The Tradeoff Between Coverage and Capacity in Dynamic Optimization of 3G Cellular Networks

Georg Hampel, Kenneth L. Clarkson, John D. Hobby, and Paul A. Polakos Bell Laboratories – Lucent Technologies 600 Mountain Ave, Murray Hill, NJ 07974, USA

Abstract— For 3G cellular networks, capacity is an important objective, along with coverage, when characterizing the performance of high-data-rate services. In live networks, the effective network capacity heavily depends on the degree that the traffic load is balanced over all cells, so changing traffic patterns demand dynamic network reconfiguration to maintain good performance. Using a four-cell sample network, and antenna tilt, cell power level and pilot fraction as adjustment variables, we study the competitive character of network coverage and capacity in such a network optimization process, and how it compares to the CDMA-intrinsic coverage-capacity tradeoff driven by interference. We find that each set of variables provides its distinct coverage-capacity tradeoff behavior with widely varying and application-dependent performance gains. The study shows that the impact of dynamic load balancing highly depends on the choice of the tuning variable as well as the particular tradeoff range of operation.

Keywords – Coverage optimization; capacity optimization; 3G cellular networks; coverage-capacity tradeoff; load balancing

I. INTRODUCTION

With the introduction of high-data-rate services, network capacity has become a crucial aspect of the performance of 3G access networks [1][2]. It therefore must be considered along with network coverage as an objective during network planning [3]–[7]. Since network capacity strongly depends on how well the actual traffic load is balanced over all cells, dynamic load balancing routines that adjust one or more cell parameters in response to changing traffic patterns can be applied after a network has gone into service [3][8][9]. Such load balancing operations, in turn, have to include coverage as on additional objective or constraint [5][6].

For GSM/GPRS and TDMA IS-136, cell load variations due to changing traffic patterns can be rebalanced by changing the number of channels per cell, and a subsequent recomputation of the frequency plan helps to mitigate the interference problems this introduces [8][9][10]. This approach is not feasible for CDMA networks because the wide carrier bandwidth prevents load balancing by swapping carriers among cells, and a frequency plan to mitigate interference does not exist. Load balancing is therefore done by moving cell boundaries and soft handoff areas via the adjustment of cell locations, antenna configurations and power levels, each of which affects both objectives simultaneously [11]–[15]. Since the functional dependence between objectives and adjustment parameters is complicated, automated optimization approaches have been proposed by various authors [7][11][12][14].

In this paper, we illustrate the general problem of simultaneously addressing coverage and capacity objectives in 3G cellular network optimization, as it occurs when loadbalancing operations are performed via coordinated adjustment of cell parameters. This analysis includes the CDMA-generic coverage-capacity tradeoff, which is caused by the traffic-load dependence of interference as discussed by Wheatley [5], and Veeravalli et al. [6].

The study is done on a four-cell model network with a square layout, which is incompatible with the optimal hexagonal pattern and therefore represents the layout irregularities found in real markets due to cell placement constraints. The small size of the model network makes it an ideal candidate for studies since performance changes can easily be attributed to particular configuration adjustments. In this paper, we only analyze the downlink, assuming it will be the weaker link in the presence of asymmetrical load as expected for 3G data services [16].

We consider optimizing antenna tilt, cell power level, and pilot power fraction as adjustment variables for the loadbalancing operation. We find that even for uniform traffic and under perfectly balanced load, optimization provides a set of network configurations with a larger coverage-capacity tradeoff than given through the interference-driven CDMA-generic tradeoff mechanism. As can be expected, the associated performance gains grow with increasing traffic non-uniformity. Our study demonstrates how the complete tradeoff characterization can be used to differentiate and rank potential tuning parameters with respect to their application-specific impact on overall network performance.

In the following, we first present the definitions of network coverage and capacity and the tradeoff between these objectives. Next, we describe the performance modeling for the four-cell network and then we present and discuss the tradeoffs obtained from optimization of the model network with respect to the different variables. Finally, we discuss the impact these variables have on overall network performance when used to load-balance the network to traffic pattern variations.



Figure 1. A network performance plot depicting network coverage and carried traffic for all network configurations (shaded area). The optimal tradeoff curve (solid and dashed line) represents the performance of all solutions of the optimization process. The dashed-dotted and the dashed lines represent solutions due to the interference-based coverage-capacity tradeoff.

II. TRADEOFF BETWEEN NETWORK CAPACITY AND COVERAGE

The outcome of a network performance analysis and optimization can be very dependent on the particular choice of objective functions. In this paper, we focus on performance enhancements in live networks. Since real markets are characterized by irregular network layout, complex propagation patterns, and inhomogeneous traffic distributions that give each cell its own shape, performance and traffic load, coverage and capacity have to be defined in terms of network properties rather than cell properties. This means that the capacity measure must capture the degree of load balancing among cells. The coverage measure should be weighted by the local traffic density to capture the network-wide fraction of users who receive adequate service.

In this study, we restrict ourselves to a deterministic traffic distribution with one service that is characterized through one average transmission data rate. The deterministic traffic distribution would correspond to a time snapshot evaluation of network performance shorter than average call time, such that traffic fluctuations due to the statistical nature of user arrival can be neglected. With these assumptions, and the given traffic distribution, the coverage and capacity definitions become rather simple: We define capacity as the maximum carried traffic that is compliant with a given coverage constraint, and we define coverage as the fraction of the offered traffic that can receive service with guaranteed minimum quality. The fraction of covered traffic can be regarded as equivalent to the trafficweighted coverage area.

Since the traffic distribution is deterministic, call blocking won't occur before at least one cell has reached maximum load. In the zero-blocking regime, lack of coverage becomes the only network-access-limiting factor, which relates the two objectives through a simple expression: With the focus set on dynamic load balancing, one could interpret call blocking as unacceptable performance degradation and hence disregard solutions with non-zero blocking entirely. In our sample-network study, we found all solutions with non-zero blocking to be sub-optimal, which allowed us to stay within the zero-blocking regime.

With these definitions, the performance of different network configurations and their individual tradeoffs derived from various coverage constraints can be compared in a twodimensional plot with network coverage and network capacity as components (Fig.1). The overall set of all potential solutions forms a cloud of points in this plot. A perfect network optimization algorithm that tries to maximize the coverage and capacity objectives would naturally select all those configurations whose performance points form the upper right boundary of this cloud (solid and dashed line in Fig.1). We call these solutions optimal tradeoff points and the curve they form an optimal tradeoff curve. The length of the optimal tradeoff curve and its shape are measures for how well network coverage and capacity can be simultaneously optimized. All other points that lie within the cloud can be considered suboptimal solutions. The optimal tradeoff curve apparently consists of two sections: an upper section (solid line in Fig.1), forming a configuration-based tradeoff, where all solutions correspond to different network configurations under maximum traffic load, and a lower section, forming an interference-based tradeoff for only one configuration (dashed line in Fig.1). This lower section of the optimal tradeoff curve corresponds to the configuration that delivers best coverage. The interference-based tradeoff is identical to the coveragecapacity tradeoff discussed by Veeravalli et al. [6].

The strength of this tradeoff analysis is that potential improvements of optimization through coordinated adjustment of cell parameters can be analyzed and compared to the interference-based tradeoff that CDMA technologies give generically. The relative improvement in coverage when going from configuration A to C in Fig.1, for instance, is rather small when evaluated at the same traffic load (ΔCov in Fig.1). The potential capacity growth available for configuration A over configuration C comes at a price of a substantial coverage reduction that is significantly larger than ΔCov .

One may argue that the tradeoff analysis shown here is based on the absence of traffic fluctuations, where zero blocking prevails until one cell reaches maximum load. However, if traffic fluctuations are included and call blocking is non-zero for any non-zero traffic load, the same approach can still be used. Here either a maximum tolerable blocking rate per cell has to be determined, which sets an upper bound to traffic load, or network blocking and coverage are summarized into one common performance metric that measures the likelihood of access to the network.

III. FOUR-CELL NETWORK PERFORMANCE MODELING

The model network consists of four cell sectors at the corners of a square (Fig.2). We will optimize the model network once for uniform traffic and once for an example of a



Figure 2. A layout for a four-cell model network with non-uniform traffic distribution. The traffic density increases diagonally across the network area.

non-uniform traffic distribution. The first case is used to study optimization of imperfect network layouts, and the second case serves to study the performance sensitivity to traffic pattern variations and the potential gain that can be achieved through network reoptimization. In the simulation, performance will be evaluated over a uniform grid of $m \times m$ points, where the traffic of the *i*th point at location (x_i, y_i) has traffic T_i . The two traffic distributions are as follows (Fig.2):

Uniform:

Non-uniform: $T_i = T_0 \cdot (1 - x_i + y_i), \quad x_i, y_i \in [0,1]$

Here, the dimension of the square layout has been set to one, and T_0 represents the total number of users per area.

 $T_i = T_0$

To give a clear picture of the coverage-versus-capacity tradeoff in 3G-network optimization, we will simplify the performance modeling so that it only includes the essential features. We will focus on the downlink using one common pilot channel and a power-controlled dedicated traffic channel with constant data rate. Power control is assumed to be perfect, and network access is assumed to be limited by power amplifier (PA) overload, as can be expected for many 3G applications.

The vertical antenna pattern of each cell sector is approximated by a \cos^2 - pattern for the main lobe, without side lobes and finite back-lobe suppression, and the azimuthal pattern does not have any structure. The antenna height is set in relative terms by the down angle from the antenna to the center of the network.

The common pilot channel coverage at location *i* is given by $(Ec/Io)_i$ of the strongest server:

Location i has pilot coverage $\Leftrightarrow \max_{k} (Ec/Io)_{ik} \ge \theta_{Pl}$ (2)

where *Ec* denotes the received common pilot channel signal strength of the k^{th} server, θ_{Pl} an absolute pilot threshold, and *Io* is the receiver RSSI. *Io* can be expressed as

$$Io = N + IF_{Ext} + \sum_{k=1}^{4} \frac{\beta_k \cdot Ec_k}{\gamma_k^{PL}} = \eta + \sum_{k=1}^{4} \frac{\beta_k \cdot Ec_k}{\gamma_k^{PL}}$$
(3)

N denotes thermal noise,

 IF_{ext} denotes interference from other cells,

 β_k denotes the PA load fraction of the kth sector,

 γ_k^{PL} denotes the pilot power fraction of the kth sector.

Since only four cell sectors are modeled here, the interference contribution of surrounding cells has to be included too. Since we have no spatial distribution for it, we just add it to the thermal noise floor and treat both terms as one common background interference term η .

 Ec_{ik} is linked to the pilot EIRP, Pc_k , of server k via the path loss L_{ik} , which will be approximated by a power law

$$Pc_{k} = Ec_{ik} \cdot L_{ik}, \quad L_{ik} = r^{\kappa}/r_{0}$$

$$\tag{4}$$

Fading can be included within this model also. The results shown in the next sections, however, have been computed without fading. This allows the depiction of cell footprints and coverage in a more transparent fashion.

The PA load fraction β_k results from the traffic-channel load analysis and will be solved self-consistently for the entire network (see below.)

We further assume that any of the four pilots can enter the active set as soon as its pilot signal strength exceeds the same absolute threshold as introduced for the coverage condition: $(Ec/Io)_{ik} \ge \theta_{Pl}$. We neglect additional relative thresholds. These approximations are not very serious since the soft handoff areas in the given scenarios are rather small.

The dedicated traffic channel will be modeled via SIR estimation. The SIR at location i with respect to server k is

$$(S/I)_{ik} = S_{ik} / (Io - \alpha \cdot \beta_k \cdot Ec / \gamma_k^{PL})$$
(5)

where S_{ik} denotes the received traffic-channel power level, and α is an orthogonality factor with respect to same-cell channelization codes. In simplex mode, the power control algorithm gives $(S/I)_{ik} = 1/PC_{dwn}$, where PC_{dwn} denotes the downlink pole capacity for this service, including EbNo, processing gain and activity factor.

In soft handoff, we assume that the UE or mobile can perform maximum ratio combining, where the channel estimation is given by the pilot channel. The resulting power control equation is

$$(S/I)_{i} = \left(\sum_{k} \sqrt{S_{ik}} \sqrt{Ec_{ik}}\right)^{2} / \sum_{k} I_{ik} Ec_{ik} = 1 / PC_{dwn}$$
(6)

where the sum is taken over all pilots in the active set. To solve the power control equation in soft handoff, we assume

where

that all contributing traffic channel power levels, P_{ik} , are synchronized:

$$P_{ik} = S_{ik} \cdot L_{ik} = P_{il} \quad \forall l, k \tag{7}$$

Downlink traffic channel coverage is achieved as long as power control stays below the maximum bound: $P_{ik} \leq P_{max}$. In the following analyses, however, coverage is always limited by the common pilot channel.

The PA load fraction vector $\underline{\beta}$ can be obtained from the sum of pilot power and the power provided to all traffic channels, the latter weighted with respect to the local traffic density

$$\boldsymbol{\beta}_{k} = \left(Pc_{k} + \sum_{i} \frac{T_{i}}{T_{0}} \cdot P_{ik}\right) / Ptot_{k}$$
(8)

where $Ptot_k$ denotes the total PA power of the k^{th} sector. Here, the sum is taken over all locations that have active-set membership with respect to the k^{th} server.

During the course of the simulation, we slowly ramp up the total traffic T_0 , and for each value of T_0 , we self-consistently evaluate network coverage and the load vector, $\underline{\beta}$. We evaluate over a dense grid starting with common pilot channel coverage and traffic channel coverage based on initial default values for β . The resulting network coverage is given by

Network coverage =
$$\sum_{i} T_i \cdot Cov_i / T_0$$
 (9)

where $Cov_i = 1$, if location or grid point *i* has pilot and traffic-channel coverage, and zero otherwise. The resulting PA load fraction vector serves for the next iteration until convergence has been achieved. The T_0 -ramp-up is performed until one of the sectors reaches maximum load, i.e. $\beta_k = 1$. For all T_0 -steps, network capacity is set through:

Network Capacity =
$$T_0 \cdot Network Coverage$$
. (10)

IV. OPTIMIZATION AND OPTIMIZATION VARIABLES

We will optimize the model network for both traffic distributions and with respect to three different types of parameters:

- Mechanical antenna tilt
- Pilot power fraction (maximum PA power stays the same)
- Maximum PA power level (pilot power fraction stays the same).



Figure 3. Optimal tradeoff curves for a uniform traffic distribution with respect to adjustment of antenna tilt (diamond), pilot fraction (dot) and power level (triangle). The open square shows the performance of the initial configuration. The solid points represent configuration-based tradeoff sections and the open points interference-based tradeoff sections. On the top, the soft-handoff areas and corresponding network configurations are shown for three tradeoff points P1, P1', and P2.

For each of these parameters, we will determine the entire optimal tradeoff curve within a reasonable window for coverage and capacity. Since all three of these tuning parameters essentially cause an area breathing of the associated cell sector, we expect them to create similar configurationbased tradeoff curves. Since only four variables have to be considered at a time, the optimization can be done by bruteforce.

V. OPTIMIZATION FOR UNIFORM TRAFFIC

The optimization for uniform traffic starts out from an initial design, in which all cells are set to maximum power, the antennas are tilted so as to point to the center of the network and the pilot fraction is set to 10% of total power. The detailed set of parameters is summarized in Table 1. This initial network configuration is symmetrical and has a symmetrical traffic distribution, resulting in the network performance shown in Fig. 3 (open square). The resulting three optimal tradeoff curves (also plotted in Fig.3) have rather distinct shapes, setting them apart in their potential impact on optimization:

- The optimal tradeoff curve for cell-power adjustment only consists of the interference-based tradeoff section derived from the initial network configuration (triangles in Fig.3). This means, that power adjustment cannot further improve network performance.
- The optimal tradeoff for tilt adjustment consists of both configuration- and interference-based section (diamonds on Fig.3). Tilt adjustment can simultaneously improve coverage and capacity with

 TABLE I.
 PARAMETERS FOR MODEL NETWORK

Length of network layout:	Normalized to 1
Antenna height:	Based on 1.5° down angle to
	network center
Antenna vertical beam width:	6° , \cos^2 -pattern, no side lobes
Antenna back-lobe suppression:	25dB
Antenna horizontal beam width:	Uniform over 90°
Maximum EIRP per cell:	Set to 0dB received signal
	strength at network center
Ext. Interference:	12dB above max EIRP
	reference
Overhead channels:	Pilot only
Pilot threshold:	-15dB
Pathloss slope:	40 dB/decade
DL pole capacity per cell:	100
Orthogonality factor:	0.5
Initial power level:	Max power
Initial pilot fraction:	10% of total EIRP
Initial tilt:	1.5°

respect to the initial configurations. This is achieved by increasing all tilt values from 1.5° to 3° (Point T1 in Fig.3), and thereby reducing the mutual interference among all cells.

• The optimal tradeoff curve for pilot adjustment only consists of configuration-based tradeoff points (dots in Fig.3). Pilot adjustment therefore allows a wide range of performance tuning. The potential gains, however, are still small. Changing the pilot configuration P1 to P2 would increase coverage only by about 0.5% for same traffic load (compare P1' and P2). The corresponding coverage plots for these two configurations are also shown in Fig.3.

The small improvements achievable through tilt and pilot optimization are due to the fact that network traffic and network layout have the same symmetry, a situation which rarely occurs in real markets.



Figure 4. Load balancing via adjustment of the pilot fraction: The black squares show the optimal tradeoff curve for uniform traffic, the diamonds for non-uniform traffic. The capacity of configuration P1 deteriorates by 26% when the traffic changes from uniform to nonuniform (P1'). Load-balancing can retrieve a major fraction of this capacity loss (P10). On the top, the soft-handoff areas of configurations P1 and P10 are shown.

VI. OPTIMIZATION FOR NON-UNIFORM TRAFFIC

When the network traffic pattern is changed from a uniform distribution to a non-uniform one, the initial performance should degrade. This development can be confirmed by overlaying the interference-based tradeoff curves of the initial configuration for both traffic distributions (dash-dotted curves through points P1 and P1' in Fig.4). The corresponding capacity degradation is about 26%. Coverage also deteriorates, although only slightly, since the interference-based tradeoff curve for the non-uniform traffic distribution lies left of the original curve.

Network re-optimization using adjustment of pilot fraction produces a new optimal tradeoff curve (Fig.4, gray diamonds). A comparison to the old optimal tradeoff curve (black squares) shows the following:

- Both curves only consist of a configuration-based tradeoff section. However, the new optimal tradeoff curve lies below the old curve indicating that the load balancing operation has effectively lost performance.
- Choosing a configuration from the new curve (point P10) that has same coverage as the initial configuration for homogeneous traffic (point P1) results in a capacity loss of 7% (arrow between both points). This capacity loss, however, is rather modest when compared to the capacity loss of 26% without load balancing.

The overall performance loss of the load-balancing operation is due to a complex combination of power control, channelization code-orthogonality and cell size variation introduced by the optimization efforts. The latter has shifted the soft-handoff zone toward the heavily loaded cell. Off-loaded users, now fully supported by the neighbor cells, cause more interference to the congested cell since their channel power is not orthogonal anymore to its own channelization codes. The



Figure 5. Same plot as Fig. 4 but for optimization with respect to tilt. The capacity deterioration through change of traffic distribution is about the same as for optimization with respect to pilot fraction. The cell footprints and soft handoff areas for the re-optimized network, however, are reshaped.

shift of the soft handoff areas for configurations P1 and P10 is illustrated on the top of Fig 4.

We have repeated the same optimization procedure using antenna tilt instead of pilot fraction (Fig.5) with the initial configuration based on the tradeoff point T1. As in the prior case, the capacity loss caused by the change of the traffic distribution is about 26%. The new optimal tradeoff curve obtained through network re-optimization has a larger configuration-based section (Fig.5, gray diamonds) than the optimal tradeoff curve for the uniform traffic distribution before (black squares). The spatial separation between both tilt tradeoff curves, however, is larger than the separation between the optimal pilot tradeoff curves, indicating that pilot fraction can better regain performance through load balancing than tilt can. Choosing a new tilt configuration (Point T10) with same coverage as before shows an overall capacity decrease of 10% opposed to 7% when optimizing with respect to pilot fraction, which confirms this notion. The absolute capacity of both T10. however, is still 2.5% higher than that of P10.

Repeating the same procedure for cell-power level as adjustment parameter shows far smaller gains through load balancing than obtained form pilot and tilt adjustment (not shown). The overall capacity decrease after load balancing is around 19%, compared to 7-10% for pilot and tilt adjustment. This shows that adjustment of cell power levels is not very well suited for load-balancing operations.

Fading has been neglected in the results above. When lognormal shadow fading is included, the optimum tradeoff curves and the associated optimization configurations change only slightly. The largest impact can be observed for configurations with large coverage; especially the coverage value of 100% cannot be achieved anymore. The insensitivity to the particular choice of the fading model shows the general validity of the presented results.

VII. CONCLUSION

We have shown that load balancing operations for 3G cellular networks need to consider the tradeoff between network coverage and capacity. As a result, performance improvements through network reconfiguration always have to be measured against the CDMA-intrinsic coverage-capacity tradeoff through interference. Within this framework, it could be shown that load balancing through the shifting of cell boundaries can improve network performance substantially. This performance improvement, however, sensitively depends on the particular tuning parameter chosen. We found that especially power-adjustments are pretty ineffective while pilot and tilt tuning can achieve good results. Further investigation including reverse link and packet data servers are underway.

REFERENCES

- T. Ojanperä and R. Prasad, "WCDMA: Towards IP mobility and mobile internet", Artech House Publishers, 2001.
- [2] H. Holma and A. Toskala, "WCDMA for UMTS: Radio access for third generation mobile communications", John Wiley & Sons, Ltd, 2000.
- [3] S.T.S. Chia, "Design and optimization for cellular access network", Electron.& Commun. Eng. J., vol. 8, no.6, pp.269-277, Dec. 1996.

- [4] S. Hurley, "Planning effective cellular mobile radio networks", IEEE Trans. Veh. Technol., vol.51, no.2, March 2002.
- [5] C. Wheatley, "Trading coverage for capacity in cellular systems: A systems perspective," Microwave J., vol. 38, no. 7, pp. 62-76, July 1995
- [6] V.V. Veeravalli and A. Sendonaris, "The coverage-capacity tradeoff in cellular CDMA systems", IEEE Trans. Veh. Technol., Vol. 48, no.5, pp1443-1450, Sept. 1999
- [7] Q. Hao, B.-H. Soong, E. Gunawan, J.-T. Ong, C.-B. Soh, and Z. Li, "A low-cost cellular mobile communication system: A hierarchical optimization network resource planning approach", in IEEE J. Select. Areas Commun., vol. 15, no. 7, pp. 1315-1326, Sept. 1997.
- [8] C. Y. Lee and H. G. Kang, "Cell planning with capacity expansion in mobile communications: A tabu search approach", in IEEE Trans. Veh. Technol., vol.49, no.5, pp. 1678-1691, Sept.2000.
- [9] S. K. Das, S. K. Sen and R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment", Wireless Networks, vol.3, pp. 333-347, Aug. 1997.
- [10] S. K. Das, S. K. Sen and R. Jayaram, "A novel load balancing scheme for the tele-trffic hot spot problem in cellular networks", Wireless Networks, vol. 4, pp.325-340, July 98.
- [11] C. Y. Lee, "A dynamic sectorization of microcells for balanced traffic in CDMA: Genetic algorithms approach", IEEE Trans. Veh. Technol., vol.51, no.1, Jan 2002.
- [12] K. Takeo and S. Sato, "Evaluation of a CDMA cell design algorithm considering non-uniformity of traffic and base station locations", IEICE Trans. Fundamentals, vol. E81, no.7, pp.1367-1377, July 1998.
- [13] D. H. Kim, D. D. Lee, H. J. Kim, and K. C. Whang, "Capacity analysis of macro/mircocellular CDMA with power ratio control and tilted antenna", IEEE Trans. Veh. Technol, vol. 49, no. 1, Jan. 2000.
- [14] R. G. Akl, M. V. Hegde, M. Naraghi-Pour and P. S. Min, "Multicell CDMA network design", IEEE Trans. Veh. Technol., vol. 50, no. 4, pp. 711-722, May 2001.
- [15] J.-S. Wu, J.-K. Chung and C.-C. Wen, "Hot-spot traffic relief with a tilted antenna in CDMA cellular networks", IEEE Trans. Veh. Technol, vol. 47, no. 1, pp. 1-9, Feb. 1998.
- [16] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, and P. Larsson, "Future-Generation Wireless Networks", IEEE Pers. Commun., vol. 8, no. 5, pp. 10-17, Oct. 2001.