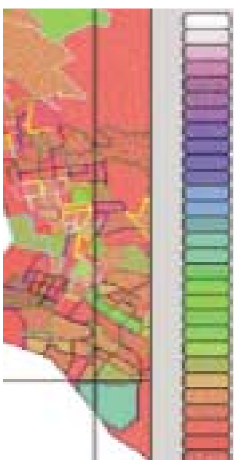
A FRAMEWORK FOR THE DIMENSIONING OF BROADBAND MOBILE NETWORKS SUPPORTING WIRELESS INTERNET SERVICES

LEILA Z. RIBEIRO AND LUIZ A. DASILVA,
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY



The integration of heterogeneous flows in wireless mobile networks requires that new methodologies for network design and resource dimensioning be developed.

ABSTRACT

The integration of heterogeneous flows in wireless mobile networks requires that new methodologies for network design and resource dimensioning be developed. In this article, we discuss a framework for the dimensioning of such networks by taking into account differentiated user and traffic profiles. Distinct QoS requirements from various applications result in different aggregate throughput per cell being achievable for the same loading factor and network layout depending on user mix. Therefore, appropriate characterization of user mixes and aggregation techniques that map these mixes into resource requirements are key in the design of third-generation systems.

INTRODUCTION

From the original systems created in the 1980s until the end of last decade, mobile communications have primarily been used for circuit-switched voice communications, with occasional low-bandwidth data connections.

The evolution toward third-generation (3G) systems addresses the increasing user need for mobile data access, mainly due to the growth of Internet use, with mobile subscribers becoming more and more demanding of a system that supports services such as wireless email, Web browsing, wireless telecommuting, and streaming video, among others. This evolution includes a migration path through 2.5G (e.g., General Packet Radio Service, GPRS), which introduces the concepts of variable bit rates and the creation of an alternate path for packet-switched data in the core network. 3G systems will expand the concept of 2.5G by supporting larger bandwidth and quality of service (QoS) provisioning.

Many 3G systems are being planned for implementation in the near future, with global harmonization efforts being led by the 3rd Generation Partnership Project (3GPP) that standardizes wideband code-division multiple access (WCDMA), Enhanced Data for GSM Evolution (EDGE), and time-division space-code-division multiple access (TD-SCDMA) technologies as the 3G evolution from current Global System for Mobile Communications (GSM) and TDMA systems. Additionally, 3GPP2 was created to standardize cdma2000 systems as the 3G evolution path from existing IS-95 systems.

The shift to a system that fully supports heterogeneous traffic imposes a number of new challenges on the design and dimensioning of network infrastructure. These challenges can be summarized in three distinct areas:

- *Traffic characterization*: Basic, well established assumptions in modeling circuit-switched voice traffic no longer hold true with the convergence of voice and data over a packet-switched infrastructure. Self-similar models may have to be adopted in order to appropriately take into account the burstiness that packet traffic is expected to exhibit in all timescales.
- *Demand characterization*: For an accurate assessment of resource requirements, one must consider the distinct demand profiles of residential users, business users, and road traffic, as well as forecasts of how the market for mobile services will evolve over the near and medium terms.
- *Achieving QoS in 3G systems*: QoS requirements at the application layer must be mapped into parameters that can be controlled at the physical (spreading factors, power) and/or data link layers (medium access control mechanisms).

In this article, we introduce a basic framework in which to combine the above considerations to arrive at an efficiently dimensioned design. This work is currently in progress and will lead to a set of design methodologies and

This research was partly supported by the National Science Foundation under the Integrative Graduate Education and Research Traineeship Program.

Portions of this article were presented at the 2001-Fall IEEE Vehicular Technology Conference.

Some of the pictures presented in this article utilize network planning software developed by CelPlan Technologies, Inc.

1070-9916/02/$17.00 © 2002 IEEE

tools for the dimensioning of 3G systems. The next three sections discuss the three main areas of interest outlined above. In them we will present a brief review on the main theoretical aspects of traffic modeling, detail a practical methodology for demand characterization, discuss the mapping of QoS requirements from the higher layers, with a focus on the physical layer; and finally summarize our conclusions.

TRAFFIC CHARACTERIZATION

The integration of heterogeneous traffic in a cellular network adds complexity to the problem of characterizing demand for bandwidth. Different applications may call for the adoption of distinct traffic models, the definition of distinct QoS requirements, even the use of distinct metrics for traffic description and performance characterization.

Traditional telephony describes traffic load in terms of call arrival rate and holding time; in 3G systems, we must also consider peak and sustained (long-term average) data rates and traffic burstiness. While for circuit-switched networks supporting voice applications the measure of QoS is reasonably defined through the use of a few metrics (e.g., call blocking probability and probability of dropping a call due to handoff), for data networks we must consider a larger set of metrics. These include bit error rate, frame loss rates, throughput, flow blocking probability, delay, and delay variability (jitter).

In this section, we discuss possible traffic models and QoS requirements for the diverse set of applications that will be offered in broadband wireless environments.

TRAFFIC MODELS

For the past several decades, traffic modeling and dimensioning of voice networks has relied on the work developed by Erlang around 1909. Call arrivals are modeled as a Poisson process, and holding times are often assumed to follow an exponential distribution. The system is assumed to have a finite number of channels, and consequently to support a maximum number of simultaneous calls. For system dimensioning purposes, there are well-known mathematical relationships that relate the total traffic offered, the number of channels provided in the system, and the probability of call attempt failure.

In these models, there are different ways to define call attempt failure, depending on the behavior of the system when a call is received and all circuits are busy. The assumption adopted by most traditional fixed telephony systems and in many cellular systems (AMPS, IS-136) is that the attempt is immediately refused, and the user immediately receives an indication of failure and could try again if (s)he so desires. An alternative assumption (e.g., used in GSM) is for the system to put the user on a queue and allow some waiting time; if the waiting time exceeds a predefined threshold, the call attempt has failed. This mechanism is known as *blocked calls delayed*, and the call blocking probability is defined as the probability of the waiting time exceeding the maximum allowed delay. The first approach is modeled using the Erlang B formula, while the second is described by the Erlang C formula [1]. The typical dimensioning problem in telephony is to calculate the required number of channels to support some traffic forecast with a given maximum call blocking probability. Both of the formulas mentioned above allow direct mathematical treatment of this problem.

The Markovian models, of which the Poisson-related are the main subset, constitute the fundamental pillar of queuing theory today. Erlang theory is one example of such an approach, but call arrivals may also be modeled as a Poisson batch model, as a Markov-modulated Poisson process, and so on. While such models have been successfully used to characterize voice call arrivals (as well as other arrival processes relating to connection-oriented data transmission, e.g., the establishment of an FTP connection), the arrival of data packets at a node is not well modeled by such processes.

More recent models rely on the observation that traffic patterns generated by data sources with variable bit rate applications tend to exhibit long-range dependence and burstiness over a wide range of timescales. These fractal-like characteristics fall under the definition of self-similarity. The seminal article on the study of self-similar traffic was published in 1993–1994 [1] and shattered the basic assumption that Poisson models were applicable to packet data traffic.

A mathematical characterization of self-similar processes would show approximately the same statistics for magnified (or aggregated) versions of the same process. This suggests that the burstiness persists on different timescales. A weaker condition is asymptotical self-similarity, which states that autocorrelation functions for compressed versions of the process keep the same form as for the original process. A consequence of this property is that, as the compression factor tends to infinity, the correlation does not tend to zero (as would be the case, e.g., for Poisson processes). This characteristic is referred to as *long-range dependence*. Self-similarity and long-range dependence are not exactly the same concept and do not necessarily imply each other, but in the context of traffic dimensioning these terms are often used interchangeably.

There has been some recent research [2] attempting to understand the physical causes of self-similarity in network traffic. Two main causality possibilities are considered. Single source causality considers the arrival pattern of a single data source; in this manner, the variability found at multiple time scales for a variable bit rate (VBR) video stream (e.g., MPEG) could be explained by the variability in time duration between two successive scene changes. Structural causality attributes the self-similarity observed at multiplexing points in the network layer to the end hosts' exchanging of objects whose sizes present a heavy-tailed distribution.

From the point of view of queuing analysis, research has focused [2] on the impact of self-similar sources on performance, as compared to Markovian assumptions. It has been found that increasing buffer sizes results in little improvement in packet loss rate for self-similar sources. For this reason, there have been proposals advo-

The integration of heterogeneous traffic in a cellular network adds complexity to the problem of characterizing demand for bandwidth. Different applications may call for the adoption of distinct traffic models, the definition of distinct QoS requirements, even the use of distinct metrics for traffic description and performance characterization.

	User profile	1	2	3	4
Service class	Service name	Videoconference	Voice	Streaming video	Web browsing
	Service class group	Conversational	Conversational	Streaming	Interactive
	Type of switching	Circuit	Circuit	Packet	Packet
	Max. bit rate down (kb/s)	64	13	144	1000
	Max. bit rate up (kb/s)	64	13	0	500
	Session arrival distribution	Exponential	Exponential	Exponential	Exponential
	Burst arrival distribution	Exponential	Exponential	Exponential	Exponential
	Burst length distribution	Pareto	Exponential	Pareto	Pareto
	Required BER down	1.0E–06	1.00E–03	1.00E-06	1.00E–08
	Required BER up	1.0E–06	1.00E–03	-	1.00E–08
	Required E_b/N_0 down (dB)	8	7	12	1
	Required E_b/N_0 up (dB)	8	7	–	1
	Tolerated delay (ms)	100	100	200	–
Propagation environment	Propagation scenario	Indoor slow	Outdoor fast	Outdoor slow	Indoor slow
	Terminal speed (km/h)	0	120	120	3
Terminal type	User terminal	Laptop	Mobile phone	Palmtop	Laptop

Table 1. *Examples of multiple user classes and performance requirements.*

cating small buffer capacity/large bandwidth resource provisioning strategies.

Provisioning for 3G wireless systems must incorporate these latest developments in traffic modeling. Analytical models will likely have to be coupled with measurement-based modeling and traffic control techniques to appropriately account for the heterogeneity of traffic sources.

QoS Requirements

Just as the traditional assumptions in traffic modeling no longer hold for heterogeneous applications, neither do performance objectives. The larger set of possible applications can be grouped into four main categories of service classes according to the Universal Mobile Telecommunications Services (UMTS) definition, which we discuss next [3].

The *conversational class* includes voice, videoconference, video games, and so on. The most important performance requirement for this class is consistency in time relations, including low delay and low jitter requirements. The idea is to preserve, to the extent possible, the source data rate; data integrity, on the other hand, is not as critical, since acceptable thresholds depend on human perception for most applications.

The *streaming class* includes streaming audio and/or video (e.g., video on demand). It requires low jitter, although low delay is not as critical as for the conversational class. Data integrity, again, is not as critical as time variation sensitivity.

The *interactive class* encompasses Web browsing, database retrieval, and remote LAN access. Round-trip delay, characterizing request-response time, is of interest for this class. Data integrity, manifested as a low loss rate, is very important for these applications.

Finally, *background class* applications include non-real-time download of emails, file transfers, and so on. There are no strong delay or jitter requirements, although data integrity is reasonably critical to minimize retransmissions.

The issues involved in incorporating QoS mechanisms in mobile environments have received considerable attention from industry as well as the research community [4–6]. Renegotiation of performance guarantees during handoff, differentiated allocation of resources through power control, and prioritized access to the medium are some of the mechanisms through which QoS may be incorporated into 3G systems.

Demand Characterization

The previous section dealt with the characterization of traffic that may be produced by distinct applications generated by each user. In dimensioning the system, we must be aware of the different *user profiles* supported, which are distinguished by a combination of attributes defining their service class, propagation environment, and equipment characteristics. The concept of user profile is explained in detail in the following paragraphs.

The service class attributes will specify the nature and volume of traffic of a user profile, as well as its QoS requirements. Looking at the many possible types of users for mobile broadband systems and their most likely applications, such as Internet access, voice, and videoconferencing, it seems natural that we should try to group users into categories and describe each category by its typical parameters.

The mobility characteristic of the system will influence the user profile by having an impact on propagation behavior, which in turn limits data rates available to the user. The mobility also determines the limits of user equipment possibilities for each profile. With regards to the mobility characteristic of a user profile, we divide users into three major groups and present below the main propagation and terminal differences between them.

Indoor users: Building construction attenuation presents the major propagation impairment, having an impact on link strength, and limiting coverage area and bandwidth. On the other hand, users are mostly fixed or at very low speed, which reduces fast fading effects. Indoor signal boosters and microcells can be used to compensate for wall attenuations. Typical applications for this type of user are Internet access and connection to office LANs using laptops.

Pedestrian users: Terminal types include handheld phones and palmtops. Typically no building attenuations are expected, but on the other hand the propagation path is stressed more due to the larger cell radius allowed for outdoor cells in the system design. Voice and Internet access and video applications are the primary focus for these users.

Vehicular users: Typical terminal types are onboard car computers, hands-free terminals, and handset phones. A fast-moving subscriber unit means that the effects of multipath fading may be severe. Expected applications include voice, videoconferencing, Internet access, and navigation systems.

Of course, each of these types of users may generate multiple simultaneous flows of different service types. In turn, the performance requirements within an application type may differ depending on the propagation environment and terminal type. For instance, Internet access from a laptop within a building will allow higher bandwidth demands than Internet access by a pedestrian carrying a palmtop.

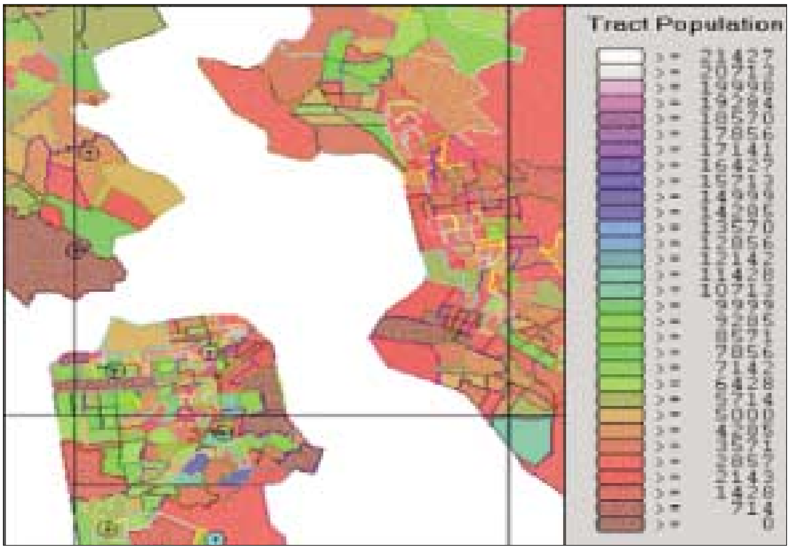
In order to characterize demand, we propose forming what we term a user profile by taking the Cartesian product of the service class (which specifies an application, performance requirements, and traffic characteristics), propagation environment, and terminal type. For instance, indoor-laptop-videoconference and pedestrian-palmtop-Web browsing would be two possible user profiles. Besides, traffic models for each class are likely to be parameterized differently for different user profiles (accounting for variations in access rates, burstiness, etc.). We show some examples in Table 1.

Accurate demand characterization should include proper description of the different geographical distributions for the users classified in each profile. The methodology presented here suggests the use of demographic database layers [7] as input to this process, and discusses possibilities of mapping these data into *user layers*. We discuss some of the main data inputs in the next subsections.

Residential Database (Census Data)

One important layer to be considered in forecasting traffic demand is the residential geographic user distribution. In the United States, this type of information usually comes from census data; the database is a collection of polygons that correspond to census tracts or census lots (a smaller subdivision of census tracts). Relevant information is available about the geographical area corresponding to each of these polygons, including:

- Residential population
- Number of households
- Statistics such as ethnicity, gender, and age distribution



■ Figure 1. *An example of a residential database (census tracts) for the San Francisco Bay area.*

Figure 1 illustrates this type of database, including information about population distribution in the San Francisco Bay area.

In addition to census data, other commercially available databases classify areas with additional attributes, such as average income, average number of years in school, and level of familiarity with "high-tech" consumer items. This information allows further levels of refinement in forecasting traffic coming from residential layers. For instance, a forecast could assign differentiated values of market penetration and average bandwidth depending on the household income in each area.

From the residential geographical distribution, using different market forecast factors, designers can create one traffic layer distribution for each applicable user profile.

Business Database

Another layer in building a geographic traffic distribution profile is a business database. This type of data may come from different sources in each market, such as chambers of commerce and GIS government databases. Due to the source diversity they do not follow any standard format but are typically either a collection of polygons or a collection of points, geographically distributed, with attributes including:

- Number of employees at the location
- Average business revenue
- Average monthly expenditures on telephone/data services
- Available office area in square feet
- Type of business

Figure 2 illustrates this type of database for a dense urban area.

To forecast traffic distribution based on this type of database, one typically relies on market penetration data per type of business and estimates the range of the required data rate depending on number of employees. For instance, a high-technology business will tend to have more employees who require wireless Internet access than a grocery store.

Block Code	Street Address	Area (1000sq.ft)	Height (ft)	Owner	Business Group
260021	46 Wall Street	493	312	46 Wall St. LLC	D9
270001	53 Wall Street	206	130	55 Wall Company NC	O1
270009	59 Wall Street	390	399	63 Wall Inc.	O3
270017	67 Wall Street	303	292	Gesellschaft Fuer MM	O4
310011	86 Wall Street	650	370	Eight-Five Wall Etc.	O4
330011	96 Wall Street	493	312	Chemical Bank	O3
330022	99 Wall Street	91	335	PA Building Co	O3
350010	107 Wall Street	890	296	Citicorp	O4

■ Figure 2 *An example of a business database in the Manhattan area.*

Road Traffic

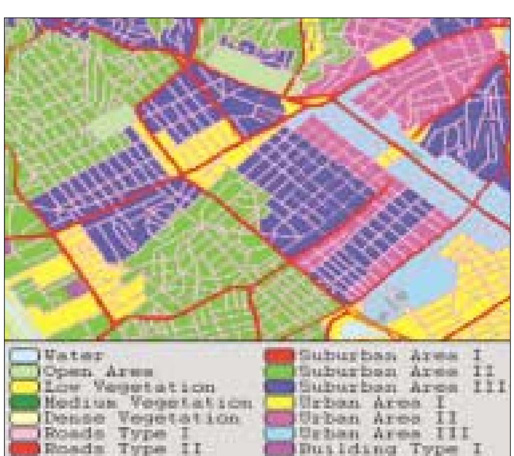
The road traffic database accounts for vehicular and pedestrian users. It locates geographically, usually as vectors, the main roads, highways and streets in a certain area. Typical information available in this database is the number of vehicles that transit in sections of the road per day. Forecast of data traffic must also take into consideration mobile traffic at the busiest hour.

Again, using different market forecast methods, one traffic layer distribution must be created for each applicable user profile related to vehicular and pedestrian users.

Morphology Weighting

As discussed above, traffic demand from residential users is typically available as a set of polygons. In performing resource planning, the designer is confronted with the question of how to geographically distribute the traffic associated with each polygon. The simplest approach would be to assume traffic to be uniformly distributed over the entire area covered by the polygon. However, further refinement is possible if information regarding land usage (morphology) is available. Morphology databases classify the terrain as to what type of land cover exists at each location. Typical terrain classifications include water, areas of low/medium/dense vegetation, urban and suburban areas, as exemplified in Fig. 3.

When distributing traffic geographically based on a residential database, one may consider different weights to account for the relative probability that a user in a certain polygon will be located in an area of dense vegetation vs. a suburban area, for instance.



■ Figure 3. *An example of a morphology database.*

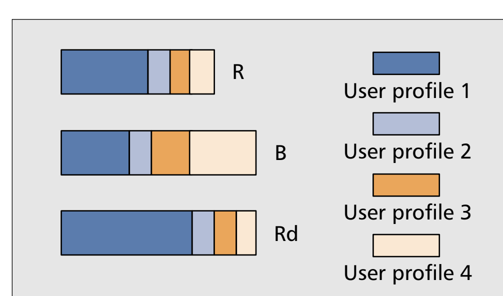
Other Demographic Layers

As discussed in the previous sections, multiple layers of traffic sources (residential, business, road traffic, etc.) must be considered individually as to how they correlate with expected load for each traffic type. Other factors that could also have an impact are event-related locations, which may be permanent (stadiums, theatres, and show arenas) or temporary (festivals, athletic games, etc.). Such considerations allow further refinement of load forecast to be used for network planning and design.

Market Penetration Forecast

Network design must take into consideration not only current requirements but also how these requirements may change over the near and medium terms. Market forecast data must be used to modify the information discussed above. Important metrics include market penetration, market share factor, and population growth from year to year.

Clearly, the network operator has a direct influence on some of these factors, since the growth of its customer base will depend on the provided service quality as well as how aggressive the marketing and pricing strategies are. On the other hand, service quality and return on investment to allow aggressive pricing will depend on a good dimensioning system that uses appropriate combinations of all factors mentioned above. Therefore, at the same time that



■ Figure 4. *Sample profiles for residential (R), business (B), and road traffic (Rd) for a given region, taking into account census data, morphology weighting, market forecast, and so on. The total load for purposes of network dimensioning is the superposition of all these factors.*

the operator can "forecast" market penetration, it can also influence the market and turn this forecast into reality.

Combining Multiple Layers

Once all the information described in the previous section comes together, it must be combined, as illustrated in Fig. 4. One should note that from a modeling point of view, this combination does not consist of a simple sum. First, each user layer (corresponding to a certain user profile) must be taken into account separately; besides, there is the question of time dependencies, discussed next.

Time Dependency

Suppose that, for each layer, the method outlined above helped us obtain the traffic distribution during the busiest hour. However, the busiest hour for each layer does not necessarily correspond to that of the other layers, as illustrated in Fig. 5.

Combining all layers at their peak hours would clearly lead to overdimensioning of the network. Time dependency weighting must be applied in order to correctly assess demand at any given time.

For instance, suppose we are evaluating system behavior at 2:00 p.m. The business layer should be considered with 100 percent weight, but the residential layer would be scaled to under 50 percent of its peak value. If we look at the combined traffic grid at this time, we would most likely find the highest bandwidth demand on those cells located in business areas, and few channels needed on the residential ones. However, if we look at the 7:00 p.m. time frame, heaviest demand for bandwidth is now more likely to come from residential areas.

Since we are looking for ways to dimension channels on a per sector basis, not only on a system basis, one cannot ignore the spatial distribution of the busiest hour over different layers, and the resulting required bandwidth per sector will be the maximum demand from looking at all time periods. To allow for all possibilities one would need to create combined class grids for all hours of the day; for simplicity, however, usually only two or three hours are selected. A typical system analysis looks into hours where most relevant layers have their peaks (e.g., residential and business in the example above).

Figure 6 illustrates one user grid created as a combination of multiple input layers (residential, business, etc.) for a certain user profile and time of day. This grid is represented by small boxes of fixed size (grid resolution), each one with a particular traffic attribute. Attributes could be expressed in terms of user density per area, absolute number of users, and so on.

The previous illustration displays the effective traffic demand load offered for a given user profile. Multiple traffic grids are combined together to characterize a demand scenario for a certain time of day, as illustrated in Fig. 7.

Figure 8 summarizes the process described in this section, where multiple user distribution layers are first processed in order to estimate user profiles and dismember them into class layers that, combined together with a certain distribu-

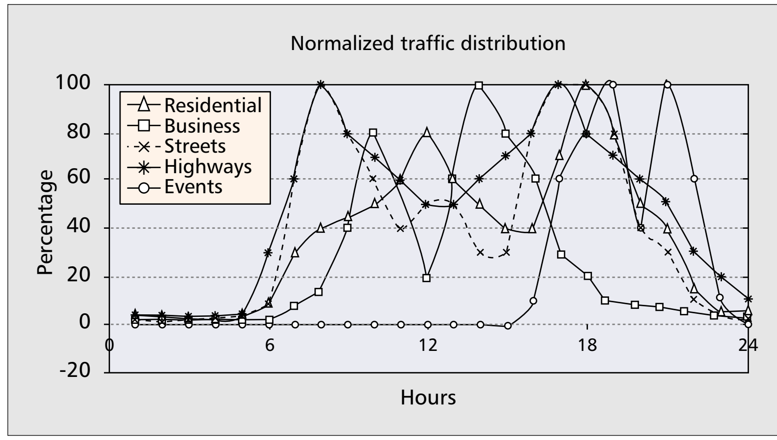


Figure 5. *An example of traffic distribution dynamics with time for each class.*

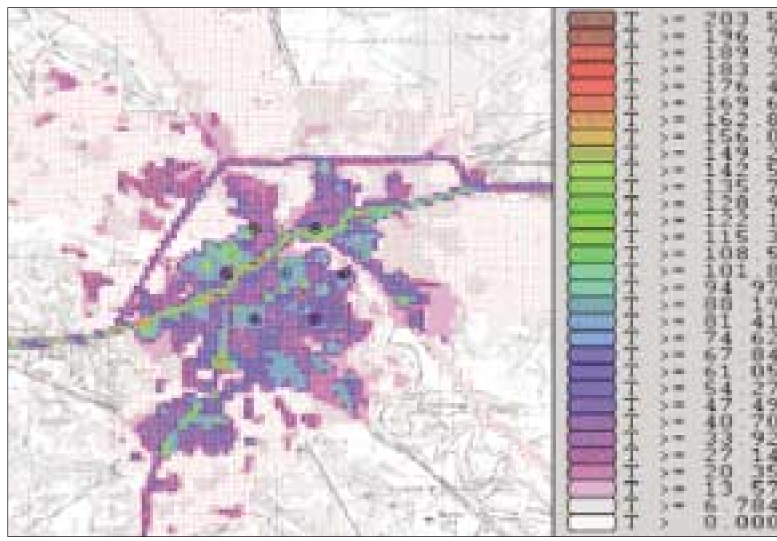


Figure 6. *An example of a one-layer traffic grid for a specific user profile and time of day.*

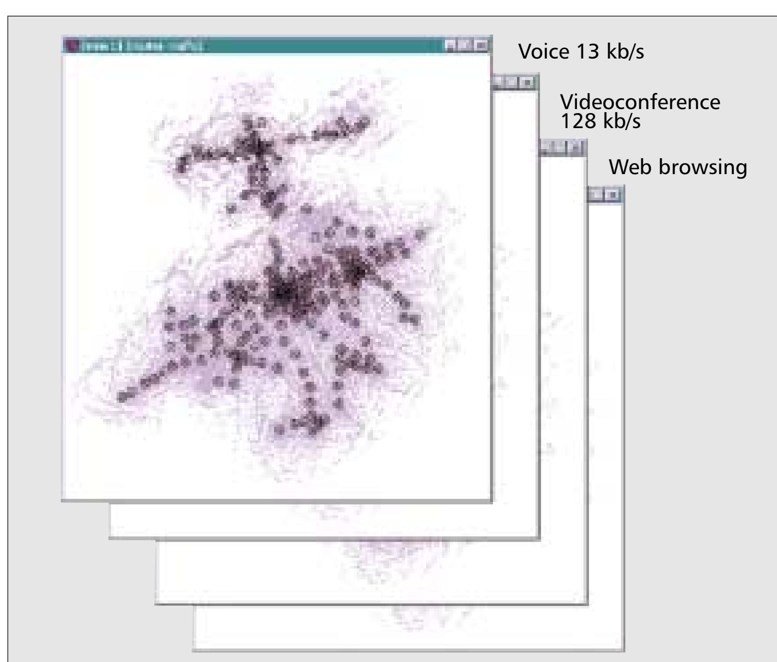
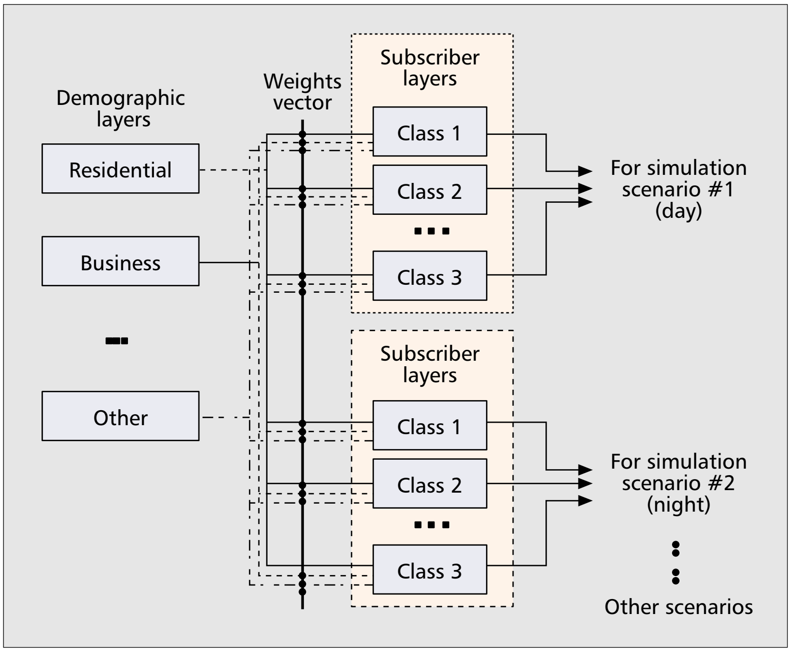


Figure 7. *An example of one demand scenario for a specific time of day, with multiple traffic (user) grids (one grid for each user profile).*



■ Figure 8. *A block diagram of demand database estimation.*

tion weighting, define one simulation scenario for the 3G system to be dimensioned.

Providing QoS in 3G Systems

Once a traffic grid is created for each user profile, as a combination of multiple demographic layers as described in the previous section, the next step is to cross this information with the cellular system design, considering cell locations and their coverage areas. Resource requirements assessment must of course consider the type of technology selected for the system, access methods, modulation schemes, and so on. System-level simulation is the most likely means to perform this assessment, due to the difficulty of conducting mathematical analysis that realistically models such complex systems. The focus of this section will be radio resource management.

The core of standards for terrestrial 3G systems uses CDMA technology for the air interface (WCDMA, cdma2000, TD-SCDMA, etc). In providing QoS for CDMA systems, there are important mechanisms that will define admission control, load control, and interference reduction in order to maximize system capacity.

Those mechanisms need to be considered in the simulation of the system when one wants to verify its response to different "user mixes" as described in the previous section.

In WCDMA systems, distinct data rates are supported on the traffic channels through the use of different spreading factors. Dedicated channels are usually allocated for real-time applications such as conversational and streaming classes, although full IP solutions for real-time applications are expected to be supported in the future. Non-real-time applications are handled as scheduled packet data over either common, shared, or dedicated channels.

In order to maximize capacity and at the same time provide QoS guarantees for the allocated services, radio resource management (RRM) needs to be in place. The main tasks related to RRM are:

- Power control: Used in the downlink and uplink to keep transmitting power at the minimum requirements that will satisfy the targets defined for each class (eventually updated by an outer loop).
- Load control: Based on received power or throughput.
- Admission control: Based on an estimate of the increase in total interference power on the uplink or total throughput being supported, the network estimates the noise rise (or throughput rise) that would result from accepting the requested connection. Admission control is performed on both downlink and uplink.
- Handover: In order to manage mobility, the system needs to support soft handoff between neighbor cells.
- Packet scheduling functionalities are also available to support non-real-time transmissions over packet data channels.

RRM functions can be implemented based on hard or soft blocking. In either case, the system will stop accepting new connections at some point in order not to sacrifice existing connections and planned coverage. Typically, soft blocking mechanisms allow more efficient system utilization.

One important step for verification of the system capacity is to include in the system-level simulation the same approach used in the real system for admission and congestion control. This includes the check on whether the system is exceeding the *target loading factors*. In this sense, the loading factor is one of the most important design parameters in CDMA systems, and maximizing capacity achievable for a given loading factor is a design objective.

The loading factor expresses the ratio of received interference divided by the sum of received interference plus thermal noise. The formula used to estimate the upstream loading factor for a multi-user multirate WCDMA system can be written as [8]

$$\ell = \left(1+\frac{o}{s}\right) \sum_{\text{All\_cell\_users}} \frac{1}{1+\dfrac{W}{\left(\dfrac{E_b}{N_0}\right)_j R_j \nu_j}}, \qquad (1)$$

where W is the chip rate, R_j is the data rate for user j, and E_b/N_{0j} is the required signal energy per bit divided by noise plus interference spectral density in order to achieve a certain quality of service for user j when connected at data rate R_j. ν_j is the activity factor for user j at the physical layer, and o/s represents other-cell-to-same-cell interference ratio as seen by the sector receiver.

From inspection of Eq. 1, we notice that different user mixes (proportion of users in different classes) would result in approximately constant total throughput values (given by the sum of all users' data rates) for a given loading factor if the required $(E_b/N_0)_j$ thresholds were all set to the same value. In other words, if the E_b/N_{0j} requirement for all classes were homogeneous, the total throughput capacity of a cell to achieve a desired loading factor would not depend on the proportions of service classes used to achieve that throughput.

On the other hand, in the more realistic scenario where different classes have different thresholds set for operation in terms of E_b/N_0 required, the total load will be given by a weighted sum of the different data rates times the required thresholds. This means that different total throughput per cell can be achieved for the same loading factor and network layout depending on the user mix that defines the demand offered to the system. That observation emphasizes the importance of appropriately describing the traffic demand on the system.

Considering that the total loading factor is used as a criterion for admission control, it determines the capacity of the cell as a design parameter that could then be appropriately calculated through simulation using as input the demand characterization framework described here.

Conclusions

In dimensioning traffic for CDMA systems, the main resource for allocation is the spectrum, that is, the operator needs to estimate and maximize the capacity that can be handled per installed carrier for a given network design. Assuming a system layout with given site locations and sector configurations, and using the methodology previously described to estimate the demand on that network, the most important dimensioning problem is to determine the amount of resources that need to be allocated at each sector in order to satisfy the demand while providing the required quality of service. In the event that the number of resources needed exceeds the spectrum availability (dictated by license agreements), a redesign of the network with addition of new cells and cell splitting is considered in order to satisfy QoS requirements. If the system design is for a new network layout, this process allows more flexibility, and many iterations are allowed in order to optimize site location and sector configuration.

Once simulation is performed reflecting the most relevant aspects of the system, different user mixes may be studied, and a better understanding of capacity performance expected from the system will be achieved.

We also argue that aggregate throughput within a cell is not by itself the metric of interest, since different user mixes will have an impact on the total rate the network is capable of supporting.

This is an ongoing research project, and the authors are in the process of exploring in further detail the impact of demand characterization on the design and optimization of 3G systems. A (primarily simulation-based) framework is being developed to facilitate the design of such systems.

This is an ongoing research project, and the authors are in the process of exploring in further detail the impact of the demand characterization in the design and optimization of 3G systems.

References

[1] T. G. Robertazzi, *Computer Networks and Systems: Queueing Theory and Performance Evaluation*, 2nd ed., Springer-Verlag, 1994.
[2] K. Park and W. Willinger, *Self Similar Network Traffic and Performance Evaluation*, 1st ed., Wiley, 2000.
[3] 3GPP, "Universal Mobile Telecommunications System (UMTS); QoS Concept and Architecture," 3GPP TS 23.107 v. 3.6.0, Release 1999, Apr. 2001.
[4] S. Marikar and L. A. DaSilva, "Resource Management in Third Generation Mobile Communication Systems Employing Smart Antennas," *Proc. INFOCOM 2002*, June 2002.
[5] S. Dixit, Y. Guo and Z. Antoniou, "Resource Management and Quality of Service in Third-Generation Wireless Networks," *IEEE Commun. Mag.*, Feb. 2001.
[6] L. Jorguseski *et al.*, "Radio Resource Allocation in Third-Generation Mobile Communication Systems" *IEEE Commun. Mag.*, Feb. 2001.
[7] L. Ribeiro and L. DaSilva., "Traffic Demand Characterization for Multimedia Mobile Networks," *Proc. IEEE VTC 2001-Fall*, vol. 2, Oct. 2001, pp. 620–24.
[8] H. Holma and A. Toskala, *WCDMA for UMTS — Radio Access for Third Generation Mobile Communications*, Wiley, 2000.

Additional Reading

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Trans. Net.*, Feb. 1994.
[2] Z. Sahinoglu and S. Tekinay, "On Multimedia Networks: Self-Similar Traffic and Network Performance," *IEEE Commun. Mag.*, Jan. 1999.

Biographies

LEILA Z. RIBEIRO (lribeiro@vt.edu) is a Ph.D. candidate at Virginia Tech at the Bradley Department of Electrical and Computer Engineering. She received her B.S.E.E. from the Federal University of Santa Catarina and her M.S.E.E. from the State University of Campinas, Brazil, in 1992 and 1995, respectively. She has worked for CelPlan Technologies, Inc. since 1993. She has also taught as an adjunct professor at George Mason University. Her research interests include wireless networks system design, traffic dimensioning, and performance optimization.

LUIZ A. DASILVA [SM] (ldasilva@vt.edu) joined Virginia Tech as an assistant professor at the Bradley Department of Electrical and Computer Engineering in 1998, after earning his Ph.D. in electrical engineering at the University of Kansas. He previously worked for IBM for six years. His research interests currently focus on performance and resource management in wireless mobile networks and QoS issues. He is currently involved in funded research projects in the areas of QoS interoperability and network policy, resource management mechanisms for the deployment of smart antennas in 3G systems, and mobile ad hoc networks, among others. He is a member of the Center for Wireless Communications (CWT) and associated faculty at the Mobile and Personal Radio Group (MPRG). He is a member of ASEE.