♦ Dynamic Optimization in Future Cellular Networks

Simon C. Borst, Arumugam Buvaneswari, Lawrence M. Drabeck, Michael J. Flanagan, John M. Graybeal, Georg K. Hampel, Mark Haner, William M. MacDonald, Paul A. Polakos, George Rittenhouse, Iraj Saniee, Alan Weiss, and Philip A. Whiting

With multiple air-interface support capabilities and higher cell densities, future cellular networks will offer a diverse spectrum of user services. The resulting dynamics in traffic load and resource demand will challenge present control loop algorithms. In addition, frequent upgrades in the network infrastructure will substantially increase the network operation costs if done using current optimization methodology. This motivates the development of dynamic control algorithms that can automatically adjust the network to changes in both traffic and network conditions and autonomously adapt when new cells are added to the system. Bell Labs is pursuing efforts to realize such algorithms with research on near-term approaches that benefit present third-generation (3G) systems and the development of control features for future networks that perform dynamic parameter adjustment across protocol layers. In this paper, we describe the development of conceptual approaches, algorithms, modeling, simulation, and real-time measurements that provide the foundation for future dynamic network optimization techniques.

Introduction

Cellular network optimization has traditionally been associated with a process that aims to adjust the network air interface to market-specific traffic and propagation conditions [11, 12]. The tuning operations in this process focus on a small set of hardware parameters such as cell-site locations and antenna configurations. The cells themselves are grouped more densely around traffic nucleation points to provide capacity, and the antennas are pointed in compliance with the local terrain and clutter to reduce signal shadows and interference. Occasionally, software parameters such as handoff thresholds and cell-power budgets are also adjusted. While hardware parameters are easily set during network installation, they are hard to change afterward. As a result, the optimization with respect to hardware parameters occurs during network planning and deployment, and it is only repeated in areas where performance problems or infrastructure upgrades are required. Since it is performed as a singular event, the optimization process is

Bell Labs Technical Journal 10(2), 99–119 (2005) © 2005 Lucent Technologies Inc. Published by Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com). • DOI: 10.1002/bltj.20096

fundamentally based upon time-averaged worst-case traffic and propagation conditions. Originally, these optimizations were performed through a manual, iterative process relying on network planning tools and drive testing. Bell Labs subsequently introduced the concept of predictive optimization in the Ocelot® optimization tool, which computes optimum network parameters directly according to well-defined performance metrics [5, 7, 8]. Its introduction has translated into faster network rollouts, improved network performance, and higher capacity.

Notwithstanding its widespread use, the static optimization approach is increasingly approaching its limits. The growth of the wireless customer base and the introduction of various new data services mandate the consideration of new objectives such as throughput, delay, latency, and quality of service (QoS). Indeed, the migration to IP Multimedia Subsystems (IMS) will promote the continuous development of new services with different resource requirements, QoS demands, and traffic characteristics. Furthermore, data services introduce demand fluctuations that are intrinsically larger than they are for voice services. The multidimensional nature of demand, its temporal dependence, and its increased dynamic range render optimization strategies based on a peak (albeit composite) loading progressively less effective at efficiently allocating and managing network resources. Additionally, the demand for increasing data rates and the falling costs for network hardware will drive network architectures toward micro-cellular structures. This development will create frequent infrastructure upgrades with the demand for fast, autonomous, and inexpensive cell integration.

Demand fluctuations have widely different origins, correlations, and characteristic times, with important implications for how best to address them. Channel fluctuations (i.e., fast and shadow fading) for individual users typically range from milliseconds to tens of seconds. Arrival, departure, and handoff/swap times for individual voice and data users typically inhabit the range between a few seconds to tens of minutes. When looking at aggregate demand variations, one observes that they typically occur on an interval from intermediate to long-time scales (i.e., seconds and above). One further observes the coexistence of both predictable (periodic) and unpredictable (random) variations in aggregate demand. Predictable variations possess characteristic times of order hours, days, and weeks, and they can possess intercell correlations. Random fluctuations occur on practically all time scales, but because of intrinsic averaging, they tend to be most pronounced at shorter time scales. They are typically uncorrelated between cells. Similar characterization occurs when one looks at spatial variations of the users within and between cells, which also contribute to demand fluctuations. (See **Figure 1**, which displays characteristic time scales for network changes and fluctuations on a logarithmic time axis spanning 12 decades. Individual per-mobile changes [upper left] occur on faster time scales. Predictable aggregate demand variations [upper right] typically dominate on longer time scales. Existing per-mobile cell controls [lower left] naturally have time scales comparable to per-mobile variations. Present manual optimization efforts to adapt the network [lower right] clearly cannot address fluctuations occurring on shorter time scales. Hourly service measurements occur in the middle and thus are handicapped in observing behavior at shorter times.)

Panel 1. Abbreviations, Acronyms, and Terms

1xEV-DO—CDMA2000* evolution—data optimized
2G—Second generation
3G—Third generation
3G1X—3G CDMA air-interface specification; CDMA2000 first evolution
CDMA—Code division multiple access
HSDPA—High-speed downlink packet access
IFHO—Inter-frequency handoff
IMS—IP Multimedia Subsystem
IP—Internet Protocol
LAN—Local area network
LP—Linear program
PA—Power amplifier
QoS—Quality of service
SINR—Signal-to-interference-plus-noise ratio
UMTS*—Universal Mobile Telecommunications System; 3G wideband-CDMA air-interface specification

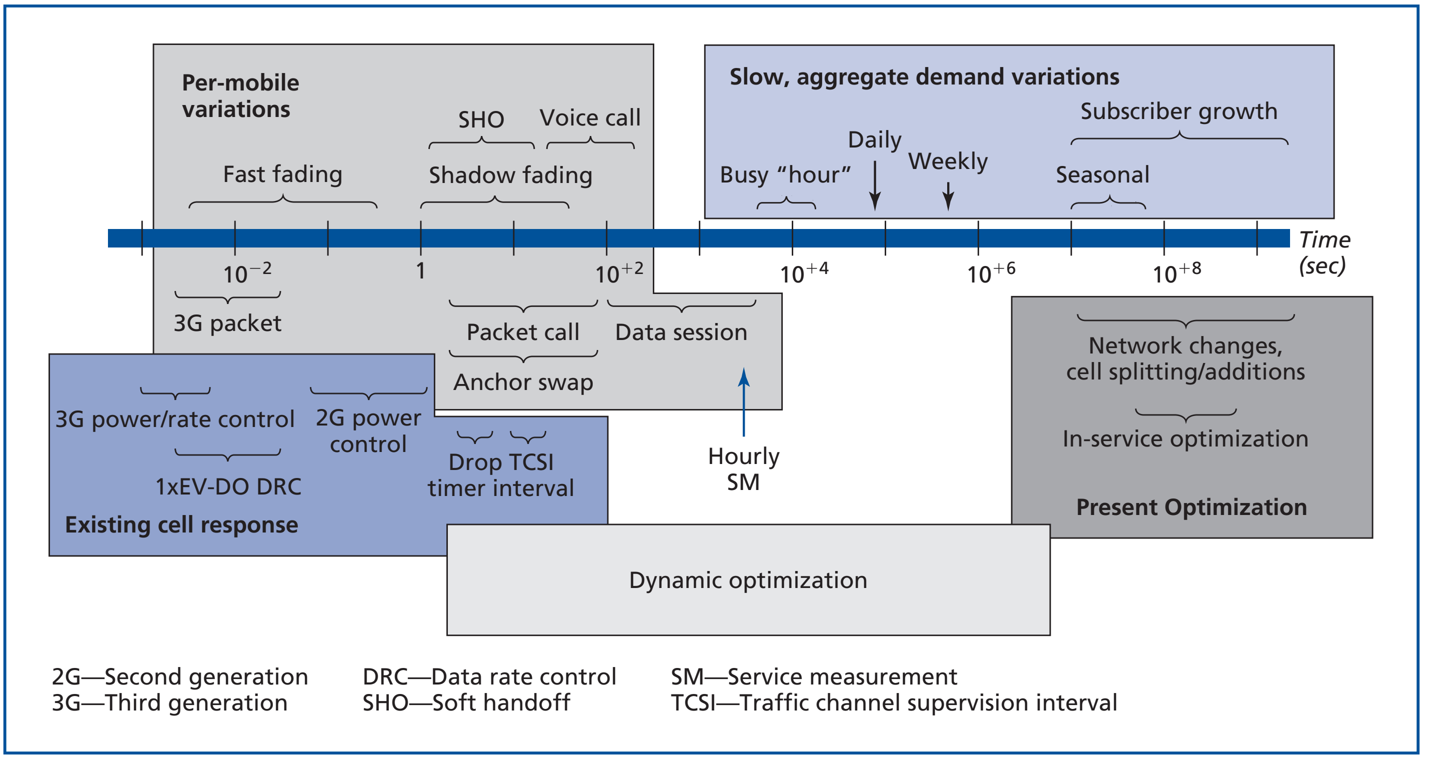


Figure 1.
Characteristic time scales for network changes and fluctuations displayed on a logarithmic time axis spanning 12 decades.

Cellular networks have always been supported by a set of fast control algorithms that aim to account for the dynamics introduced by variations in channel conditions and traffic loading through adjustment of cell-, service-, or protocol-layer-specific parameters. Examples are power control, handover, scheduling, and congestion control algorithms. With the evolution of wireless standards, the response times of these control algorithms have continuously been improved. One key point is that the existing controls can have intrinsic limitations that limit their overall network response options to certain types of demand variations. At a high level, one can group most second-generation (2G)/third-generation (3G) control mechanisms as autonomous control algorithms of two types: autonomous per-call controls, such as power control and soft-handoff, and autonomous aggregate controls, typically on a per-cell-sector basis (e.g., overload and congestion controls). By comparison, few existing control mechanisms work in a coordinated manner across neighboring cell sectors or the network.

As a consequence, we see a growing need for additional dynamic optimization mechanisms with the following capabilities:

- State- and time-dependent control parameters to help the network adapt the coverage and capacity tradeoff for multiple services in response to spatio-temporal demand variations;
- Coordinated (and potentially additional autonomous) load-balancing mechanisms that can address demand and traffic fluctuations by optimally "smoothing out" uncorrelated demand peaks between neighboring cells and even between differing wireless technologies; and
- Active measures to address rare but undesirable events, such as reducing dropped and blocked calls.

For these controls to safely coexist with the existing per-user network control mechanisms (e.g., power control) while not overburdening call processing capabilities, we anticipate that many of these new mechanisms will act at intermediate time scales (i.e., time scales of seconds and up).

These features will fill the gap between present control algorithms and the current network optimization process (potentially overlapping at either end) regarding operation space, time scale, adjustment-parameter space, and performance objectives. This will translate into benefits for service providers and mobile users such as improved coverage, fewer dropped calls, better QoS, and higher throughput. Service providers will also benefit from reduced maintenance and operation costs, the ability to capitalize from higher network capacity, and quicker launch of new services.

From the research perspective, the development of these dynamic optimization features follows a roadmap that combines short-term aspects that improve the performance of present markets and longer-term strategies. This allows the experiences gained from early market trials to refine algorithms planned for later releases. This incremental development also permits sequential rollout of these techniques into the marketplace for more immediate benefits to both service providers and mobile users.

In this paper, we introduce several facets of Bell Labs' current research efforts in the area of dynamic network optimization. These illustrate the complexity and multidisciplinary nature of the underlying research. In the section immediately below, we describe the near-term research effort to establish dynamic optimization capabilities in current networks. The approach capitalizes on automated algorithms for the traditional network optimization process developed specifically to operate within the constraints of current network-measurement reporting capabilities. A subsequent section examines the necessity and significance of real-time network measurements in the development of dynamic optimization strategies and algorithms. Such measurements provide detailed data and invaluable insights on traffic fluctuations, user mobility, data traffic models, and network response and performance characteristics under actual user conditions. The succeeding section presents an example of a dynamic optimization algorithm for future cellular networks as well as comparative performance results based on dynamic simulations. The last section concludes with a summary of our current efforts in this area.

Dynamic Optimization in Current Networks

Historically, network optimization efforts have been highly static, and hence a "worst-case" design view has been dominant. The time constants of the individual network changes have been on the order of years because subsequent antenna changes are expensive. Further, to simplify network operations, parameter uniformity (e.g., pilot fractions and hand-off parameters) has been emphasized throughout the entire network.

The above ideology is still acceptable to a service provider in the earliest days of the network launch; it is also currently acceptable if there is sufficient capital and spectrum for carrier additions and sufficient capital and site availability for cell additions. With the increasing maturity of networks, however, limited spectrum, site scarcity, and limited available capital for network expansion make it necessary to employ more sophisticated network optimization techniques.

As discussed in the introduction, automation of the current static optimization process is the first step in this direction. This approach has already proven indispensable in network rollout and for re-optimization efforts of 2G and 3G technologies. It has created the fundamental understanding on how to realize algorithm-guided adjustment of large network clusters capturing the complex interdependence between network performance and the large number of tuning parameters through the specific traffic distribution and propagation conditions. Such automation has also shown the possibility of better performance faster. Although this approach still has to be considered "static," it sets the basis for the near-term realization of dynamic optimization [1][2].

The dynamic optimization effort described in this section involves changes to a relatively small number of existing control parameters in time scales on the order of hours (as set by current performance reporting mechanisms) that can be realized as a near-term network product feature.

Hourly time-scale variations in a customer market are shown in **Figure 2** (where the x-axis shows the measurement of time in a 24-hour format and the y-axis shows the resulting time-averaged traffic). This motivates the need for a "many-case" design approach

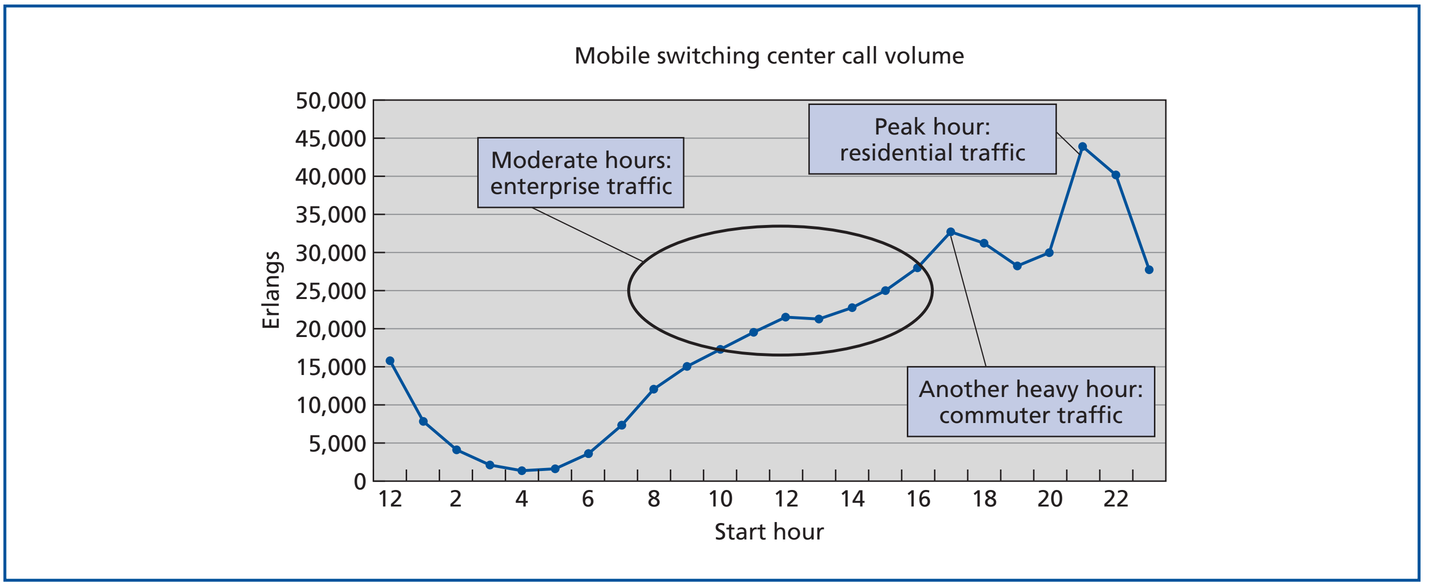


Figure 2.
Call volume of a time-varying network.

instead of the present worst-case design approach. The many-case design scenarios are in part differentiated by varying aggregate call volumes and in part by changes in traffic distributions (e.g., users at home between 8:00 p.m. and 8:00 a.m., at work between 8:00 a.m. and 5:00 p.m., commuting between 5:00 p.m. and 8:00 p.m.).

Since coverage is one of the major drivers of churn in a market, the objective of near-term dynamic optimization is to load a network configuration for each hour of the day (or each collection of hours) that both maximizes coverage and satisfies the estimated traffic demand (based on the analysis of recent service measurement data).

High-Level Architecture and Implementation Details

This subsection addresses the high-level architecture and the changes that need to occur in order to realize the dynamic optimization feature. The expected impacts on architecture and implementation are discussed in detail. A high-level view of the near-term dynamic optimization can be viewed as having two parts: open loop and closed loop.

Figure 3 shows the typical setup of the current network monitoring architecture with an external user interface to the network (to control, e.g., cell-power budget and other radio resource management

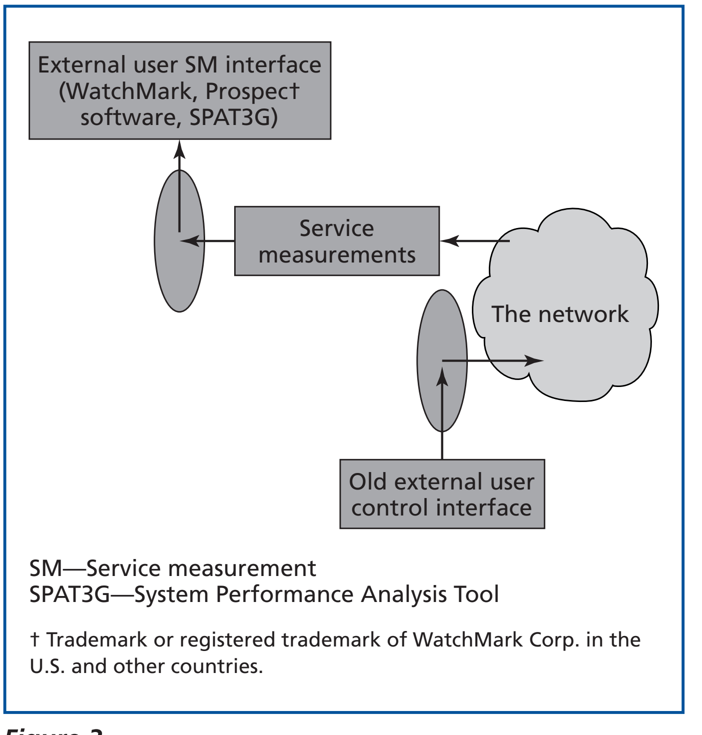


Figure 3.
Current network monitoring architecture.

[RRM] parameters). These control parameters can be viewed to define a particular network configuration where all other network aspects (e.g., base station locations and antenna orientations) remain constant.

In response to the specified network configuration, offered traffic and other network aspects (e.g., propagation conditions), the network generates service measurement data that is then filtered using the external user service measurement interface. A more detailed description of the network performance and conditions can be obtained locally using other tools and techniques as described in the "Real-Time Measurements" section of this paper. In this current mode of operation, an operator monitors the external user service-measurement interface and makes changes via the external user control interface. In practice, any single control parameter is rarely changed. Because of this, the specified network configuration is effectively "one size fits all" and is expected to serve the network 24 hours a day in spite of the well-known variations in offered traffic across this time scale.

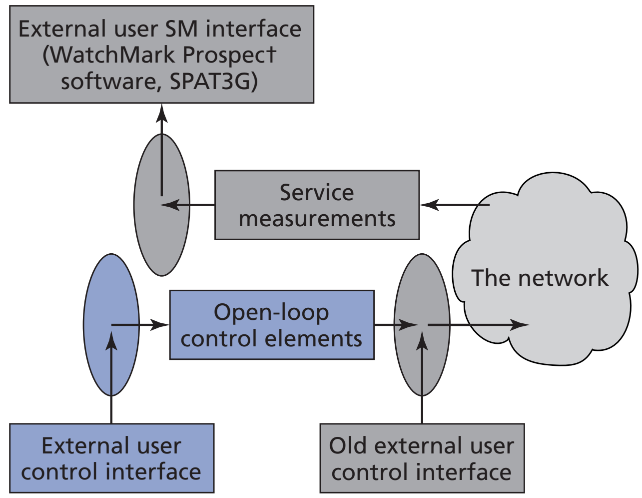
Open-Loop Dynamic Optimization

The open-loop mode of operation is a natural step between the current operator control and the more aggressive closed-loop control (which greatly reduces operator involvement). This mode of operation specifically addresses variations in network conditions and permits "many-case" network design. The operation of the open loop element can be best described using a pre-programmed approach: the ability to automatically switch to different settings during different times of the day based on a set of well-known and predictable inputs. In a similar manner, the input to open loop dynamic optimization would be a set of operator-prescribed network settings for different hours of the day/days of the week. Based on this input, the new network settings are automatically loaded during the corresponding hours of the day and days of the week.

The control elements that can be changed in the open-loop mode of operation include:

- Cell-power budget,
- RRM parameters,
- Neighbor lists,
- Inter-frequency handoff (IFHO) parameters, and
- Electronically controllable antennas (to adjust tilt and/or azimuth).

Some of the advantages of open loop dynamic optimization are that it:



SM—Service measurement
SPAT3G—System Performance Analysis Tool

† Trademark or registered trademark of WatchMark Corp. in the U.S. and other countries.

Figure 4.
Network control architecture for open-loop dynamic optimization.

- Facilitates preliminary testing of the full closed-loop system,
- Allows for an intermediate realization of the near-term dynamic optimization features,
- Avoids stability issues, and
- Permits configuration management and integrated operation (unlike scripts and cron jobs)

Figure 4 shows the network control architecture for open-loop dynamic optimization. Notice that the "old" external user control interface is replaced with a new external user interface that provides input to the open-loop control elements with the recommended values for cell-power budget, RRM parameters, neighbor lists, IFHO parameters, and electronically controllable antennas. The new interface also permits specification of start times for different network configurations (e.g., 4:28 p.m. on Wednesdays).

Closed-Loop Dynamic Optimization

Figure 5 shows the architecture of closed-loop dynamic optimization. The only additional element

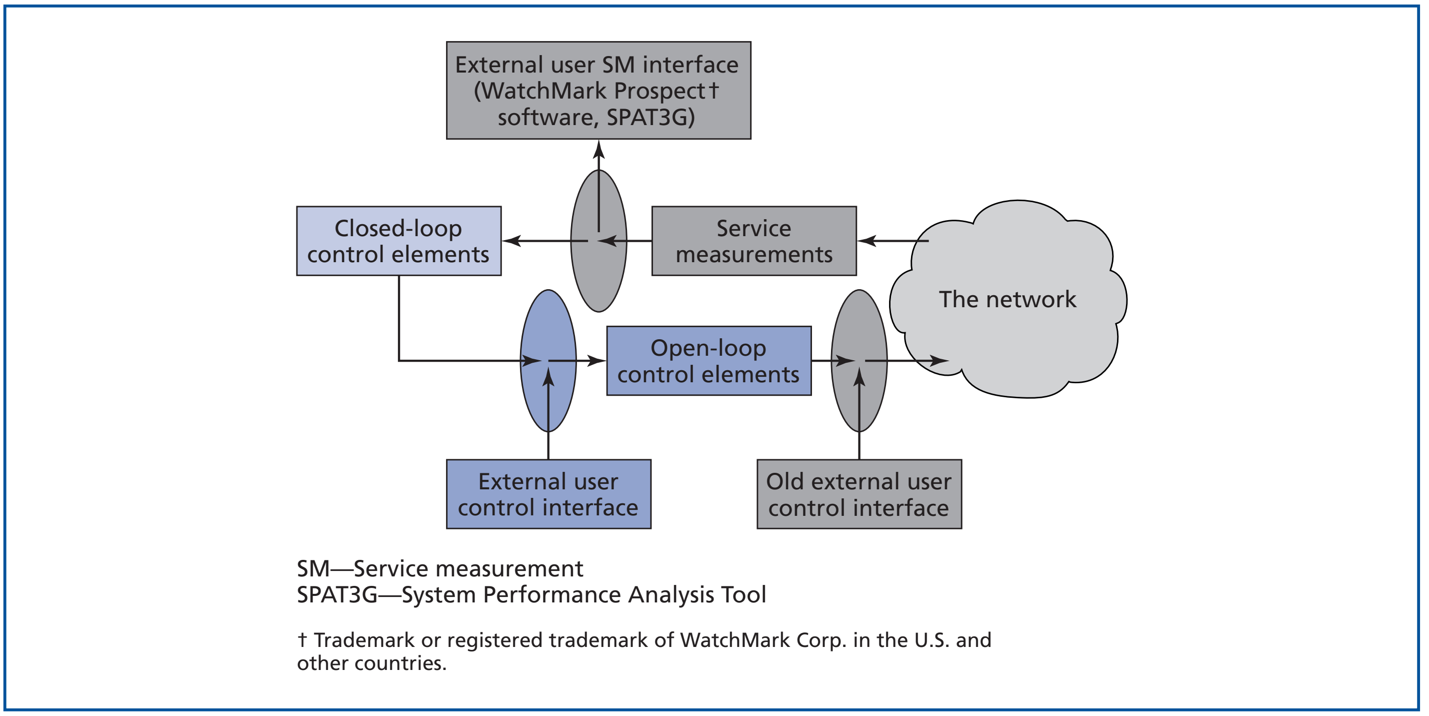


Figure 5.
Architecture of closed-loop dynamic optimization.

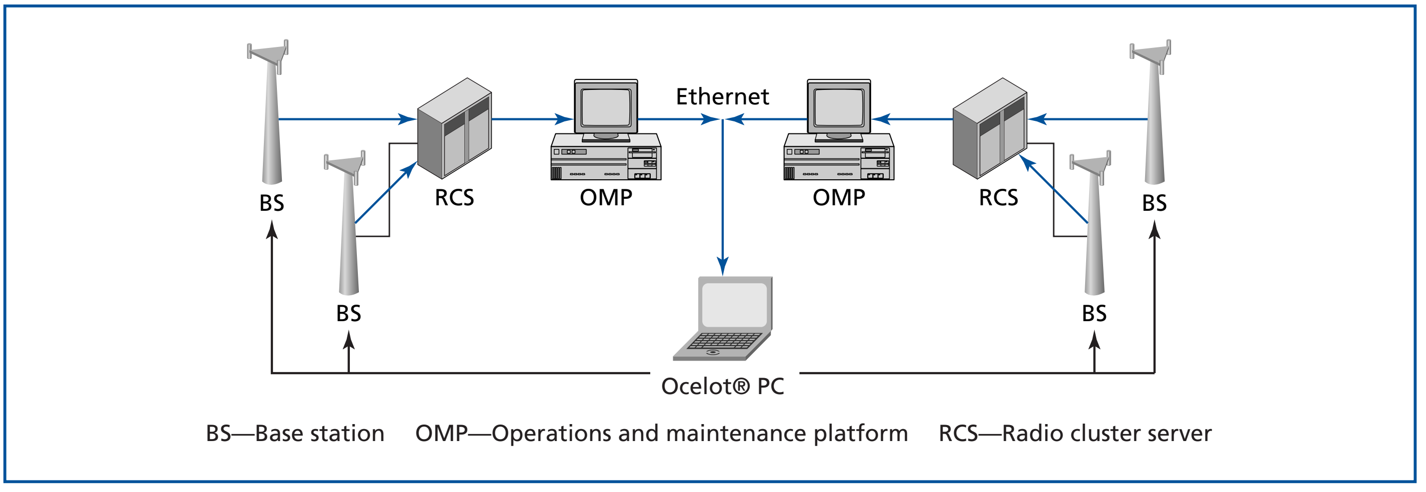


Figure 6.
High-level physical architecture of closed-loop dynamic optimization.

beyond open-loop dynamic optimization is the feedback loop that interfaces with the closed-loop control elements. The input to the closed-loop control elements is a sliding window of service measurement data. This window ranges from some point in the reasonably recent past (e.g., several weeks) to upwards of minutes ago. The use of older data allows longer-term trend assessment as well as mitigation of short-term fluctuations in the input data. The use of very recent data facilitates responses to dramatic network events. The output of the closed-loop control elements identifies changes to control parameters including cell-power budget, RRM parameters, cell neighbor lists, IFHO parameters and electronically controllable antenna parameters.

Figure 6 shows a high-level physical architecture of the closed-loop dynamic optimization system. For each optimization period, the autonomous Ocelot

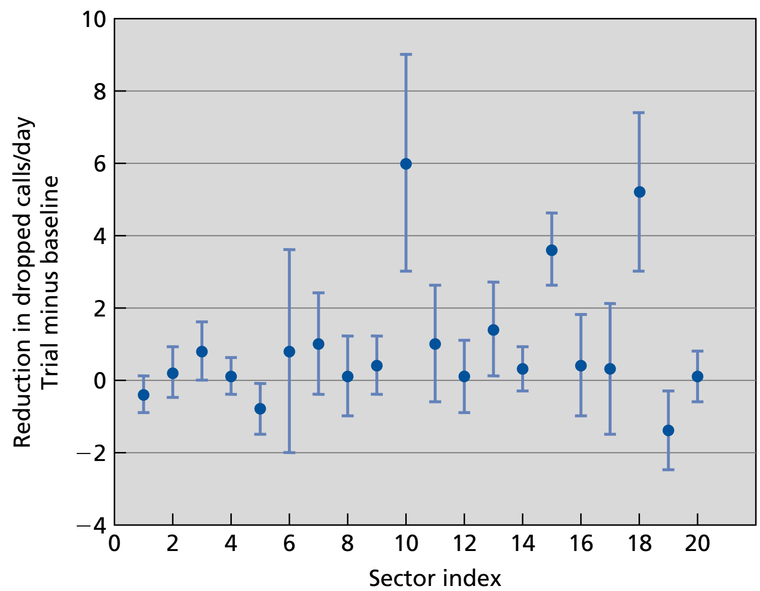


Figure 7.
Reduction in dropped calls per day.

software would perform an optimization and then choose a network solution that would maximize coverage and still meet the expected capacity demands. The blue arrows shown in Figure 6 indicate the service measurement input that is fed back to the Ocelot software module. The output from this module is then sent to the base station (shown conceptually by the black arrows in Figure 6).

Initial Trial Results

Field trials have been conducted using the open-loop optimization approach to evaluate near-term dynamic optimization methodologies and performance improvements. An example of performance improvements (measured by dropped calls) is shown in **Figure 7**. Each point on the graph represents the average and uncertainty in the reduction of dropped calls for each sector measured.

The change in the number of daily dropped calls (trial minus baseline) is shown as a function of sector index along with measurement uncertainties. Sector 10 experienced 6 fewer drops per day during the open-loop dynamic optimization trial. Sectors 15 and 18 also revealed significant improvements. For the entire cluster under study, the dropped call rate and blocked call rates improved by approximately 10%. The dynamic optimization experiment also resulted in improved frame-error rates and increased served traffic.

Real-Time Measurements

The operation of modern cellular networks is extremely complex. In order to extract the highest possible performance from the network while simultaneously avoiding any associated instabilities, dynamic optimization algorithms must be based upon a deep and accurate understanding of actual mobile and network behavior. Detailed measurements on operating networks are critical to building this understanding for several reasons. First, they provide a mechanism for quantifying important user behavior and patterns. Such data consequently constrains the parameter space that has to be explored with simulations, thereby avoiding unnecessary and/or erroneous assumptions regarding user and system behavior. Second, detailed network measurements provide a powerful tool for accurately quantifying underlying relationships between user QoS/performance and network parameters. Third, network measurements provide a means for discovering unrecognized phenomena, problems, and inefficiencies resulting from the interplay between the multiple users and the network.

Traditional network service measurements are typically binned or accrued over an hourly basis. Such service measurements are well suited to performance monitoring situations where the relevant quantity is deterministic and readily measurable, often associated with a particular network sub-element. An example might be the peak number of Walsh codes in use during the hour. As long as this peak demand does not exceed the maximum number of available codes, one can safely conclude that no performance degradation directly resulted. However, the dominant fraction of wireless network response occurs on time scales much finer than those captured by traditional service measurements (see Figure 1). Hence a deeper understanding of actual system behavior often requires much more detailed information than service measurements can supply. In order to draw statistically valid conclusions, one must correlate measurements of individual mobiles as well as network elements upon very fine time scales (often $<$1 second). Two examples include:

- Understanding and managing the interactions between simultaneous voice and data users on the same carrier, and
- Gaining a deeper statistical understanding regarding the typical confluence of events that generate dropped calls.

We have found that excellent results can be obtained by collecting detailed network performance data upon actual user mobiles, rather than from intentionally introduced drive test mobiles. This technique generates statistically valid data sets far faster than traditional drive testing, as there are large numbers of simultaneously active user mobiles in the market. Connection to a single cell can provide detailed information on several tens of thousands of calls per day. Perhaps more surprisingly, we have observed that measuring actual user mobiles can in some circumstances be more accurate in that they capture actual user behavior such as mobility mix, traffic distributions, building penetration mix, and data traffic models.

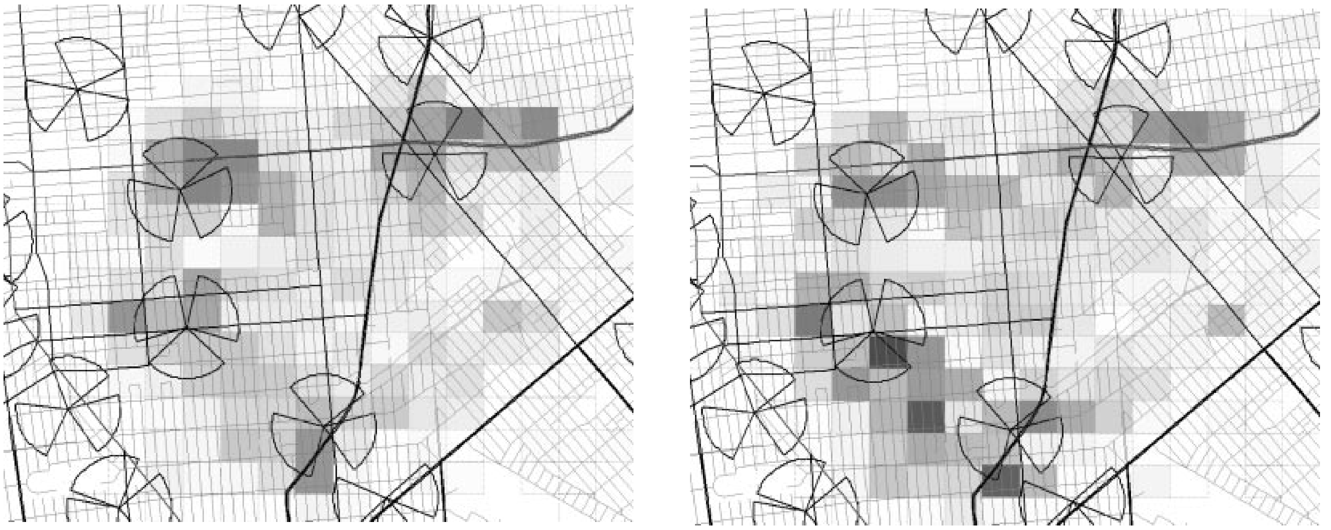
We have developed a flexible cellular measurement tool named Celnet Xplorer [3] that spans the full range of relevant time scales and operates with negligible impact upon a fully loaded network. Our experiments have focused upon several issues, including:

- Measurements of the spatio-temporal fluctuations in user traffic distributions and their impact upon network performance.
- Statistical data regarding user soft handoff and data anchor behavior, including how these may be influenced by system behavior.
- Understanding the primary sources of lost calls, with a sharp focus upon identifying possible measures that could be taken to reduce call failures.
- User traffic behavior, particularly for data users. A key focus is identifying steps that could be taken to enhance data performance.

Figures 8 and **9** display geolocation of user mobile density and lost calls in a cellular system via measurements made by Celnet Xplorer. Network measurements allow us to simultaneously track all user calls on the cluster with an average accuracy of ~150–200 meters. The figures show the base stations as “pie slices” with a common vertex. Each pie slice represents a particular sector direction. The road map is also shown. Figures 8a and 8b present 10 minutes of data aggregated into 250 m spatial bins, with color intensity proportional to the total user connection time per spatial bin. Substantial traffic density variations appear even on intermediate time scales.

Abnormally terminated calls are also observed and geolocated, and the histories of those calls as well as the associated network state are carefully analyzed. In Figure 9, for example, we observe that the dropped call spatial density is quite different from the user density—i.e., the dropped call likelihood has significant spatial dependence (in that the majority of these drops are in the interior of the cluster, near the cell boundaries). Taken together with the local propagation conditions, this information provides insight into the specific user environment prone to abnormal call termination in this cluster.

Another relevant area of user behavior is packet-data traffic models for mobile cellular users. While



(a) Measured mobile density: 7:20–7:30 p.m.

(b) Measured mobile density: 7:30–7:40 p.m.

Figure 8.
Measured mobile density of data aggregated in 250 m spatial bins.

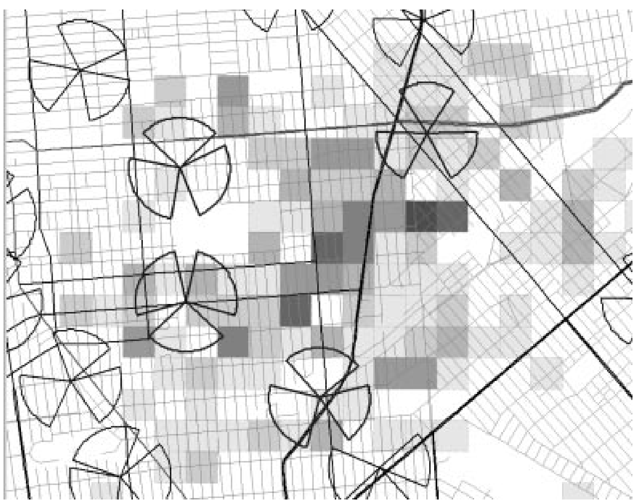


Figure 9.
Total dropped calls collected over 12 hours of data, again aggregated into 250 m spatial bins (monitored cells in bold).

extensive packet data traffic models exist in the literature, the vast majority of these are based upon measurements of TCP-IP traffic between computers connected via hard-wired LANs. However, the statistical character of cellular data traffic can be very different, particularly when viewed at layer 1. Other potential sources for changes in packet data traffic characteristics include different application mixes for wireless users, particularly those users with hand-held mobile phones with small screens. The measured supplemental channel activity for two simultaneous 3G1X data users is shown in **Figure 10**. Interactions between these users are discernible in their traffic patterns.

As our last example of the power of detailed network measurements, we consider the impact of forward-link transmit power upon abnormal call terminations (drops). Here we have associated user mobile connection time against their strongest serving cell sector (as seen by the mobile) and simultaneously recorded the total forward-link transmit power of that server versus time. If a call is dropped, we similarly note the total transmit power on the strongest serving sector. From these data, we can infer a drop probability per unit user-mobile connection time (i.e., the probability that the call will drop in the next second) as a function of the total forward power on the strongest serving sector. The result of one such measurement is displayed in **Figure 11**, which quantitatively demonstrates the role of increased interference upon dropped call rates. The positive slope displays

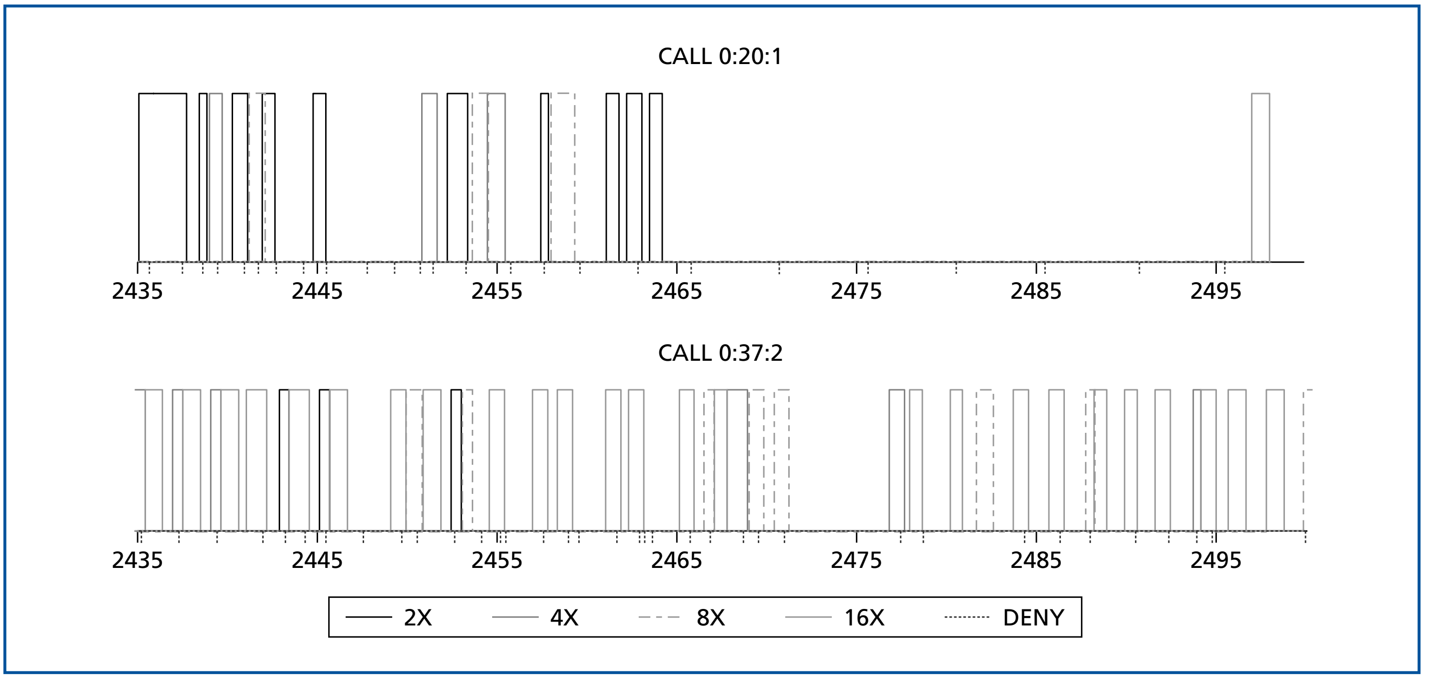


Figure 10.
3G1X supplemental channel activity for two simultaneously active data users on a particular cell sector.

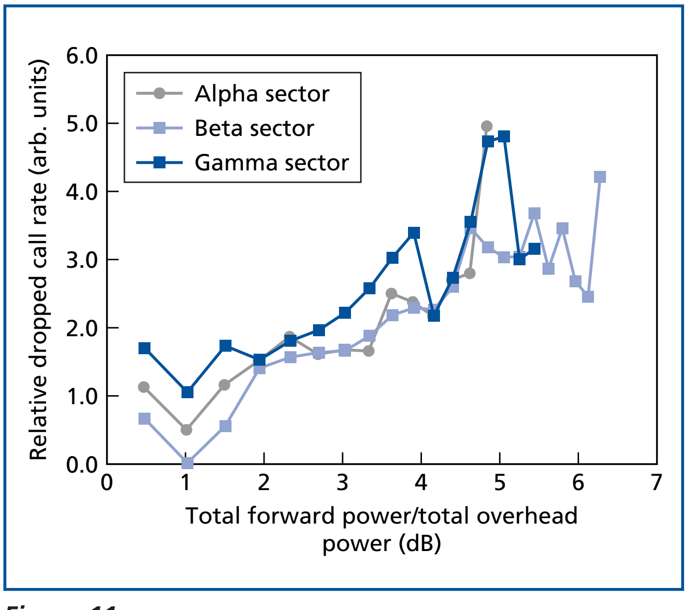


Figure 11.
Relative drop probability per unit connection time as a function of the total forward power on the strongest serving sector.

the impact of forward loading. These results were inferred from five days of measurement upon a specific cell in an actual customer market. The potential relevance of such results to load-balancing algorithms (such as discussed in the next section) is clear.

Control Algorithms for Future Networks

One class of dynamic control mechanisms in future networks defines algorithms that respond to fluctuations in both traffic load and channel conditions while respecting network-wide performance objectives. For the development of this class of control mechanisms, it is important to understand the origins of fluctuations in traffic load and channel conditions, their impact on network resources, and the advantages and shortcomings of present mechanisms that were not designed to tackle dynamic phenomena comprehensively.

For example, fluctuations in traffic load and channel conditions translate into fluctuations in the demand for resources, which in turn affect network and/or cell capacity and QoS due to the limitation of resource supplies. Current networks provide a limited set of control algorithms that attempt to distribute the available resources appropriately. They reside at different protocol layers and network elements and perform operations locally. Examples for such control algorithms are power control, scheduling, handover, channel assignment, and congestion control.

Local control routines have the advantage that they can react quickly. Speeding up the existing control mechanisms has led to significant performance improvements as seen, for instance, in the upgrade of downlink power control from cdmaOne* to CDMA2000*, or in the technology evolution for wireless packet data services from UMTS* R99 to high-speed downlink packet access (HSDPA). In the most recent technology upgrades, however, the speed of power and rate control has reached the spectral limit of channel fluctuations and further acceleration would promise only little improvement.

Handling load and channel fluctuations through local algorithms risks missing out on global performance goals (i.e., at the cluster or network level). A mobile utilizing a HSDPA technology interface, for instance, may be well served through fast rate control combined with scheduling by its anchor cell. From the network perspective, however, it may be better to serve this user with a slightly lower rate by a peer that carries substantially less load and can therefore provide more time slots. A user enjoying a conversational service at high speed may suffer from performance reductions associated with the high handoff rate and the lack of soft-handoff robustness, which is not supported on the HSDPA traffic channel. In such a case, a dedicated channel may be better suited to achieve call performance criteria.

The realization of resource utilization according to cluster- or network-performance objectives is a complex task, which may involve multiple parameters and performance objectives as well as additional messaging. Furthermore, the distribution of performance measurements and adjustments over various network nodes makes it difficult to develop features that can react sufficiently fast to follow the desired fluctuations. Accurately estimating the potential gains is quite challenging in view of the network complexity and the need for underlying assumptions. In order to overcome these challenges, a researcher needs to follow a systematic methodology that involves mathematical modeling and simulation, based upon real-time measurements of actual traffic behavior and network response as illustrated in the previous section.

There have been various publications that demonstrate the potential performance gains through dynamic optimization to network-wide performance objectives [1, 4]. They all focus on particular optimization tasks involving a predefined set of parameters and objectives. Often, the problem of implementation is not addressed, or a centralized control system that manages intercell or cross-layer adjustment in a coordinated fashion is assumed.

To quantify the benefits of resource allocation according to network-wide performance objectives, we present an example methodology for dynamic optimization algorithm development.

A Methodology for Algorithm Development

An example methodology we have used follows three discrete steps:

- *Formulation of the optimization task.* This addresses the determination of the adjustment parameters and network-wide objectives. Objectives need to be selected carefully since they critically determine the gain and relevance of the proposed mechanisms in the final implementation.
- *Comparative simulation studies.* A mathematical model is developed for the network objectives and with respect to the particular technologies of interest. This step allows an evaluation via simulation of the potential gains that are achievable through network optimization as compared to the (best) performance of current systems. The latter is obtained through detailed dynamic simulations that can capture the network-related aspect of the objective being optimized. Ideally, the modeling should capture the relevant time scale of operation for the optimization process, which should represent an optimistic limit to what is achievable under best circumstances in practical implementations.
- *Feature development.* Feature development can be guided by the knowledge gained from the mathematical modeling and simulation stages. For example, a centralized optimization solution may be realized in a distributed form only, which allows approximating the optimum solution without excessive communication. Alternatively, good candidate algorithms can be “guessed” based on observable patterns in the optimum solution based on engineering experience. In both cases, the outcome must be ranked against the bounds

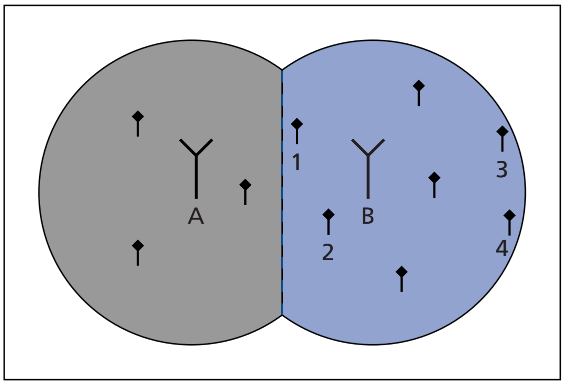
obtained in the comparative evaluation process, which helps in determining the expected gain from feature development.

The uplink power control algorithm for CDMA technologies is one example of an algorithm that converges to optimum network performance in the limit of perfect operation and sufficiently large soft handoff areas [9]. In the following subsections, we investigate if there is a corresponding downlink power control mechanism that optimizes network-wide performance criteria. This mechanism regulates the fraction of power provided by all active-set servers to each user and how user cell assignment reflects the particular temporary cell load levels. Although this problem has relevance only for dedicated channels in 3G1X and UMTS R99, it illustrates the above methodology with a mathematical model that can be solved in straightforward fashion. This example further shows how the specific nature of the mathematical optimization can lead to the development of a distributed implementation.

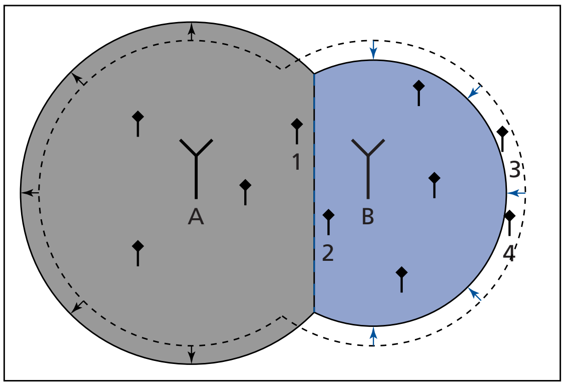
Formulation of Optimization Task: Global Downlink Power Control

Fast downlink power control as provided by 3G technologies is a powerful feature that guarantees sufficient resources for every mobile while minimizing the power-amplifier (PA) load of each cell for the assigned set of users. The assignment of users to cells is determined by an independent process, which is uncoordinated with channel power allocation and can therefore lead to large variations in power demand from one cell to the next. As shown in numerous predictions and field trials, network capacity can be substantially improved by balancing these load variations through proper reassignment of users from overloaded to lightly loaded cells [5, 6, 8]. Such a procedure is performed during the network optimization process, where cell boundaries are shifted through adjustments of antenna configuration or pilot power to balance the long-term average loading of cells. These procedures can achieve capacity improvements of up to 25%. Obviously, one can expect larger gains when load balancing compensates temporal fluctuations in resource demand.

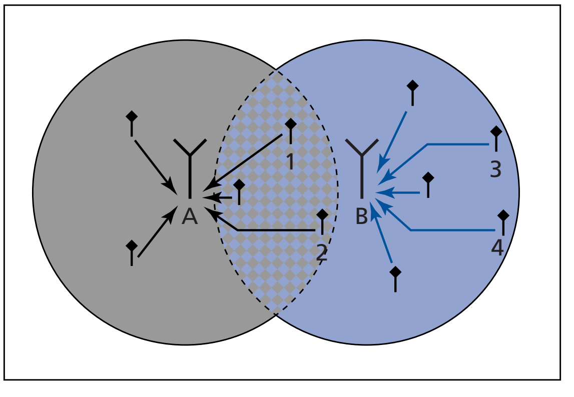
It can be shown that combining power control with fast adjustment of cell boundaries may not lead to best performance. **Figure 12** shows two different methods of load balancing. Figure12a shows an example of a few mobiles that are unevenly distributed



(a) Cell B, with seven users, more heavily loaded than cell A



(b) Expanding cell A and shrinking cell B, with two users dropped



(c) Balanced load, with no users dropped, through individual assignment of users

Figure 12.
Different methods of load balancing.

over the access areas of two cells. In the overlap region, users are assigned to the strongest-serving cell. As a result, cell B ends up more heavily loaded than cell A. Adjusting the cell access areas to rebalance the load has the undesired side effect that some users in the shrinking cell may get dropped or suffer from temporary reduction in QoS, as shown in Figure 12b. Figure 12c shows an alternative approach, where load balancing is accomplished by individually assigning users to cells without causing the undesired side effect. Mathematically, this alternative approach translates into a network-wide cell mobile resource assignment problem. It is obvious that this problem has to be solved with sufficiently high frequency to account for channel fluctuations, user mobility, and both call arrivals and departures.

When power is the limited resource, the network-wide cell mobile resource assignment is represented by the matrix p_{mc}, which determines the amount of power cell c provides to mobile m. Since p_{mc} allows several cells to provide power to one mobile, it implicitly supports soft handoff. The following conditions define the best solution for p_{mc} in form of an optimization problem, P:

- The PA load of the most heavily loaded cell should be minimized. This condition defines an objective function.
- Every user should be provided with sufficient QoS. This condition defines a constraint.

The objective performs the task of load balancing, while the constraint pursues the same goal as power control. The optimization problem can be written in the following form:

$$P : \min_{p_{mc}} \max_c L_c \qquad (1)$$

subject to

$$\sum_{c=1,C} \alpha_{mc} p_{mc} \geq \gamma\left(\eta + \sum_{c=1,C} \alpha_{mc} L_c\right) \quad \forall 1 \leq m \leq M \qquad (2)$$

$$L_c = L_{OH} + \sum_{m=1,M} p_{mc} \quad \forall 1 \leq c \leq C \qquad (3)$$

with $p_{mc} \geq 0$ and $L_c \geq 0$.

Here, α_{mc} is the propagation loss, γ is the target SINR, η is the thermal noise floor, L_c is the load of amplifier in cell c, and L_{OH} the load of overhead channels. Expression (1) represents the objective function and (2) translates the above-cited constraint into an SINR requirement; (2) mimics code division multiple access (CDMA) downlink power control with maximum ratio combining for signals from different cells. In this formulation, we have neglected bounds to power control as well as same-cell channel code orthogonality for the sake of simplicity. We further consider only one service since all users share the same value for γ.

The optimization problem defined by (1) subject to (2) and (3) represents a linear program (LP). This has the advantage that a solution can easily be obtained using standard solvers. It further allows insight on how to derive a distributed implementation as outlined in the "Development of Distributed Implementation" subsection below.

Comparison of Results from LP to 3G System

The optimum solution provided by the LP is compared to the performance of a 3G1X system through simulation. To capture fluctuations in load and channel conditions, a time-driven simulator is employed. The simulation is performed on a network layout consisting of four cells located at the corner points of a square coverage region (**Figure 13**). This scenario is sufficiently large to capture network-level aspects while keeping the complexity of intercell relations at a comprehensible level. We restrict the simulation to one circuit-switched service, which matches the above formulation of the LP. We permit calls to originate according to a spatially uniform Poisson process with exponentially distributed holding times. Users can move with constant speed and randomly varying direction over the network area, and they are bounced back as soon as they hit a boundary.

In this example, we assume that cell power is the only limited resource and neglect the uplink. Due to uniform call arrival and mobility properties over the coverage area, load fluctuations are solely caused by dynamics of arrival process, user motion, and channel fading characteristics, which cover time scales between milliseconds and minutes. Due to the symmetry of layout and traffic conditions, all cells have the same (i.e. balanced) long-term properties. This

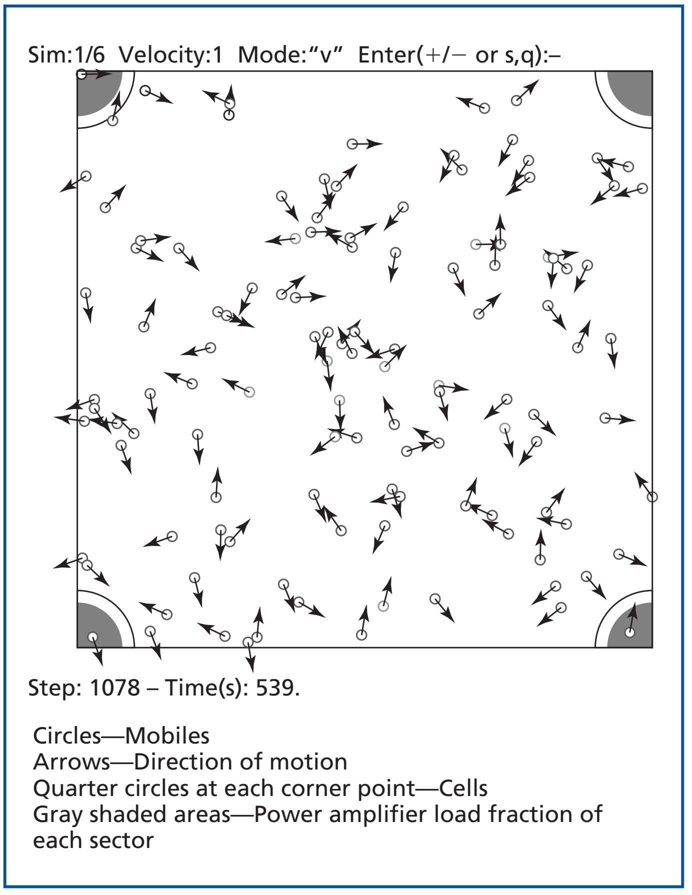


Figure 13.
Simulation of 3G cellular network with four cells.

guarantees that the gains can entirely be attributed to dynamic adjustments through the LP.

One critical parameter in the simulation process is the time granularity used in the evaluation. It should be guided by an optimistic estimation of the response time for a realistic implementation of the LP. At this point, we anticipate the implementation introduced in the next subsection, which accomplishes the LP solving process through two interdependent control algorithms: one resembles power control and can follow rapidly changing channel conditions; the other requires an iterative process involving intercell communications. We can optimistically estimate the settling time for the latter mechanism to a few hundred milliseconds. Since the convergence of both algorithms is necessary to achieve the full gain of the LP over a common 3G system, we set the granularity of the simulation to 500 ms and solve all faster processes in closed form at each time step.

The fast processes include channel variations due to multipath fading. For the 3G system, the interrelation between cell load level, interference, and channel power defines a fix-point problem, which is solved in iterative fashion (see [9]). Arrival process, user mobility and shadow fading are evaluated explicitly from one step to the next. For the 3G-system, the soft-handoff configurations are set according to the IS95B standard with respect to time-step-averaged channel conditions. We further make the approximation that all servers in the active set provide the same transmit power. For the LP, the soft handoff configurations are derived during the solving process.

The final comparison is based on statistical averages over a large number of simulation snapshots. Since the LP provides the optimum solution to a snapshot condition, it is agnostic to the dynamic flow of processes. Hence, it does not include mechanisms that provide link robustness to unaccounted channel fluctuations; 3G systems support some features of that kind (e.g., as soft-handoff), and they come at the price of capacity. To ensure that favorable results for the LP are not driven by the omission of robustness features, we have repeated the comparison for a variety of scenarios, where certain robustness-providing or robustness-demanding factors were turned off. The scenarios read as follows:

- *Scenario A.* This scenario captures channel fluctuations due to shadow fading and soft handoff in close proximity to real implementations. Soft handoff is supported for up to four servers. It includes a drop timer of 5s, which is invoked before servers are dropped from the active set. Obviously, multiple-leg soft handoff and drop time ensure link robustness in the presence of channel fluctuations. Figure 13 shows a simulation snapshot.
- *Scenario B.* The drop timer is eliminated. As a result, soft handoff legs are dropped from the active set as soon as their pilot signal strength falls under the required minimum threshold. This reduces robustness, since channels that have undergone a fade for only a brief moment need to be set up again.
- *Scenario C.* Soft handoff and drop timer are eliminated. This step further reduces robustness to fast

Table I. Comparative studies via simulations of the linear programming solutions and the distributed dual-ascent implementations compared to standard 3G systems.

Scenario	Data rate (arb. units)	Load reduction in worst cell: LP compared to 3G1X					
		Exact LP solution			Distributed implementation		
		min	mean	max	min	mean	max
A	10	16%	**60%**	99%	13%	**60%**	99%
B	10	6%	**46%**	99%	3%	**45%**	99%
C	10	7%	**52%**	96%	5%	**52%**	95%
D	10	3%	**22%**	48%	2%	**21%**	48%
A	20	14%	**63%**	98%	8%	**62%**	98%
B	20	5%	**48%**	98%	2%	**47%**	98%
C	20	1%	**52%**	97%	0%	**51%**	97%
D	20	1%	**31%**	93%	0%	**29%**	93%
A	40	21%	**62%**	98%	13%	**61%**	98%
B	40	5%	**49%**	97%	0%	**48%**	97%
C	40	3%	**52%**	97%	0%	**51%**	97%
D	40	1%	**34%**	96%	0%	**32%**	96%

3G1X—CDMA2000† first evolution
LP—Linear program

†Registered trademark of the Telecommunications Industry Association (TIA-USA).

channel variations unless captured by other means (e.g., as through dual polarization diversity at the mobile receiver). Such a solution has been discussed for 1xEV-DO, and it would also be feasible for 3G1X.

- *Scenario D.* In addition to soft handoff and drop timer, shadow fading is eliminated. This would represent a scenario of rather immobile data users. If we assume that robustness is achieved by other means, this scenario will definitely set the highest bound to capacity in present 3G systems.

We have evaluated the performance gains for the four scenarios and with respect to three different data rates leading overall to 12 trials. The data rate is reflected in the value of γ in (2), which translates into effective power consumption at given link conditions. In order to keep the long-term average PA load for the various data rates at approximately the same level, we have reduced the call arrival rate proportionally. For each of the 400 simulation snap shots of each trial, we solved the corresponding LP using a commercial solver [10].

The trial results are summarized in **Table I**, columns 3–5. The "gain" is measured by the amount the LP can reduce the PA load of the most heavily loaded cell over the 3G1X model system. The improvements due to the LP approach are shown for the mean, the minimum and the maximum for the 400 snapshots taken for each trial compared to 3G mechanisms. Inspecting the results, we observe that the percentage gains from the LP range between 0% and 99% in the various scenarios, with an average around 50% for scenarios A, B and C. The elimination of robustness factors such as active-server drop timers or soft handoff apparently has a rather small impact on the resource utilization of 3G systems. This indicates that potential implementations of the LP would substantially improve efficiency without sacrificing robustness. With an average around 25%, scenario D shows substantially lower gains. Since this scenario spares out on the rather fast channel fluctuations caused by shadow fading, one can conclude that dynamic optimization can especially improve on fast fluctuations in resource demand. Finally, the results show slightly larger improvements for higher data rates, which can be attributed to the enhanced fluctuations encountered with fewer users.

We want to emphasize that the improvements through dynamic optimization shown are entirely due to fluctuations occurring on time scales of minutes and below. In real networks, the LP solution can show substantially higher improvements since traffic fluctuations on slower time scales can be accounted for as well.

Development of Distributed Implementation

While the LP may provide a considerable load reduction, it relies on fast, centralized processing, which is hard to accomplish in real systems. It would therefore be appropriate to find a method that allows achieving the same or similar gains through a set of distributed control algorithms. To accomplish this task, two different approaches can be pursued. One approach starts out from existing systems and tries to improve the available control algorithms to approach the best solution provided by the LP. Since there is no straightforward mathematical recipe, such an approach would be guided by engineering judgment combined with evaluation through simulation.

We want to illustrate another approach, which utilizes the mathematical structure of the LP to derive a distributed implementation. For that purpose, we consider the dual linear program, D, of the optimization problem in (1)–(3), which is given by

$$D : \max_{\lambda} \gamma\eta \sum_{m} \mu_m \tag{4}$$

subject to

$$\alpha_{mc}\mu_m \leq \gamma \sum_{c} \alpha_{mc}\mu_m + \lambda_c \quad \forall 1 \leq m \leq M,\ 1 \leq c \leq C \tag{5}$$

$$\sum_{c} \lambda_c = 1 \tag{6}$$

with $\lambda_c \geq 0$ and $\mu_m \geq 0$.

As is well known in optimization theory, this companion problem gives the same optimal solution from a different perspective. It also provides the means for a disaggregated solution to the LP that uses local (i.e., cell-level) information to perform part of the processing on a fast time scale and enables definition of a new protocol using intercell communications to balance the load among cells on a slower time scale. The local part of the problem is represented by (5), which resembles the uplink power control problem of CDMA systems and can be handled through an equivalent control algorithm for a given vector of dual variables, λ_c. Additionally, (4) and (6) require an algorithm that relies on intercell communications and sets the particular values of λ_c. The solution of the dual LP can be translated into an economical problem, where λ_c represents the "price of power" offered by each cell, and μ_{mc} is equivalent to the "payment" made by mobile m to cell c for purchase of this power. In this picture, balancing load translates into maximizing total revenues by appropriately setting (normalized) prices. We refer to this solution approach as *dual ascent,* in reference to the fact that a solution based on this interpretation carries out an ascent (maximization) via use of dual "prices" described above. Detailed considerations of this model are given in [2].

To determine the quality of the dual approximation as described above, we also implemented the distributed version of the dual LP and repeated the 12 trials introduced in the prior section. The corresponding results are included in Table I, columns 6–8. They are almost identical to those obtained for the LP. This result confirms that distributed implementation of the LP optimization can be realized for the network-wide optimization problem introduced without loss in performance.

Conclusions

We have presented several facets of the Bell Labs Research program on dynamic network optimization that provides cellular networks with the capabilities to respond to fluctuations in traffic and resource demand. We have presented applications to current networks, where network trials have demonstrated improved network accessibility through dynamic response to recurring traffic patterns, as well as longer-term efforts that focus on the development of faster, coordinated response mechanisms that can successfully adapt to load fluctuations across cells.

We have further demonstrated that real-time measurements are a fundamental ingredient to the development of dynamic control mechanisms since they reveal the inefficiencies in cellular networks and provide information on actual traffic characteristics and user behavior. In addition, they identify important interrelations among network properties, such as per-call QoS, resource demand, and traffic load, which in turn drive the development of future control algorithms. Information on actual traffic behavior is already being applied in current dynamic network trials for optimization of recurring traffic patterns. Finally, real-time measurements become an integral part of future dynamic optimization features for diagnostic and monitoring purposes.

The interplay of these facets is critical to our dynamic optimization roadmap. Features developed for present networks validate our dynamic optimization approach and thereby facilitate the rollout of future control mechanisms. We are convinced that dynamic optimization algorithms will continuously drive improvements in key performance metrics, such as throughput, dropped-call rate, and resource utilization.

Acknowledgments

The authors are grateful to numerous colleagues for interesting and useful discussions. Particular thanks are due to Alvaro Diaz, S.J. Han, David James, Diane Lambert, Luca Marinelli, Aris Moustakas, Bhanumathi Ravishankar, Jay Srinivasan, and Debasis Mitra for helpful comments and encouragement and to Andrew Mackay, Shaun Morton-Jones, and Gabriel Wong for field support in the network optimization trials.

*Trademarks

CDMA2000 is a registered trademark of the Telecommunications Industry Association (TIA-USA).

CdmaOne is a registered trademark of the CDMA Development Group.

UMTS is a trademark of the European Telecommunications Standards Institute (ETSI) under auspices of the ITU.

WatchMark Prospect is a trademark or registered trademark of WatchMark Corp. in the United States and other countries.

References

[1] A. Bedekar, S. C. Borst, K. Ramanan, P. A. Whiting, and E.M. Yeh, “Downlink Scheduling in CDMA Data Networks,” Proc. Globecom ’99 (Rio de Janiero, Braz., 1999), pp. 2653–2657.

[2] S. Borst, G. Hampel, I. Saniee, and P. Whiting, “A Shadow-Priced-Based Distributed Approach to Dynamic Load Balancing in CDMA Networks,” Handbook of Optimization in Telecommunications (P. M. Pardolas and M.G.C. Resende, eds.), Springer Verlag, Norwell, MA, forthcoming 2005.

[3] A. Buvaneswari, B. Ravishankar, J. Graybeal, M. Haner, and G. Rittenhouse, “New Optimization and Management Services for 3G Wireless Networks using CELNET Xplorer,” Bell Labs Tech. J., 9:4 (2005) 101–115.

[4] S. Das, H. Viswanathan, and G. Rittenhouse, “Dynamic Load Balancing Through Coordinated Scheduling in Packet Data Systems,” Proc. IEEE INFOCOM (San Franscisco, CA, 2003), 786–796.

[5] L. M. Drabeck, M. J. Flanagan, J. Srinivasan, W. M. MacDonald, G. Hampel, and A. Diaz, “Network Optimization Trials of a Vendor-Independent Methodology Using the Ocelot® Tool,” Bell Labs Tech. J. 9:4 (2005), 49–66.

[6] K. Goto, T. Suzuki, and T. Hattori, “Cell-Size Adaptation in W-CDMA Cellular Systems,” Proc. Veh. Technol. Conf. (Birmingham, AL, 2002), 444–448.

[7] G. Hampel, D. Abush-Magder, A. Diaz, L. M. Drabeck, M. J. Flanagan, J. M. Graybeal, J. D. Hobby, M. MacDonald, P. A. Polakos, J. Srinivasan, H. Trickey, L. Zhang, and G. Rittenhouse, “The New Paradigm for Wireless Network Optimization: A Synergy of Automated Processes and Human Intervention,” IEEE Commun. Mag., 43:3 (2005), s14–s21.

[8] G. Hampel, K. L. Clarkson, J. D. Hobby, and P. A. Polakos, “The Tradeoff Between Coverage and Capacity in Dynamic Optimization of 3G Cellular Networks,” Proc. Veh. Technol. Conf., 58 (Orlando, FL, 2003), 927–932.

[9] S. V. Hanly, “An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular Spread Spectrum Capacity,” IEEE J. on Selected Areas in Commun., 13:7 (1995), 1332–1340.

[10] ILOG, Inc, CPLEX 6.0, Incline Village, NV, 1998.

[11] W. C. Y. Lee, Mobile Communications Design Fundamentals, John Wiley, New York, NY, 1993.

[12] The Members of Technical Staff, Bell Labs, Handbook of CDMA System Design, Engineering, and Optimization, (K. I. Kim, ed.), Prentice Hall PTR, Upper Saddle River, NJ, 2000.

(Manuscript approved April 2005)

SIMON C. BORST is a member of technical staff in the Mathematics of Networks and Systems Research Department of the Mathematical Sciences Research Center at Bell Labs in Murray Hill, New Jersey. He received the M.Sc. degree in applied mathematics from the University of Twente and the Ph.D. degree from the University of Tilburg, both located in the Netherlands. His main research interests are in the mathematical analysis of resource allocation problems in communication networks and computer systems. Dr. Borst, who has authored over eighty publications in journals, conference proceedings, and book chapters, serves on the editorial boards of several journals. He holds ten patents and has four patent applications pending in the area of networking and wireless communications.

ARUMUGAM BUVANESWARI is a member of technical staff in the Networking Infrastructure Research Department of the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. She holds a B.E. degree in electronics and communication from Thiagarajar College of Engineering in Madurai, India, and an M.Sc. degree in electrical communication engineering from the Indian Institute of Science in Bangalore. Her research interests are in optimization of wireless and landline networks through data capture and analysis, wireless high-speed packet data techniques, and embedded systems. She has a number of publications in the area of digital signal processing algorithms and firmware.

LAWRENCE M. DRABECK is a member of technical staff in the Wireless Technology Research Department of the Wireless Research Laboratory at Bell Labs in Holmdel, New Jersey. He holds B.S., M.S., and Ph.D. degrees in experimental condensed matter physics from the University of California at Los Angeles. Dr. Drabeck is currently working on wireless network optimization.

MICHAEL J. FLANAGAN is a technical manager in the Open Innovations Laboratory of the Mobility Solutions business unit in Whippany, New Jersey. He holds a B.S. degree in electrical engineering from the Stevens Institute of Technology in Hoboken, New Jersey, and M.S. and Ph.D. degrees, also in electrical engineering, from the California Institute of Technology in Pasadena. Dr. Flanagan is currently working on wireless network optimization.

JOHN M. GRAYBEAL is a member of technical staff in the Wireless Technology Research Department of the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He received a B.S. (with highest honors) from the University of Texas at Austin and a Ph.D. from Stanford University in Palo Alto, California, both in physics. Dr. Graybeal's present research interests are in predictive and dynamic optimization algorithms for cellular wireless networks and also in experimental measurements upon and analysis of operating wireless networks. Originally a condensed matter experimental physicist, he previously served on the physics faculties at the University of Florida and the Massachusetts Institute of Technology. He is a past recipient of a Sloan Research Fellowship and a Young Investigator award from the Office of Naval Research.

GEORG K. HAMPEL is a member of technical staff in the Wireless Technology Research Department of the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He received M.S. and Ph.D. degrees in physical sciences from J.W. Goethe Universitat in Frankfurt am Main/Germany. Dr. Hampel has provided major contributions to the establishment and realization of automated network optimization. His current research interests include dynamic network reconfiguration for 3G and 4G wireless technologies and algorithm development for cross-layer network control. In the past, he has also worked in the areas of network growth planning, intelligent antenna systems, and supercon-ductor filter technology for the base station front end.

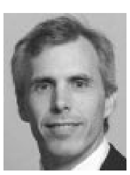
MARK HANER is a director of Networking Infrastructure Research in the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He holds B.S., M.S., and Ph.D. degrees in electrical engineering and physics from the University of California at Berkeley, where he also holds a Miller Research

Institute fellowship. Dr. Haner has focused his research on broadband access and fixed and mobile wireless systems. He helped establish the Wireless Broadband Networks Division in Silicon Valley as part of Mobility Solutions within Lucent Technologies, and he has served on advisory committees for both the Defense Advanced Research Projects Agency and the National Science Foundation.

WILLIAM M. MACDONALD is a member of technical staff in the High Speed Circuits and Systems Research Department of the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He received a B.S. degree in physics from the University of Notre Dame in South Bend, Indiana, and M.S. and Ph.D. degrees, also in physics, from the University of Illinois at Urbana-Champaign. Dr. MacDonald has worked on optical communications as well as wireless communications.

PAUL A. POLAKOS is director of Wireless Technology Research in the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He received a B.S. degree in physics from Rensselaer Polytechnic Institute, and M.S. and Ph.D. degrees from the University of Arizona Department of Physics and Astronomy. He spent more than 10 years in elementary particle physics research at Fermilab (Chicago, Illinois), CERN (Geneva, Switzerland), and the Max-Planck Institute for Physics and Astrophysics (Munich, Germany) prior to joining Bell Labs. During his 20-year tenure in physics and wireless research at Bell Labs, Dr. Polakos has been responsible for a number of key initiatives in wireless communication systems, including intelligent antennas, advanced BTS and RAN architectures, enhancements for high data rates and capacity, and wireless network optimization.

GEORGE RITTENHOUSE is vice president of the Wireless Research Laboratory at Bell Labs in Murray Hill, New Jersey. He holds a B.S. degree in physics from the University of California at Los Angeles and M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology. He developed a high-speed 0.1 μm NMOS process for optical networking and headed several projects, including MIMO system development, network optimization, wireless IP networks, and fourth-generation wireless. His research has focused on RF front-end radio architectures and cellular system engineering. A Bell Labs Fellow, Dr. Rittenhouse has numerous publications and patents in the areas of wireless systems and circuits.



IRAJ SANIEE is director of Mathematics of Networks Systems Research in the Mathematical Sciences Research Center at Bell Labs in Murray Hill, New Jersey. He received his B.A. and M.A. degrees in mathematics and his Ph.D. in operations research and control theory, all from Cambridge University in the United Kingdom. At Bell Labs, the emphasis of research in his department is on the modeling, analysis, and optimization of emerging processes in wireless, optical, and data networks. Recent joint projects with Lucent businesses include new schemes for scheduling and network control algorithms. Dr. Saniee is past chair of the Telecom Section at INFORMS and a member of IEEE and the IFIP Working Group 7.3 on Computer Performance Modeling and Analysis. He is currently an associate editor of Operations Research. He led the 1994 runner-up team for INFORMS' Edelman International Prize in Operations Research, and he has published numerous articles in IEEE, INFORMS, and SIAM journals and proceedings.

ALAN WEISS is a distinguished member of technical staff in the Fundamental Mathematics Research Department of the Mathematical Sciences Research Center at Bell Labs in Murray Hill, New Jersey. He earned B.S. degrees in physics and mathematics from Case Western Reserve University in Cleveland, Ohio, and M.A. and Ph.D. degrees in mathematics from the Courant Institute, New York University, in New York City. Most of his work relates to stochastic processes or the control of distributed systems. Dr. Weiss' specialty is the theory of large deviations, especially as applied to performance analysis, queueing, and computing.

PHILIP A. WHITING is a member of technical staff in the Mathematics of Networks and Systems Research Department of the Mathematical Sciences Research Center at Bell Labs in Murray Hill, New Jersey. He received his B.A. in mathematics from the University of Oxford, his M.Sc. in probability and statistics from the University of London (both in the United Kingdom),

and his Ph. D. in queueing theory from the University of Strathclyde in Glasgow, Scotland. After post-doctoral work at the University of Cambridge, Dr. Whiting's interests centered on wireless. He subsequently participated in the Telstra trial of Qualcomm CDMA in southeastern Australia and later joined the Mobile Research Centre at the University of South Australia in Adelaide. His main interests are the mathematics of wireless networks, particularly stochastic models for resource allocation and information theory. His current research includes large deviations asymptotics for balls and urns models, the application of stochastic control to scheduling in wireless data networks, and modeling and schemes for performing hybrid ARQ. ◆