes in (11),

$$d\rho = d\rho \times u_1 + d\rho \times u_2 + d\rho \times u_3, \quad (16)$$

equation (14) follows.

Similarly, for any pair of cells t and t' , we can consider the type of boundary distortion shown in Figure 6, case 2, where the boundaries $A_t^{(t)}|A_t^{(t,t')}$ and $A_{t'}^{(t')}|A_{t'}^{(t,t')}$ are distorted so that the loads $\rho_t^{(t,t')}$ and $\rho_{t'}^{(t,t')}$, increase by the same amount $d\rho$. Again, by equating the increase of the term $\max\{\rho_t^{(t,t')}, \rho_{t'}^{(t,t')}\}$ by $d\rho$ to the total decrease of the terms $\rho_t^{(t)}$ and $\rho_{t'}^{(t')}$,

$$d\rho = d\rho \times w_{t,t'} + d\rho \times w_{t',t}, \quad (17)$$

equation (15) follows.

Remark 3 In addition to the single condition (14), there are three independent conditions of type (15). Altogether, these equations set four conditions on the unknown constants.

The above conditions suffice to determine the optimal division of the cells into service areas. As noted before, in the general case there are 15 boundaries between intracell service areas, i.e. 15 unknown constants. Remark 2 provides six equations between the constants, Condition 3 gives five equations, and Condition 5 four equations.

That is, we have exactly as many equations as there are unknowns, and all the constants can be solved. This is illustrated by a numerical example in the next section.

V. NUMERICAL RESULTS

In this section we demonstrate the application of the previous results for finding the optimal boundaries in a three-cell system with an inhomogeneous traffic load. The base stations are assumed to be located at the corners of an equilateral triangle, with positions \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 .

We assume that the feasible rate depends on the signal-to-interference-plus-noise-ratio (SINR) through the Shannon formula:

$$C_t^\tau(\mathbf{x}) = B \log_2(1 + \text{SINR}_{\mathbf{x},\tau}),$$

where B is the bandwidth of the channel and $\text{SINR}_{\mathbf{x},\tau}$ is the SINR of a user located at \mathbf{x} when the transmission mode is τ . For an \mathbf{x} belonging to a cell $t \in \tau$, this is given by

$$\text{SINR}_{\mathbf{x},\tau} = \frac{g(|\mathbf{x} - \mathbf{r}_t|)}{n + \theta \sum_{t' \in \tau, t' \neq t} g(|\mathbf{x} - \mathbf{r}_{t'}|)},$$

where θ is the CDMA orthogonality factor and n is the normalized noise power, $n = (N_0 B)/P$, with N_0 being the spectral noise power and P the transmission power of a base station (assumed to be the same for all base stations). The function $g(r)$ defines the attenuation of the signal power as a function of distance r . For this we use the standard model $g(r) = r^{-\alpha}$. The attenuation exponent α depends on radio propagation conditions, typically being in the range $\alpha = 2, \dots, 4$. In this study, we use the numerical values $n = 0.05$, $\theta = 1$, and $\alpha = 3$.

The traffic distribution $f(\mathbf{x})$ is assumed to be a linear function of position \mathbf{x} , i.e. to have a constant gradient. The distribution is depicted in Figure 7, where a dark shade of grey represents a low and a light shade of gray a high traffic density. The values of $f(\mathbf{x})$ at the corners

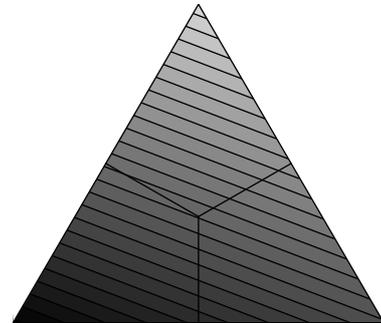


Fig. 7. Traffic distribution with a gradient. Light and dark shades of gray correspond to high and low traffic densities, respectively.

of the triangle in our example are 0.5, 0.9 and 1.6, starting from the left bottom corner in counterclockwise direction, with the average value of $f(\mathbf{x})$ equalling 1.

In each cell, there are four different service areas, separated by five boundaries. Each boundary is defined by a single parameter. So altogether there are 15 unknowns in our problem (the parameters u_t , $v_{tt'}$ and $w_{tt'}$). By equations (13), (14) and (15), 10 of these unknowns can be eliminated by expressing them in terms of the five variables, e.g., u_1 , u_2 , w_{12} , w_{23} , and w_{31} ,

$$u_3 = 1 - u_1 - u_2,$$

$$\begin{aligned} v_{12} &= \frac{u_1}{w_{12}}, & v_{21} &= \frac{u_2}{1 - w_{12}}, \\ v_{23} &= \frac{u_2}{w_{23}}, & v_{32} &= \frac{1 - u_1 - u_2}{1 - w_{23}}, \\ v_{31} &= \frac{1 - u_1 - u_2}{w_{31}}, & v_{13} &= \frac{u_1}{1 - w_{31}}, \end{aligned}$$

$$w_{21} = 1 - w_{12}, \quad w_{32} = 1 - w_{23}, \quad w_{13} = 1 - w_{31}.$$

The five unknowns can finally be solved from the five traffic load equations (12).

The solution of the traffic load equations is somewhat tricky but we were able to do it numerically with the aid of *Mathematica*. The resulting optimal division into different service areas is shown in Figure 8, where the areas served under a given transmission mode are shown using the same shade of gray. The three separate areas close to the corners are all served under the transmission mode (1,2,3), i.e. with all the transmitters simultaneously on. The three areas in the central region are all served with just the closest base station on in order to avoid interference. For the transmission modes where two of the base stations are on simultaneously, the service areas consist of two pieces such that for each piece the closest, serving base station is on and the closest interfering station is off.

Corresponding to the fact that traffic intensity is highest at the top of the triangle, the service areas $A_3^{(1,2,3)}$, $A_3^{(1,3)}$, and $A_3^{(2,3)}$ are smaller than their counterparts in other cells, in order to satisfy the traffic load equalities (12).

VI. CONCLUSIONS

In this paper we have studied the problem of optimal inter-cell coordination (scheduling) in two-cell and three-cell cellular systems, with the aim to maximize the system's capacity, i.e. the overall load that can be carried

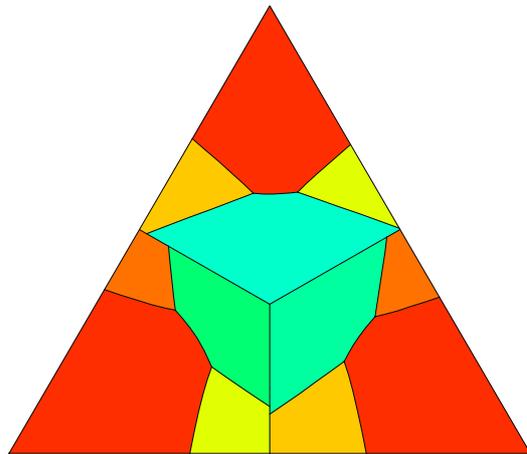


Fig. 8. The optimal division of a three-cell system into service areas with the inhomogeneous traffic load of Figure 7.

retaining stability. In particular, we allow a general distribution of the offered traffic within the considered area. A set of necessary conditions satisfied by the optimal division of the cells into service areas (served under different transmission modes) was formulated, along the lines set forth by Bonald et al. in [6], [7]. Together, these conditions suffice to define the optimal service area division uniquely.

For the case of a two-cell system our study is exhaustive. For the three-cell system we focussed on the case where all the potential service areas do indeed exist. Considering all the possible cases of vanishing areas leads to a rather tedious analysis.

The optimal service area division and the associated schedule for different transmission modes define a static operating policy for the system that stretches the capacity limit as far as possible. As noted in Introduction, a static policy itself is not practical. First, below the capacity limit its performance is suboptimal. At the extreme, when there is only one flow in the system, the corresponding base station should be wholly devoted to serve the flow, keeping other base stations switched off. This is in contrast to the static policy where different transmission modes are allocated fixed time shares. Second, at high load, a static policy is sensitive to the exact knowledge of the traffic distribution. The main point in our study, however, is that the capacity limit calculated with the static policy defines the ultimate limit which cannot be exceeded by any policy. Well-designed dynamic policies, where the resource sharing and scheduling are state dependent, achieve this limit, i.e. are stable when the load is below the limit, without the need to know the traffic distribution in advance.

A topic outside the scope of the present paper is the evaluation of the performance of a system operating below the capacity limit. The analysis of the performance under dynamic operating policies is usually much more difficult than just finding the capacity limit. However, when the resource sharing is based on the principle of balanced fairness, this kind of analysis is facilitated for some systems as discussed in, e.g., [10]-[12].

In our study, we have assumed that the cell boundaries are fixed. We have made this choice primarily for the sake of clarity of presentation. It is rather straightforward to let also the cell boundaries to be subject to optimization, i.e. in addition to interference avoidance by coordinated scheduling to do load balancing between the cells. Again, the optimal boundaries can be characterized by the property of constancy of certain feasible rate ratios. In particular, in the central area of the three-cell system, between the service areas served with a single base station on, the assumed mid-plane boundaries actually are the optimal ones.

In this paper we have considered only small systems of two or three cells. In a real network, the number of cells is, of course, much larger. Though a larger system can, in principle, be analyzed using the same kind of approach, such an analysis for a general network becomes easily very complex when the number of base stations increases since the number of possible transmission modes grows exponentially. In practice, one cannot hope to get any useful results except for regular grids where symmetry principles can be exploited. A very nice work in this direction has been started in [6].

ACKNOWLEDGMENTS

This work was financially supported by the Academy of Finland (grant no. 202204).

REFERENCES

- [1] A. Bedekar, S.C. Borst, K. Ramanan, P.A. Whiting and E.M. Yeh, *Downlink scheduling in CDMA data networks*, Proc. of Globecom'99, 1999, 2653–2657.
- [2] J.M. Holtzman and S. Ramakrishna, *A scheme for throughput maximization in a dual-class CDMA system*, IEEE J. Sel. Areas Commun. 40, 1998, 830–844.
- [3] S. Ramakrishna, *Optimal scheduling of CDMA systems*, Ph.D. thesis, WINLAB, Rutgers University (1998).
- [4] T. Bonald, S.C. Borst and A. Proutière, *How mobility impacts the flow-level performance of wireless data systems*, Proc. IEEE Infocom 2004, 2004, 1872–1881.
- [5] S.C. Borst, *User-level performance of channel-aware scheduling algorithms in wireless data networks*, Proc. IEEE Infocom 2003, 2003, 636–647.
- [6] T. Bonald, S.C. Borst and A. Proutière, *Inter-cell scheduling in wireless data networks*, Proc. of European Wireless'05, 2005, 566–572.

- [7] T. Bonald, S.C. Borst and A. Proutière, *Inter-cell scheduling in wireless data networks*, CWI report, 2004.
- [8] M. Armony and N. Bambos, *Queueing networks with interacting service resources*, Proc. 37th Allerton Conf. Commun. Control, Comp., 1999, 42–51.
- [9] J. Mo and J.C. Walrand, *Fair end-to-end window-based congestion control*, IEEE/ACM Trans. Netw., 2000, 556–567.
- [10] T. Bonald, L. Massoulié, A. Proutière and J. Virtamo, *A Queueing Analysis of Max-Min Fairness, Proportional Fairness and Balanced Fairness*, to appear in Queueing Systems, 2006.
- [11] J. Leino, A. Penttinen and J. Virtamo, *Flow Level Performance Analysis of Wireless Data Networks: A Case Study*, to appear in the proceedings of the 2006 IEEE International Conference on Communications, ICC 2006, Istanbul, 2006.
- [12] T. Bonald, A. Penttinen and J. Virtamo, *On Light and Heavy Traffic Approximations of Balanced Fairness*, to appear in the proceedings of Sigmetrics/Performance 2006, Saint-Malo, 2006.