D. Grillo, R. A. Skoog, S. Chia, and K. K. Leung, "Teletraffic Engineering for Mobile Personal Communications in ITU-T Work -- The Need for Matching Practice and Theory", to appear in IEEE Personal Communications

Abstract

Guidance on traffic engineering for fixed networks and telephone service has been successfully developed by ITU-T (formerly CCITT). This guidance has been based on a communication paradigm which has included regulated operation environment, operation domains matched to national boundaries, centralized control and predictable service quality. These assumptions are being more and more challenged with the introduction of new communication modes, the increasing popularity of mobile and personal communications services, and deregulated operation. As a result, new paradigms are emerging for the 21st century telecommunications. In such a framework, continued user satisfaction and operator revenue growth with personal communications services require that suitable traffic engineering methods be devised. These methods should help in reconciling the expected service quality with cost-effective dimensioning and operation of networks and infrastructure for supporting a range of services as well as provide means for capitalizing on investments in traditional/existing telecommunication infrastructure. The paper notes that in order to arrive at sensible ITU-T Recommendations on traffic engineering for personal communications networks, practice and theory should be mutually supportive.

Teletraffic Engineering for Mobile Personal Communications in ITU-T Work -- The Need for Matching Practice and Theory

----

DAVIDE GRILLO, FONDAZIONE UGO BORDONI (+39.6.5480 3430, grillo@fub.it)
RONALD A. SKOOG, AT&T LABS (+1.732.949 7915, skoog@att.com)
STANLEY CHIA, AIRTOUCH COMMUNICATIONS (+1.925.210 3470, stanley.chia@airtouch.com)
KIN K. LEUNG, AT&T LABS (+1.732.345 3153, kkleung@research.att.com)

Traffic engineering has as its ultimate objective the cost-effective dimensioning of network resources to handle the user's demand for telecommunications services - and hence the induced user information and signaling traffic streams. At the international level, traffic engineering is being guided through the activity of ITU-T[1] (formerly CCITT[2]) where the related procedures are cast in Recommendations covering such issues as traffic characterization, service quality targets, dimensioning methods, and measurements. As opposed to Recommendations dealing with the procedural aspects of interconnection/interworking in a multi-vendor and multi-operator environment, Recommendations on traffic engineering provide advice on "good practice" for network operation and are not binding. However their recognized value is based on input from Administrations, ROA's (Recognized Operating Agencies) and scientific bodies who collectively provide a broad range of experiences and knowledge, and who have developed an understanding of the relationships (by a given technological scenario and communications paradigm) between such interrelated aspects as traffic demand, cost of communications infrastructure, sustainable service quality, and acceptable tariffs. The bulk of ITU-T work on traffic engineering assumes the existence of operation domains comprised within national boundaries for supporting telecommunications services between users located anywhere in the world. This work has been instrumental in enabling (national) operators to conduct business on a global scale, and has performed well in ensuring user satisfaction and continued revenue growth. In particular, traffic engineering for fixed communications, and PSTN-based services, has a long tradition in ITU-T. The underlying communications paradigm has been characterized, in addition to regulated operation and operators' domains matched to geographical boundaries, by centralized network control, fixed bandwidth allocation, and predictable service quality - among others. This paradigm which has held throughout the 20th century, is now being challenged, [1], since: i) its cornerstones are either no longer valid (spread of deregulated operation, ownership of network segments and Points-of-Presence beyond national boundaries) or, ii) new services and other operation modes are flanking the traditional ones (e.g. multi-party communications, highly variable

---

[1] *International Telecommunication Union Telecommunication Standardization Sector.*

[2] *International Telegraph and Telephone Consultative Committee.*

bandwidth requirements, on-demand bandwidth allocation, terminal and personal mobility support[3]).

An intensive study activity is being undertaken to understand and model the key aspects of the emerging telecommunications systems associated with this changing scenario so as to arrive at a sensible traffic engineering methodology for them and continue to successfully operate telecommunications services under the developing paradigms. The progress of these studies in ITU-T shows a variegated picture, with areas where investigation is approaching saturation complemented with substantial traffic engineering standardization activity (ISDN, ATM), areas where traffic engineering is progressing (IN and signaling systems), and areas where traffic engineering has been started but progress is not yet adequate to the advances in system deployment and service penetration (mobile networks).

Personal communications[4], plays a vital role in shaping the 21st century communications paradigms. Indeed, the associated dimensions of space/time dependence of traffic demand, "hostile" operation environment, unpredictable quality, and changing user's network attachment point represent major deviations with respect to the traditional communications paradigm. Although mobile services have been commercially available since the late seventies, traffic engineering for personal communications has been based - and to a great extent still is - on "current practice" of mobile operators, and has been dominated until recently more by radio transmission and coverage considerations rather than by classical traffic loading and service quality arguments. The justification for this has been that an (elite) customer base has traded-off impairments of service quality against ubiquitous service that is perceived as a major value. In addition, there are many factors which have been hiding the inefficiency of network planning for operators. As a matter of fact, to circumvent poor network quality, users have been inclined to develop a pattern of complex avoidance behavior in order to work around the perceived network congestion and impairment. Furthermore, in order for an operator to meet both wide-area and in-building coverage requirements of network roll-out, a large spare capacity has been implicitly built into the network from an early day of operation. This process is further exaggerated by the large "build-ahead" margin adopted by most operators which is necessary to meet the predicted intensive subscriber growth and to alleviate long delay in site acquisition for building new base sites. Finally, infrastructure vendors have introduced many temporary capacity relief features such as "directed retry", "cell load sharing", [2], and "queuing" to dynamically exploit the (temporary) unused resources due to short-term traffic fluctuations[5]. Naturally, all these processes and techniques have adverse implications to the cost and efficiency of a network.

In the last decade mobile services have been experiencing an accelerated penetration which has culminated in the explosive growth to mass market dimensions over the last few years, with outlooks for continued growth. Accordingly, mobile-related traffic is forecast to be comparable in volume with that related to fixed networks in a not too distant future. Even with the prediction of such a strong growth in the mobile market, increasing competition in service provision has meant that mobile operators have to be more prudent in their cost-management and vigilant in maximizing their income. While over-dimensioning of a network is equivalent to poor capital investment, congestion at busy hours could mean lost calls and lost revenues. These factors, combined with the impact that mobile-related traffic may have on the fixed infrastructure, and the convergence of mobile and fixed services, drive towards a rationalization of the resource allocation and management procedures (both inter- and intra-operators' domains) and make it urgent to address traffic engineering for personal communications at the international standardization level. Furthermore, with the introduction of mobile data services such as General Packet Radio Services (GPRS) for GSM and IS-707 for cdmaOne (also known as IS-95, [4]), the integration of voice and data traffic is becoming a reality. This adds yet more dimensions to the complexity of traffic engineering. Table 1 tries to illustrate the challenges in the planning and operation of personal communications networks and the expected benefits from a traffic engineering activity.

To accommodate the need for standardization on traffic engineering for personal communications, ITU-T

---

[3] *Personal mobility refers to the ability of users to have flexible access to telecommunication services from any terminal, fixed or mobile, to meet the users requirements. These requirements may then be relocated from terminal to terminal. Personal mobility involves the network capability to locate the user on the basis of a unique personal telecommunication identity (i.e., "personal" number) for the purposes of addressing, routing and charging of the users calls.*
Terminal mobility involves the ability of the user to be in continuous motion whilst accessing and using telecommunication services and the capability of the network to keep track of the user's terminal. This requires the telecommunication services to be available as the terminal moves within the radio coverage and ideally at all times (ITU-T Recommendation I.114, [3]).

[4] *The term personal communications will be used to refer to the collection of terminal (PCS, Personal Communications Services) and personal (UPT, Universal Personal Telecommunication) mobility services.*

[5] *"Directed retry" means a terminal is redirected by the network to setup a call with a base station which is not the best server due to network congestion reasons. "Cell load sharing" is to dynamically move the boundary of the cells so that calls in progress can continue in cells which are less congested. "Queuing" is used to temporarily buffer calls e. g. on a dedicated control channel when all the traffic channels are congested.*

TERMINAL MOBILITY		
Key Aspects & Problems	**Requirements & Solutions**	**Aspects of Traffic Engineering & Challenges**
Limited amount of radio spectrum	Spectrum re-use (cellular lay-out architecture) to meet capacity requirements. Interference management, dynamic channel allocation (DCA), use of smart antennas	Spectrum partitioning between cell layers, considering traffic overflow to "umbrella cells" (or mutual overflows in non-hierarchical cell lay-outs). DCA performance models.
Hostile transmission environment for wireless communications	Channel quality monitoring and recovery to combat adverse transmission conditions. Smart antennas, advanced equalization, handover and admission control, high frequency re-use.	Handover (combining) handling and priority service; Call Admission Control for trading off service quality against carried traffic; modeling of co-channel interference as dependent on traffic load.
Mobility behavior of the customer base	Location, registration and authentication of the customer base to track users and provide a seamless communications space shielding from mobility implications and roaming technicalities.	Mobility models. Space/time traffic demand dependence modeling, mapping user mobility into signaling traffic. Signaling network dimensioning.
Long term subscriber growth, intensive in-building coverage and wide-area coverage	Minimum infrastructure build and just-in-time delivery of capacity.	Description of the traffic characteristics and modeling of the traffic processes. Generic dimensioning methods complemented with network traffic dimensioning for specific mobile technology classes.
Traffic demand fluctuation	Capability of moving spare resources around in the network to accommodate periodic and instantaneous traffic variations.	Traffic modeling for dynamic trunking of shared resources.
Integrated voice and data services	Providing additional resources for data services on a dedicated basis or shared basis.	Service quality requirements for data services. Traffic dimensioning for packet data.
PERSONAL MOBILITY		
Key Aspects & Problems	**Requirements & Solutions**	**Aspects of Traffic Engineering & Challenges**
Registration on and interaction from current user equipment	Authentication of user data, negotiation capabilities to cope with actual equipment characteristics/performance, impact of actual service provider offerings and visited network constraints.	Determination of traffic mix resulting from user-network negotiation actions. Signaling network dimensioning.
Interaction with and personalization of the user profile	Manipulation capability of user data to provide a familiar (VHE, Virtual Home Environment) and user-friendly communications space.	Modeling IN services; mapping user mobility into signaling traffic. Signaling network dimensioning.

Table 1. *Key aspects and problems in personal communications and related traffic engineering implications.*

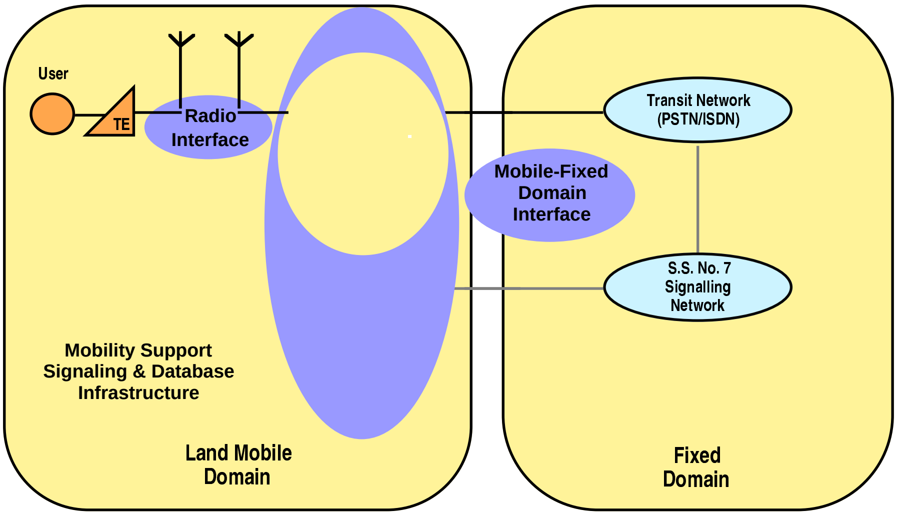
has started a dedicated series of Recommendations, the E.750 series, **[5]**. Having first addressed framework setting aspects for a traffic engineering activity - such as reference configurations and GOS[6] parameters, and target values - the E.750 series is now giving priority to studies on traffic demand modeling and dimensioning methods.

In parallel with the increasing penetration of personal communications, a range of related studies on architecture, service offerings and performance aspects have been proliferating in the open literature, also fueled by the international efforts aimed at designing advanced, "third generation" mobile systems such as IMT-2000, **[7]**, **[8]**, developed in ITU and UMTS, **[9]**, **[10]**, studied in ETSI. Although these studies do not necessarily exhibit the combination of simplicity and coverage of key aspects which matter for traffic engineering, and frequently favor the analysis of specific technical solutions rather than relate fundamental parameters, they have a great potential for traffic engineering work. The current operators' practice and this wealth of literature are, in a sense, two extremes that traffic engineering (and related ITU-T activity) has to reconcile in order to provide a sensible guidance for a cost-effective use of network resources while meeting users expectation on service quality.

Although personal communications implies both terminal and personal mobility, this note concentrates on terminal mobility for which Fig. 1 gives two fundamentally different supporting architectures. As shown in the figure, the key difference lies in the organization (signaling and database arrangement) of the mobility management functions resulting in "stand-alone" (or separated mobile and fixed network) and in "integrated" architectures. The interfaces and functionality shown in the figure, in addition to having

---

[6] *Grade of Service (GOS): "A number of traffic engineering variables used to provide a measure of adequacy of a group of resources under specified conditions; these grade of service variables may be probability of loss, dial tone delay, etc.", ITU-T Recommendation E.600,* ***[6].***



logical significance, also indicate the scope for teletraffic engineering.

The thread followed in the paper is as follows. Initially, the specifics of mobile related traffic demand are briefly introduced. Obviously, these specifics are key to the whole traffic engineering process. Since mobile services have been supported before a systematic traffic engineering activity had been initiated in ITU-T, the current practice on which operation of mobile networks is based is then described. This description highlights the relationship between the dimensions associated with radio transmission and network planning, and considers some typical actions to be taken for estimating and accommodating the traffic demand assuming that the radio coverage problem is solved. Subsequently, the organization of the E.750 series is briefly reviewed and it is indicated that the underlying traffic modeling and dimensioning studies try to provide tools to be used for rationalizing key phases comprising the operator's practice. In the face of the scope of teletraffic engineering as envisaged in the E.750 series, a selection of representative theoretical contributions addressing both mobility and traffic demand modeling as well as typical traffic engineering issues is surveyed. The survey is intended to provide an overview on the kind of theoretical support currently available for the standardization activity. The paper finishes by listing and commenting on some commonly accepted assumptions underlying theoretical work, and stating the ITU-T needs for progressing a useful standardization activity.

Specifics of Mobile Related Traffic

In a sense, user mobility has always been considered also for fixed networks and telephone traffic. As far as the origination and destination of calls is concerned, a fixed network may be viewed as a collection of communication devices rigidly associated with network attachment points. "Fixed" users, i.e. users of fixed devices, may change their location over a specific area, for example when moving from home to the working place and vice versa, and then initiate (place or receive) calls from the fixed device location they have reached.

For traffic engineering purposes, the traffic load during "busy hours" for a geographic region is often referred to. The location of the busy-hour during the day depends on, for example, whether the region is used as a business or residential area. Although users move from one place to another, the peak-hour traffic load implicitly considers such mobility. This, combined with the condition of sufficiently large ("infinite") population has led to models of the fixed traffic demand based on characterizing call arrival statistics (e.g., exponential law between call arrivals) that capture the aggregate behavior of the users in a geographical area.

To dimension transmission, switching and processing resources bound to a geographical area, the other two needed elements are the distribution of the call length and the arrival rate of calls during the reference ("busy hour") period. Again, under the above conditions, the call length distribution - at least in the case of telephony - has been found to be only dependent on aggregate models of user "behavior" which has resulted in a robust, parametric and universal model with very little dependence on user class, country, etc. In conclusion, although the basic assumptions underlying traffic engineering of fixed networks are continuously challenged and tested (for example see [11], [12] and [13]), for any practical purpose the fixed traffic demand can be expressed in

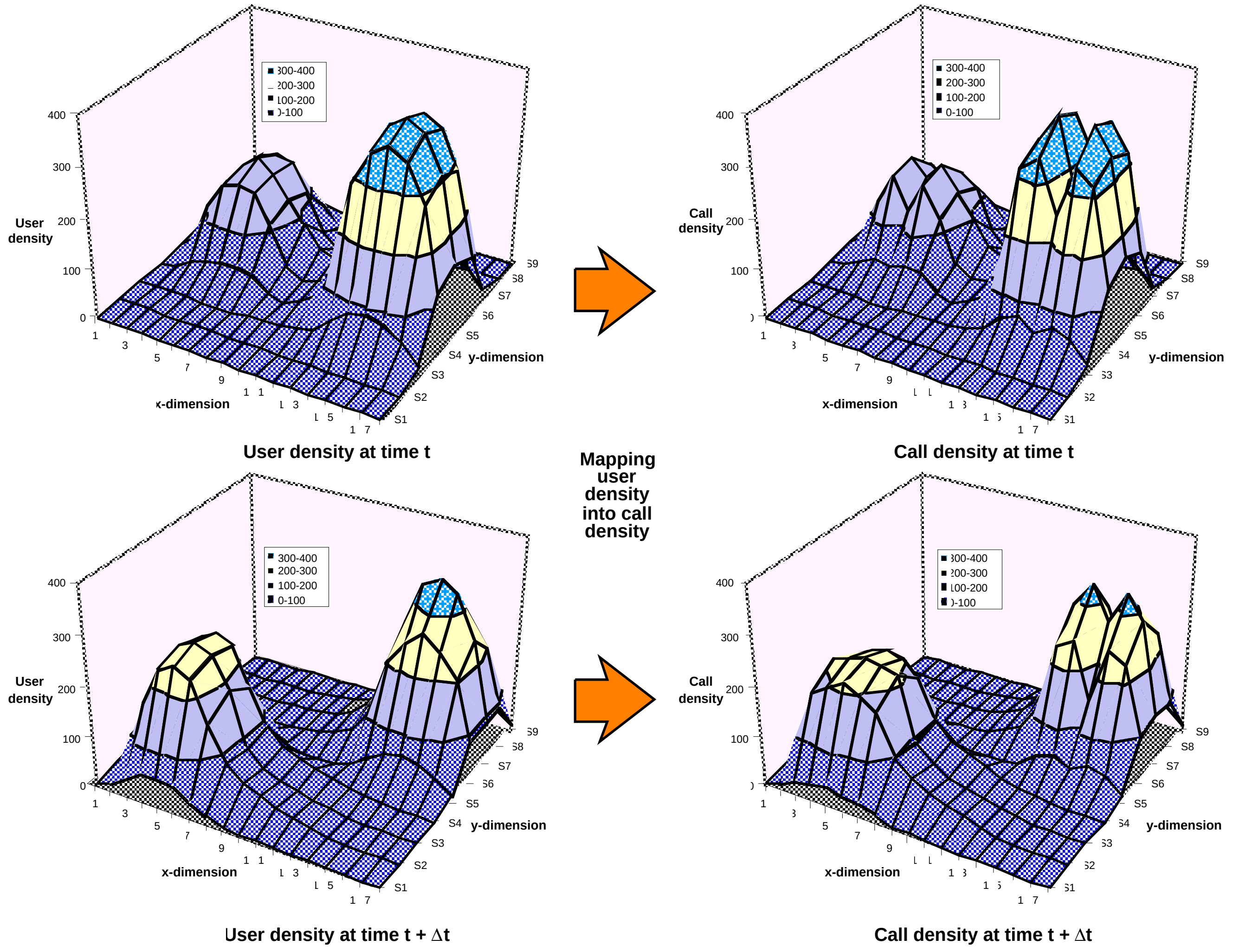


Figure 2. *Illustrative example of distribution of user density at different times and related terminal mobility traffic demand for a generic service.*

terms of the distribution of call inter-arrival times, the call arrival rate and the distribution of call duration, with virtually no dependence on space per se and a dependence on time dictated by the general level of daily and/or seasonal activity.

By contrast, as concerns the initiation of calls, a mobile network may be viewed as a collection of communication devices in no rigid association with network attachment points (radio ports), and normally traveling together with the (mobile) users. To appreciate the specifics of mobile related traffic one has to note that:

- The user mobility is no longer confined to possibly traveling within a few places with long intervals of stationariness (as in the case of fixed users), but is generally characterized by wider range user movements and more frequent location change;
- The association between calling device and network attachment point is dynamic (in-between call attach/detach) and may also change during the same call (in-call association re-arrangement due to handover or combining);
- The resources to be dimensioned (e.g., radio channels carrying user and signaling traffic, and the fixed infrastructure for supporting mobility) continue to be bound to a geographical area.

As a consequence, the underlying user behavior has a higher traffic impact (both in space and in time) than in the fixed network case. For example, the distribution of population density may be stationary, but calls are initiated while users are "on the move" thus giving rise to mobility related traffic loads and "traffic volatility", i.e.: i) the call initiation sites are scattered and dynamically changing over a geographical area; ii) bandwidth associated with a connection may have to be provided to different sites throughout the call, viz., changing radio cells during a call.

For terminal mobility traffic, Fig. 2 illustrates an example of the variation of the user density in space and time over a geographical area. The same figure also suggests that the mapping of the user density into traffic demand may not be simply a matter of scaling but may require more elaborate considerations since the call initiation rate may not be linearly dependent on the user density. For example, a high user density as a result of a traffic jam may also trigger users to place calls at an extraordinarily high rate. Moreover, in the case of cellular systems this mapping may be impacted by deep fading and/or insufficient radio coverage causing tides in the distribution of the traffic demand not present in the user density. In fact, the traffic demand in such systems may be conditioned (to a varying degree) by the provisions made by the network operators and service providers (e.g. deployment of communication facilities, roaming agreements, tariff structure, marketed service features, etc.). This conditioning, however, shall not impact the logical framework for carrying out the traffic engineering activities.

Finally, the in-call behavior of mobile users does not necessarily align with that related to the fixed network, for example as concerns the average length and length distribution of the calls. This area is the object of intensive study and investigation, **[14].**

Current Practices for the Operation of Cellular Systems

As engineering procedures for mobile systems have only recently started to be standardized, the operation of cellular systems is frequently based on simple rules for traffic demand estimation and resource allocation, complemented with monitoring and tuning the system performance in the field as the network evolves, **[15].** To illustrate this process, it is instructive to assume a "green field" situation, although in many instances similar issues could be faced by operators who are still evolving their infrastructure. Key aspects are summarized in Fig. 3. In the figure, a simplified view of the complex relationship between the radio planning and the capacity dimensioning process is also shown. In addition, a numerical example of the dimensioning process is shown in Table 2.

The dimensioning process starts with an estimation of the user population by using the density of the inhabitants in a specific area together with an anticipated service penetration rate. Radio and network planners continue with the identification of sites where the cellular infrastructure has to be laid down (typically, base transceiver stations, base station controllers and mobile switches), and the mapping of user density into traffic demand. The process then finishes with the allocation of the traffic (radio) channels making judicious use of the available spectrum[7].

The accomplishment of the dimensioning cycle requires that numerous optimization problems be solved. These range from minimizing the number of the base sites while guaranteeing sufficient coverage and acceptable service quality, to planning the re-use of spectrum so as to accommodate the traffic demand while ensuring stable system operation and user satisfaction - to name just a few.

---

[7] *It should be noted that the ultimate traffic capacity of a base site is highly dependent on the exact locations of the base sites. While every effort is usually made during the planning stage to ensure that a base site can be well positioned, the actual site acquisition process is subject to many factors including the physical location, real estate cost, the height of the location, the availability of equipment and antenna space, etc.*

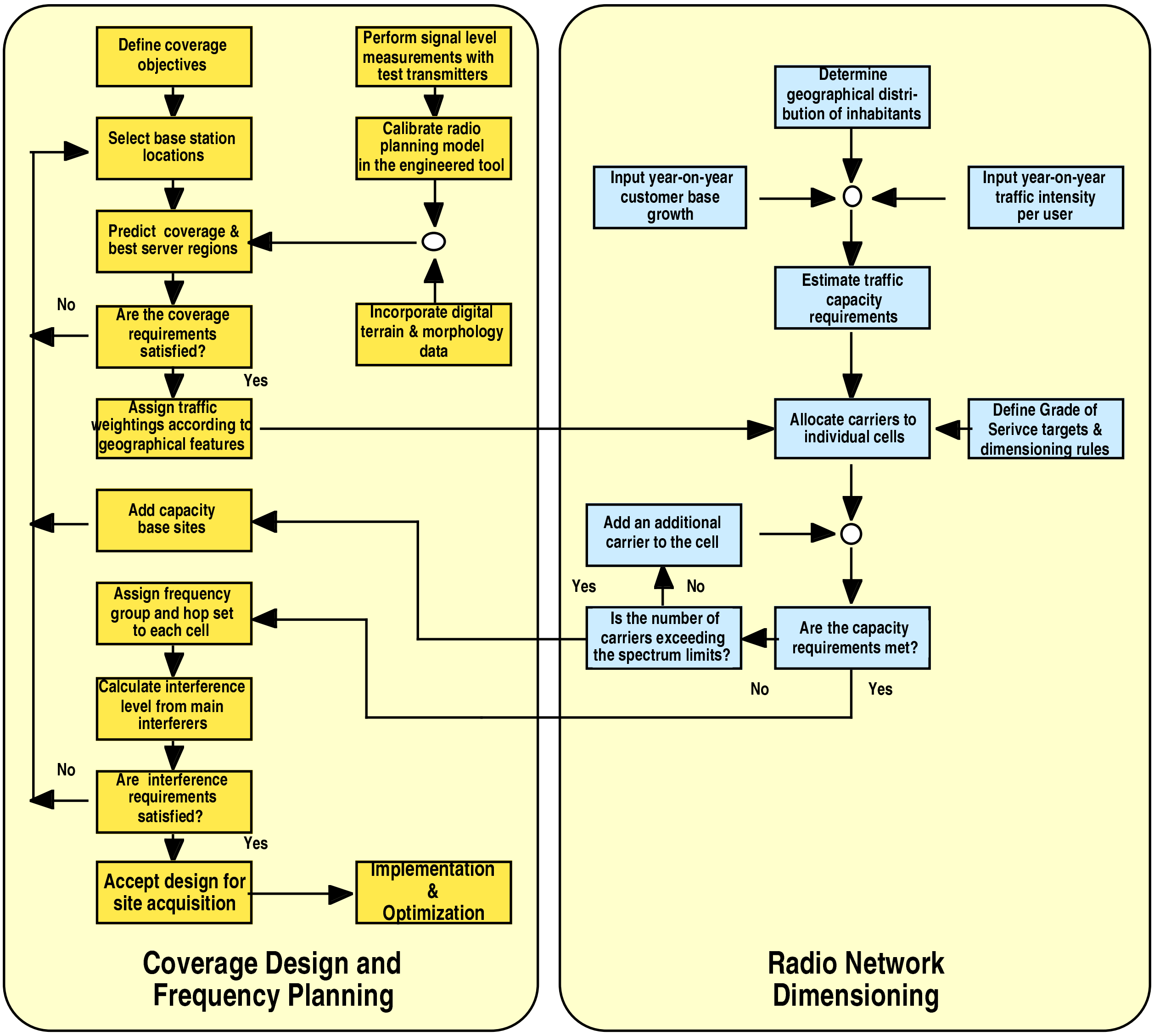


Figure 3. *A summary of the radio planning and capacity dimensioning processes.*

The success in operating a system is assessed, among others, by the degree of control exercised over such phenomena as dropped calls, repeated call attempts and handover cases. All these phenomena will penalize user expectations about good service quality and, more often than not, many of these shortcomings are due to poor balancing of operation parameters.

The following sections are intended to illustrate the current approach to the operation and management of cellular systems (together with some deficiencies). They provide a reference list of key issues to be covered in future standardization work on traffic engineering in ITU-T.

Figuring out the Traffic Demand

For traffic engineering of cellular systems, information on geographical population distribution is of vital importance to an operator. For new entrants to a market, this information can only be estimated using published census information. This may vary in resolution, with the better ones being rather detailed, and may resolve down to municipal or district level. From the census database and the size of the geographical area, it is possible to estimate the population density for the location. Together with the year-on-year user penetration forecast and the average traffic intensity per subscriber, the traffic demand can be obtained. (Occasionally, road traffic information may also be available. This information is either too detailed or very specific, which renders its use difficult. The reason is that this information is collected usually in the form of number of vehicles passing a specific junction per 24 hour period rather than the number of vehicles using the road. As urban roads could have many turn-offs, extrapolating the information to give traffic volume for a coverage area is by no means a simple task).

In parallel with the traffic engineering process, radio coverage planning is also performed to enable network infrastructure roll-out. Based on the terrain database and the morphology database together with the desired signal level necessary to provide suitable in-building and outdoor services, base site locations are identified. Frequently, contiguous coverage is required and, hence, the coverage area of base sites are packed closely together in order to eliminate coverage gaps as far as possible. In reality, the terrain is rather undulated. In order to eliminate the majority of the coverage gaps and to provide adequate in-building and in-vehicle services, one needs to significantly overlap the coverage between base stations. Thus the dividing lines defined by the equal signal level from two or more base sites will form the boundary of the "best server" region for individual base sites. In other words, when a mobile is within the best server region of a specific base site, it will receive the strongest signal from that base site even though the signal from other base sites may still be adequate for communications. By associating mobile stations with the base site of a best server region, the highest downlink carrier-to-interference ratio can be obtained. In

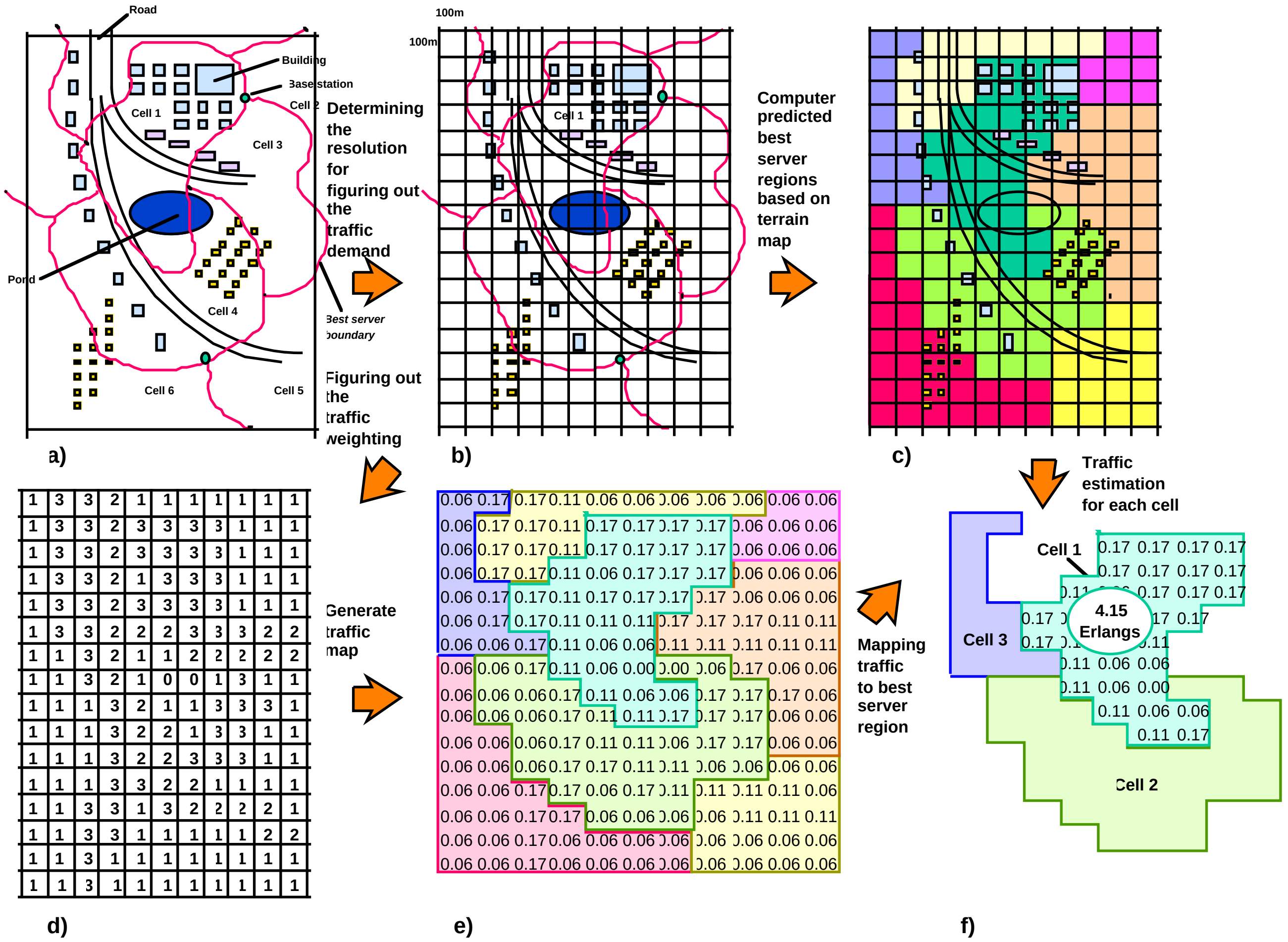
		Year 0 through 10	Year 0	Year 5	Year 10
Traffic demand estimation					
Population		1,000,000			
Penetration			10%	20%	50%
Number of subscribers			100,000	200,000	500,000
Traffic per subscriber	E		0.02	0.012	0.012
Traffic demand	E		2,000	2,400	6,000
Coverage design					
Size of service area	km^2	1,000			
Traffic distribution		Uniform			
Nominal cell radius	km	2			
Nominal cell area (assuming circular)	km^2	12.6			
Number of base sites for coverage		80			
Number of cells / site (3-sector sites)		3			
Frequency planning and network dimensioning					
Spectrum allocation	MHz	7.4			
Radio carrier spacing (as an example)	MHz	0.2			
Number of radio carriers		37			
Nominal frequency reuse		12			
Average number of carriers / cell		3			
Average number of voice channels / cell (8 channels per carrier, excluding control channels)		22			
Call carrying capacity per cell (2% blocking, Erlang B)	E	14.9			
Capacity dimensioning					
Total call carrying capacity	E	3,575			
Average base site traffic efficiency due to terrain)		60%			
Actual call carrying capacity	E	2,145			
Spare capacity (negative means insufficient capacity)	E		145	-255	-3,855
Additional capacity					
Additional capacity	E		0	255	3,855
Limiting factor			Coverage	Capacity	Capacity
Number of additional base sites			0	6	87
Adjusting for base site traffic efficiency			0	10	145
Total number of base sties			80	90	225
Infrastructure increase			-	12.50%	181.25%

Table 2. *A simple numerical example of the radio planning and capacity dimensioning process.*

addition, when power control is used in the uplink, a minimum transmit power is required which could in turn prolong battery life of the handsets.

As the best server regions are rarely regular in shape due to terrain undulation and geographical features, the traffic capture ability of each base site could be quite different. By mapping the best server region into the population density map, a first order estimation of the traffic demand per sector can be realized. However, this may lead to substantial inaccuracies since users are generally congregated along roads and buildings and rarely located in open spaces. A method to obtain a more accurate estimation of the traffic demand is to polarize the population into areas where it is most likely to be located.

This can be achieved by assigning weightings to different geographical features. Based on these weightings, the traffic for each cell can be more accurately estimated. Evidently, for a cell which contains lots of open spaces, the amount of traffic is expected to be very low. By contrast, for a cell which contains buildings and shopping areas, the traffic density is expected to be high. Fig. 4 tries to illustrate this principle. Specifically, Fig. 4a shows the geographical features and the best server regions of an area under study. Fig. 4b shows a 100 x 100 m^2 grid overlay on the area. This represents the resolution of the digital terrain database which will eventually determine the resolution of the best server map as well as the traffic distribution map. It should be noted that both the best server map and the traffic maps now become an approximation of the real life situation. Assuming that census information indicates that this 1.1 x 1.6 km^2 area has a population of 17,600 people, this corresponds to a



1	3	3	2	1	1	1	1	1	1	1
1	3	3	2	3	3	3	3	1	1	1
1	3	3	2	3	3	3	3	1	1	1
1	3	3	2	1	3	3	3	1	1	1
1	3	3	2	3	3	3	3	1	1	1
1	3	3	2	2	2	3	3	3	2	2
1	1	3	2	1	1	2	2	2	2	2
1	1	3	2	1	0	0	1	3	1	1
1	1	1	3	2	1	1	3	3	3	1
1	1	1	3	2	2	1	3	3	1	1
1	1	1	3	2	2	3	3	3	1	1
1	1	1	3	3	2	2	1	1	1	1
1	1	3	3	1	3	2	2	2	2	1
1	1	3	3	1	1	1	1	1	2	2
1	1	3	1	1	1	1	1	1	1	1
1	1	3	1	1	1	1	1	1	1	1

0.06	0.17	0.17	0.11	0.06	0.06	0.06	0.06	0.06	0.06	0.06
0.06	0.17	0.17	0.11	0.17	0.17	0.17	0.17	0.06	0.06	0.06
0.06	0.17	0.17	0.11	0.17	0.17	0.17	0.17	0.06	0.06	0.06
0.06	0.17	0.17	0.11	0.06	0.17	0.17	0.17	0.06	0.06	0.06
0.06	0.17	0.17	0.11	0.17	0.17	0.17	0.17	0.06	0.06	0.06
0.06	0.17	0.17	0.11	0.11	0.11	0.17	0.17	0.17	0.11	0.11
0.06	0.06	0.17	0.11	0.06	0.06	0.11	0.11	0.11	0.11	0.11
0.06	0.06	0.17	0.11	0.06	0.00	0.00	0.06	0.17	0.06	0.06
0.06	0.06	0.06	0.17	0.11	0.06	0.06	0.17	0.17	0.17	0.06
0.06	0.06	0.06	0.17	0.11	0.11	0.17	0.17	0.17	0.06	0.06
0.06	0.06	0.06	0.17	0.11	0.11	0.06	0.17	0.17	0.06	0.06
0.06	0.06	0.06	0.17	0.17	0.11	0.11	0.06	0.06	0.06	0.06
0.06	0.06	0.17	0.17	0.06	0.17	0.11	0.11	0.11	0.11	0.06
0.06	0.06	0.17	0.17	0.06	0.06	0.06	0.06	0.11	0.11	0.11
0.06	0.06	0.17	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
0.06	0.06	0.17	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

KEY

a)	Geographical area with a population of 17600 and a predicted traffic of 17.6 Erlangs
b)	Overlaying a grid of 100x100 m^2 on the geographical area for traffic weighting
c)	Signal level computer prediction of the best server region
d)	Traffic weighting map for the geographical area
e)	Traffic distribution map based on the traffic weighting
f)	Predicted traffic for individual cells in the geographical area

Figure 4. *Estimating traffic demand from population distribution, city lay-out and radio coverage arrangements.*

population density of 1 person per 100 m^2. With a service penetration rate of 5%, there will be 880 subscribers in the area. Given that each user will generate 20 mErlangs of traffic, the total traffic in the area amounts to 17.6 Erlangs. This corresponds to an average traffic density of 1 mErlangs per 100 m^2. As the resolution of the digital terrain database is accurate to 100 x 100 m^2, computer prediction of the signal level for the best server region will be quantized into bins of the same size.

Feature	Weight
Road	2
Open	1
Water	0
Road and building	3
Open and road	2
Open and building	3
Open and water	1

Table 3. *Weighting factors for the offered traffic in relationship to the geographical features.*

This is shown in Fig 4c. Finally, also assume some weighting factors for the traffic load in relationship to the geographical features as represented in Table 3.

By mapping the best server region to the traffic bins, the traffic demand for each cell can be calculated. Applying the weighting to the geographic area as shown in Fig 4b, a weighting map as shown in Fig 4d can be obtained. Summing the total weights in the area and knowing the total traffic, the traffic demand for each individual bin can be apportioned as shown in Fig 4e. Finally, mapping the best server region to the traffic map, the traffic prediction for each cell can be obtained, Fig. 4f.

To show the importance of using the weighting factor, consider the demanded traffic of Cell 1. Without the weighting factor, a traffic load of 3.3 Erlangs is predicted. However, with the weighting, a traffic load of 4.2 Erlangs (30% higher) can be anticipated. It should be noted that the weighting factors shown in this example are indicative and for a real application more calibrations are necessary to ensure good accuracy. The boundary effects between geographical areas will also have to be accounted for but this is beyond the scope of this discourse.

Evidently, the example shown here is related to a green field deployment where an operator has no "a priori" knowledge of the actual traffic and mobility pattern of the users. As the network evolves, with traffic statistics collected through the mobile switches over time, a much more accurate picture of the traffic distribution down to the resolution of a cell can be built. With this information, the above technique can be applied again to refine the network optimization. Furthermore, as the radio network is always dimensioned for the peak traffic together with a safety margin, variations of the daily traffic due to the mobility pattern of the users are usually adequately taken into account as well as extraordinary events such as major incidents or scheduled gatherings. In the latter instances, operators have to employ special measures to cope with the surge in traffic demand. An example of these temporary measures is the "Cell Site on Wheels" deployed by an operator following the Los Angeles earthquake in 1994.

Sizing the Channel Capacity of Cell

Based on the knowledge of the traffic for each cell, the number of traffic channels - and hence of transceivers - can be determined. Using the GSM system as an example, the relationship between offered traffic and the number of transceivers is indicated in Table 4. Specifically, for GSM one transceiver supports one carrier which in turn supports eight time slots. The time slots can be assigned as traffic channels or control channels depending on the specific configuration of a network. (The results in Table 4 consider the requirement of control channel assignment and assume a 2% probability of blocking for fresh calls using the Erlang B model).

Offered traffic per cell [Erlangs]	Number of traffic channels (or time slots)	Number of transceivers
2.3	6	1
8.2	14	2
14.9	22	3
18.4	26	4

Table 4. *Examples of allocation of transceivers to a cell as a function of offered traffic.*

As a network is normally dimensioned for growth and traffic fluctuations, it is not uncommon that a three to six months build ahead is incorporated in the traffic dimensioning plan in order to accommodate safety margins, **[16]**. These margins have proven to be adequate in most cases.

In addition to normal calls, handover requests also require radio resources especially for a "make-before-break" scheme as in some implementations of the GSM system, **[17]**. In a real network, the number of handovers per call is dependent on the length of the calls as well as the mobility pattern of the users. The contribution of handovers to the total traffic loading is difficult to predict and is normally assumed as an acceptable overhead to a system. When an operator detects an exceptionally high volume of handovers in a cell, measures will have to be taken to bring it under control. Typical techniques at the disposal of an operator are to: Increase the hysteresis margin; change the handover thresholds; and reduce the overlap between adjacent cells by using narrower beam antennas. As an example, 60° antennas are frequently used among three sector cell sites in dense urban areas for minimizing the number of repeated handovers. Otherwise, the small cell size and large coverage overlap between neighboring sectors coupled with highly variable shadow fading in an urban area might trigger an excessive number of handover requests.

	Handover Triggering Reason			Outcome Of Handover Handling			
HO's per cell	Signal criterion	Bad UL quality	Bad DL quality	Success		Failure	
				Success	Failure: Total	Failure: Returned to old channel	Failure: Dropped calls
	[%]	[%]	[%]	[%]	[%]	[%]	[%]
377	77.19	16.18	6.63	46.95	53.05	39,52	12.53
407	66.34	15.23	18.43	86.49	13.51	11,3	2.21
481	70.48	10.81	18.71	84.41	13.59	12,47	3.11
1661	90.01	5.72	4.27	52.44	47.56	46,42	1.44
648	93.21	2.01	4.78	54.94	45.06	44,6	0.46
865	56.65	22.31	21.04	86.01	13.99	9,595	4.39
618	54.53	8.41	37.06	83.33	16.67	13,27	3.39
314	37.90	3.50	58.60	86.94	13.06	9,554	3.50
583	83.53	7.20	9.26	89.02	10.98	3,087	7.89
302	76.16	18.87	4.97	50.33	49.67	47,02	2.64

Table 5. *Examples of classification of handover cases for a set of ten cells with exceptionally high volume of handover (courtesy of AirTouch International).*

The ability to prioritize handover handling with respect to fresh calls is infrastructure equipment specific. In dense urban environments where cells are heavily overlapped with each other in outdoor areas due to their short inter-site distances and the requirement for indoor coverage, it is generally not necessary to prioritize handover handling as calls can be maintained even if the mobile moves away from the best server region. This large overlap can be visualized by understanding that the indoor penetration margin for urban areas is typically required to be 15 dB or more. A handover failure will not normally lead to a dropped call as calls, which are not successfully handed-over, can be reverted back to the originating base station. Evidently this will degrade the service quality due to the increased uplink interference as calls are dragged outside the best server region.

Similar to handover prioritization, the handover algorithm and parameters are typically infrastructure equipment specific and are usually omitted from standardization. For instance, for GSM, it is possible to employ received signal level, received signal quality, timing advance as well as traffic reasons for initiating handovers, **[18]**. For a properly engineered network, the majority of the handovers should be initiated because of insufficient signal level, i.e. the system operation should be power-limited. A high volume of handovers due to poor received signal quality normally indicates that there are interference problems in the network (interference-limited operation). In this situation, the dropped call rate is expected to be high as well. An operator will have to minimize this by re-tuning the frequency plan and optimizing the base sites in the vicinity of the affected region.

As an example, Table 5 shows a break-down of handover cases according to initiating reason for a GSM system. In the table, "signal criterion" signifies that the handover is triggered by insufficient signal level, whereas "UL quality" and "DL quality" stand for handover due to excessive error rate on the uplink and downlink, respectively. In GSM, signal quality is directly related to the bit error rate measured prior to channel decoding. There are eight quality levels defined and the threshold for poor quality is between 5 and 6. Evidently, all handover thresholds are operation parameters. As shown in the table, not all unsuccessful handovers lead to dropped calls, many are reverted back to the original channel where a repeated handover request may be initiated.

As it may be observed from the table, in an evolving cellular network, on the average about 25% of the handover cases could be due to interference impairments though the scatter around this value is very significant. This is similar to what observed in **[19]**. It should be noted that the example is for a typical European capital city and only the cases which are in excess of 300 handovers per observation period are cited.

More recently, with the proliferation of the use of frequency hopping and power control in the GSM community, it was found that quality based intra-cell handover can help to improve the robustness of the radio link significantly. Up to now, operators have normally used this more as a safety feature rather than as an active approach to optimize the frequency reuse. The dimensioning rules and the relationship with interference are still not fully understood.

As for code division multiple access (CDMA) systems such as cdmaOne, **[4]**, the dimensioning rules are, in principle, similar in many respects to other cellular systems. For instance, handovers from base station to base station will be largely based on the power level measurements from adjacent base stations. However, a CDMA system has the flexibility of soft quality degradation which allows more room for facing traffic fluctuations. As CDMA is inherently an interference-limited system, soft handover (by combining or selection) is required to control interference. To this end, additional radio hardware is required at the base station. This amounts to a 40% to 70% increase in channels and this is included as part of the hardware dimensioning rules. Thus, there is a subtle tradeoff between the system loading, mobility and capacity.

In order to maintain QOS[8] objectives, CDMA admission control can be either enforced by physically

---

[8] *Quality of Service (QOS): "The collective effect of service performances which determine the degree of satisfaction of a user of the service. The quality of service is characterized by*

limiting the number of user codes which can be used, or it can achieved by allowing the interference level or noise rise in the system to determine the capacity naturally. However, as most of the CDMA systems are still in their early phase of deployment, the effectiveness and the adequacy of these techniques need to be better understood. At present, where capacity is expected to be a problem, usually more carriers are added. Thus, the relationship between admission control and QOS control is still very much an undetermined issue from a practical network operation point of view.

Adjusting the System Dimensioning

As a network evolves, the number of subscribers will grow in time while the average traffic intensity per subscriber will gradually decline over time. This is because at network start up, the initial customers are generally business customers who generate a high amount of “minutes of use”. However, as time progresses, more (non-business) subscribers are added and the minutes of use by them are much lower than the business users. This will serve to dilute the average traffic intensity per subscriber across the network. For instance, at network startup, the traffic per subscriber is typically around 18-20 mErlangs per subscriber and this declines to 12-13 mErlangs as the network matures. As for the net effect, although the subscriber growth may induce a decline in minutes of use, the growth in the volume of traffic demand is still quite significant. An operator has to closely monitor the traffic statistics of the busy hour traffic channel utilization for all the cells. This information is obtained from the mobile switch statistics on an ongoing basis. When the level of blocking reaches a pre-determined threshold, action has to be taken to increase the number of transceivers per cell. Of course this is subject to the constraint of the spectrum allocation and availability. Once the maximum number of transceivers for a specific reuse pattern is reached, the frequency reuse has to be tightened in order to enable an operator to increase the number of carriers per cell. This normally takes time to plan the frequency re-tune and the installation of new equipment at the appropriate base sites. The overall optimization cycle for network capacity is summarized in Fig. 5.

When spectrum and frequency reuse become the limiting factors, for the longer term capacity relief, it will be necessary to increase the number of base sites in order to cell split[9], or to introduce a hierarchical cell structure (HCS) with microcells overlaying the macrocell network. The traffic dimensioning rules for estimating the number of traffic channels required become more complex for HCS. For example, directed retry, [21], i.e. attempting to serve a call in the second best server cell when the best cell has no channels available, could improve the trunking efficiency of the microcells but not the macrocells[10].

A recent approach, [22], may enable efficient reuse of frequencies between microcell and macrocells, although the new technique is yet to be tested in the field. In addition, the effectiveness of speed sensitive handover algorithms, [23], [24], [25], etc., could also impact the traffic capturing ability of the microcells. At the time of writing, operators and infrastructure suppliers are still actively testing the viability of the handover algorithms and the appropriate channel allocation strategies.

As mentioned before, for shorter term capacity relief, it is possible to temporarily borrow the spare capacity in the neighboring cells. In practice, the temporal distribution of the traffic among the base sites in a dense urban area could be rather uneven and not all the cells neighboring to a congested cell are simultaneously congested. Evidently, cells have to be significantly overlapped in order to be able to share capacity with each other.

In general, there are many optimization issues surrounding a cellular radio system. Most of these have multiple variables and constraints. Precise mathematical formulations are often difficult if not intractable. For instance, the traffic loading for a specific cell is dependent on the traffic distribution, the mobility pattern, spectrum allocation, handover algorithm, switch parameters setting, and so on. However, automated planning tools are beginning to emerge to assist engineers to plan and optimize their network. An example is the proliferation of automatic frequency assignment tools which put theory, [26], [27], [28], into practice. Yet there is still a long way for tools to evolve to become fully effective for the many other optimization problems of more direct relevance to radio network dimensioning.

---

the combined aspects of service support performance, service operability performance, service integrity and other factors specific to each service.", ITU-T Recommendation M.60, [20].

[9] *This is different from tightening the frequency reuse as cell splitting is to increase the spatial reuse of the frequency set rather than tightening the frequency reuse pattern. By tightening the frequency reuse pattern, the number of carriers in the frequency group per cluster is reduced. For instance, a nominal frequency reuse pattern for GSM is a 12 carriers reuse. However, operators with a generous spectrum allocation may relax the reuse pattern to 15 carriers or more. Although, to increase the network capacity, a tighter reuse pattern of 9 carriers or lower may be required after the first transceivers.*

[10] *Assuming a split band arrangement for the microcell and the macrocell, the coverage of the microcells is only a subset of the macrocell. Splitting the spectrum allocation would imply that the number of carriers for the macrocell will be reduced. In areas where there is microcell implementation, it is possible to have better trunking efficiency as the mobiles in the microcell will be able to access both the microcell and the macrocell. However, in areas beyond the coverage of the microcell, the macrocell will have less carriers and can only offer a lower volume of traffic.*

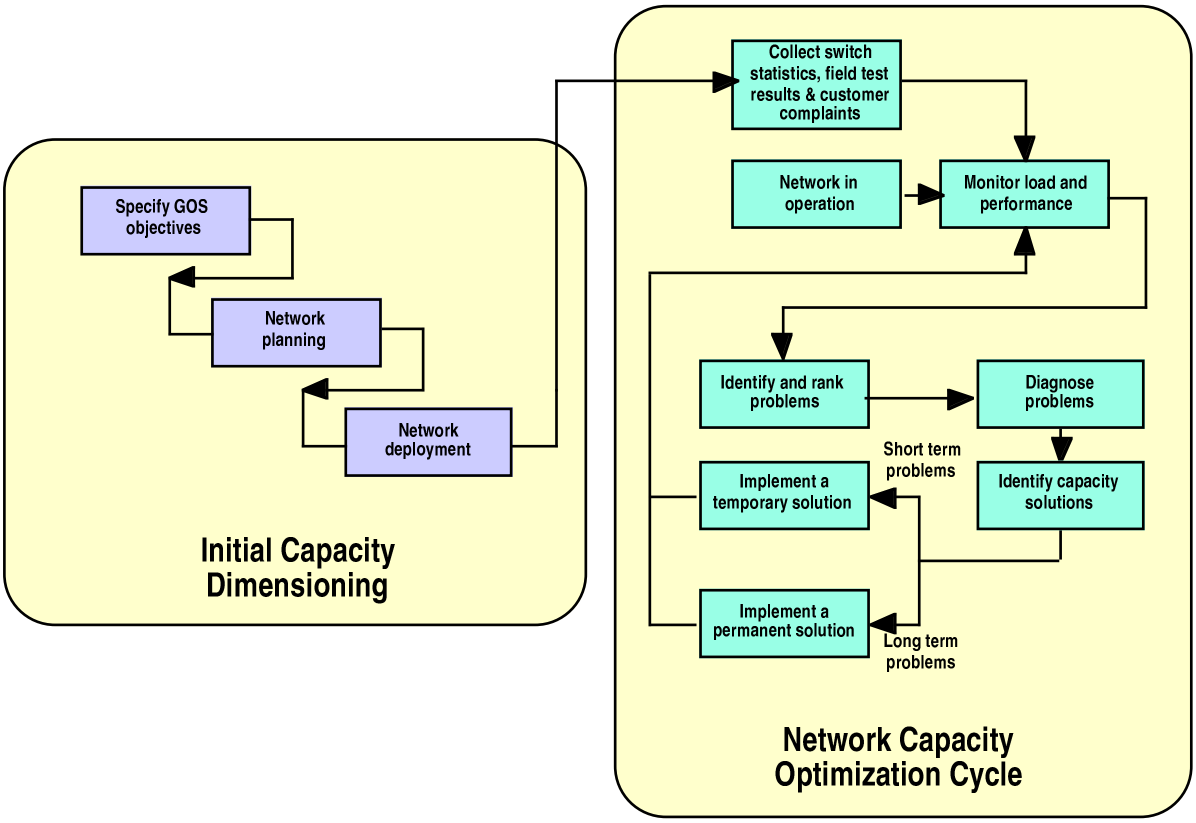


Figure 5. *The network capacity optimization cycle.*

ITU-T Framework for Traffic Engineering of Personal Communications

As already mentioned, traffic engineering for networks supporting mobile and UPT services is addressed in the ITU-T E.750-series of recommendations. This series covers traffic engineering for the user plane, while traffic engineering for the control plane is handled under separate ITU-T recommendations, typically those related to common channel signaling systems and IN (Intelligent Network)[11]. The characterization of personal communications user traffic demand is expected to deliver input to the demand processes for traffic engineering of the control plane.

The E.750-Series of Recommendations

The goal of the E.750-series is to recommend procedures for cost-effectively dimensioning network resources for terminal and personal mobility support. The scope of the E.750-series covers land, maritime and aeronautical services, as well as terrestrial and satellite based networks. To achieve the objectives of the E.750-series the following study areas need to be addressed:

- Mapping of user density and mobility for typical operating scenarios into user traffic demand (offered traffic);
- Definition of GOS parameters with user perception significance and related target values to set objectives for traffic engineering;
- Development of methods for meeting the GOS targets for specified demand patterns;
- Specification of measurement procedures for monitoring the GOS target attainment in the service operations environment.

Correspondingly, the series is organized into five major groupings (general aspects, traffic modeling, Grade of Service, dimensioning methods, and traffic measurements). Orthogonal to this grouping is the organization of the series into recommendations

---

[11] *A separation between the definition of the user plane and the control plane is that the former is associated with user data handling, whereas the latter relates to signaling traffic normally consisting of short messages and packet data.*

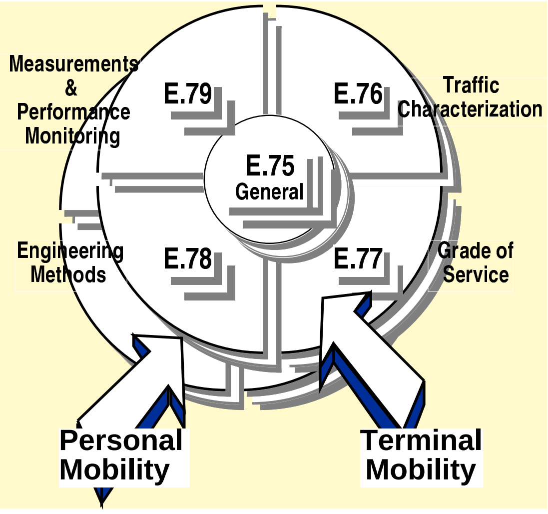


Figure 6. *Organization of the E.750-series of Recommendations.*

addressing terminal and personal mobility[12]. The organization of the E.750-series is summarized in Fig. 6.

The development of the series has been mainly focused until recently on reference connections and GOS aspects for both land and maritime/aeronautical services, with particular stress on terminal mobility issues. The E.750-series consists currently of twelve recommendations whose status ranges from approved/revised to draft recommendations to be finalized.

The Traffic Engineering Cycle

The approach taken in ITU-T is that for both fixed and mobile services the traffic demand should be characterized based on those aspects which are under no or little control of network operators and service providers. In the case of terminal mobility, these aspects relate essentially to the environmental propagation conditioning (e.g., indoor/outdoor, metropolitan/urban), mobile terminal speed, and user base characteristics (e.g., mobility behavior, fresh and repeated call arrival processes). As an example, Table 6 lists some parameters which characterize the operating scenarios envisaged for IMT-2000.

Application	Delivery Mode	Physical Attributes/ Propagation	Level of Mobility
Public cellular	Terrestrial	Indoor and/or outdoor	Stationary (0 km/h)
Private business	Satellite	Outdoor in urban, suburban, rural, hilly or coastal areas	Pedestrian (up to 10 km/h)
Residential cordless		Terrestrial or satellite operation	Typical vehicular (up to 100 km/h)
Fixed subscriber loop replacement		Land, maritime, or aeronautical operation	High speed vehicular (up to 500 km/h)
Residential neighborhood			Aeronautical (up to 1,500 km/h)
Mobile base station			Satellite (up to 27,000 km/h)
Paging			

Table 6. *Operating scenarios for IMT-2000,* **[29]**.

A key task in characterizing the traffic demand is how to capture the space-time relationship with due consideration given to the scope of ITU-T traffic engineering. Once the traffic demand is defined, traffic engineering for mobile systems should proceed by exercising the dimensioning methods with specific GOS/QOS objectives and cost constraints. Possibly, the dimensioning methods have to be repeatedly exercised to meet in the field GOS/QOS objectives and to cope with the necessary adjustments of the many operation parameters which, by necessity, cannot always be explicitly accommodated in the dimensioning procedures (e.g. power control, interleaving depth, frequency hopping patterns, source/channel coding, etc.). With a focus on radio resources for systems supporting terminal mobility, Fig. 7 schematically shows the envisaged traffic engineering study areas to be covered in the E.750-series. Note that these areas are closely related to the stages followed in the practice of radio planning and capacity dimensioning as depicted in Fig. 3. As a matter of fact, existing and envisaged recommendations in the E.76x (Traffic Characterization), E.78x (Engineering Methods), and E.79x (Measurements and Performance Monitoring) decades, relate, respectively, to the actions described under “figuring out the traffic demand”, “sizing the channel capacity of cells”, and “adjusting the system dimensioning” described in the section on the current practices for the operation of cellular systems. As indicated in the figure, the tasks associated with traffic engineering of cellular systems are part of a complex cycle which includes radio coverage design and frequency planning as key components.

---

[12] *Recommendations on terminal mobility apply to both existing, - e. g. GSM (Global System for Mobile communications, Europe), NADC (North American Digital Cellular, North America), and PDC (Personal Digital Cellular, Japan) - and developing mobile systems, e. g. IMT-2000 and UMTS (Universal Mobile Telecommunication System, ETSI).*

Scope of Traffic Engineering for Networks Supporting Terminal Mobility

The scope of traffic engineering for networks supporting terminal mobility can be classified according to two inter-related dimensions, i.e. the geographical/operation-domain dimension and the functional dimension. As for the former dimension, as a general rule, traffic engineering in ITU-T has been related to the international segments in the communication path. However, with the trend towards deregulation, competition and proliferation of roles in the provision of telecommunication services and the incurred dependency of service performance on interworking of an increasing number of subsystems in the communication path, the scope of traffic engineering has widened. Correspondingly, the span of traffic engineering for networks supporting terminal mobility ranges from metropolitan to international areas, as indicated in Fig. 8. As for the functional dimension, three major network segments can be identified (see also Fig. 1): i) the radio interface, ii) the fixed infrastructure of the mobile network; and, iii) the fixed core network. The radio interface comprises the transport and signaling functionality between the mobile terminal and the Base Transceiver Station (for simplicity in Fig. 1, BSS, Base Station System). The fixed infrastructure of the mobile network, spanned between the Base Transceiver Station and Mobile Switching Center (MSC), comprises the functionality for exercising/activating control on radio channel quality and availability as impacted by user population activity and mobility. Finally, the fixed core network, extending beyond the MSC, includes the functionality for user location, tracking, location updating, and call routing. For the case of separated fixed and mobile networks, [30], the figure also shows the allocation of mobility management functions (MMF) within the mobile network, as is typically the case with second generation mobile systems (e.g. GSM). Depending on the actual implementations and traffic requirements, the MSC (Mobile Switching Center) can be connected with the fixed network at the LE (Local Exchange) or TE (Tandem Exchange) level. This is succinctly indicated in Fig. 8 through the combination LE/TE. As a matter of fact, the allocation of MMF has a range of possible options including the arrangements resulting from integrated mobile and fixed network, [31]. The figure indicates two obvious teletraffic interfaces at which traffic demand has to be characterized for traffic engineering purposes. One traffic demand relates to the radio interface and has collected most of the contributions in the literature. The other is associated with the characterization of mobile related traffic which requires fixed network resources.

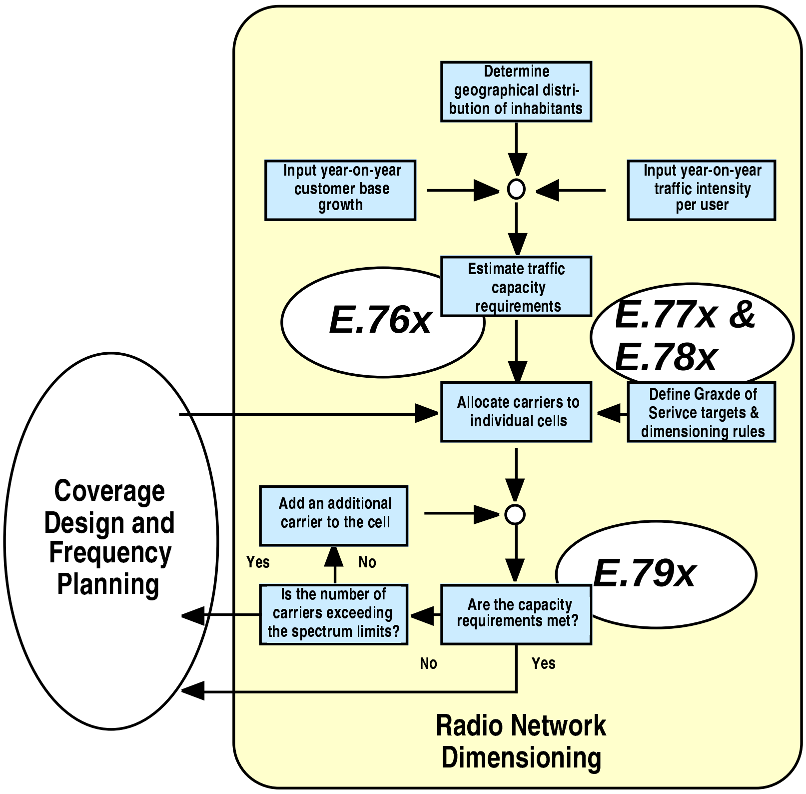


Figure 7. *Traffic engineering stages of E.750 for cellular systems (radio aspects and resources).*

The traffic engineering tasks for networks supporting terminal mobility relate to all three above functional segments. For the traffic engineering of the radio interface, the key GOS parameters are “probability of call blocking” and “probability of handover failure”, [32], [33]. The traffic engineering problems that must be addressed include (see also [34]):

- A tradeoff of radio spectrum reuse with call blocking and handover failure;
- Estimating the signaling load and allocating adequate bandwidth to handle mobility related signaling functions (e.g., paging and location updates);
- Radio resource allocation policies;
- Admission control strategies.

For the fixed part of the network, the key GOS parameters related to mobility are:

- Post selection delay (consisting of authentication delay, paging/alerting delay, time to obtain the routing number, and the fixed network ISDN/PSTN delays);
- Call blocking and lost signaling transactions;
- Profile lookup and update response times.

Fig. 8 shows the two parts of the fixed network: The fixed infrastructure of the mobile network (the middle band in Fig. 8 consisting of the BS, MSC, MMF and associated trunking and signaling) and the core network (the upper part of Fig. 8).

The mobility related traffic engineering problems that must be addressed for the fixed network are:

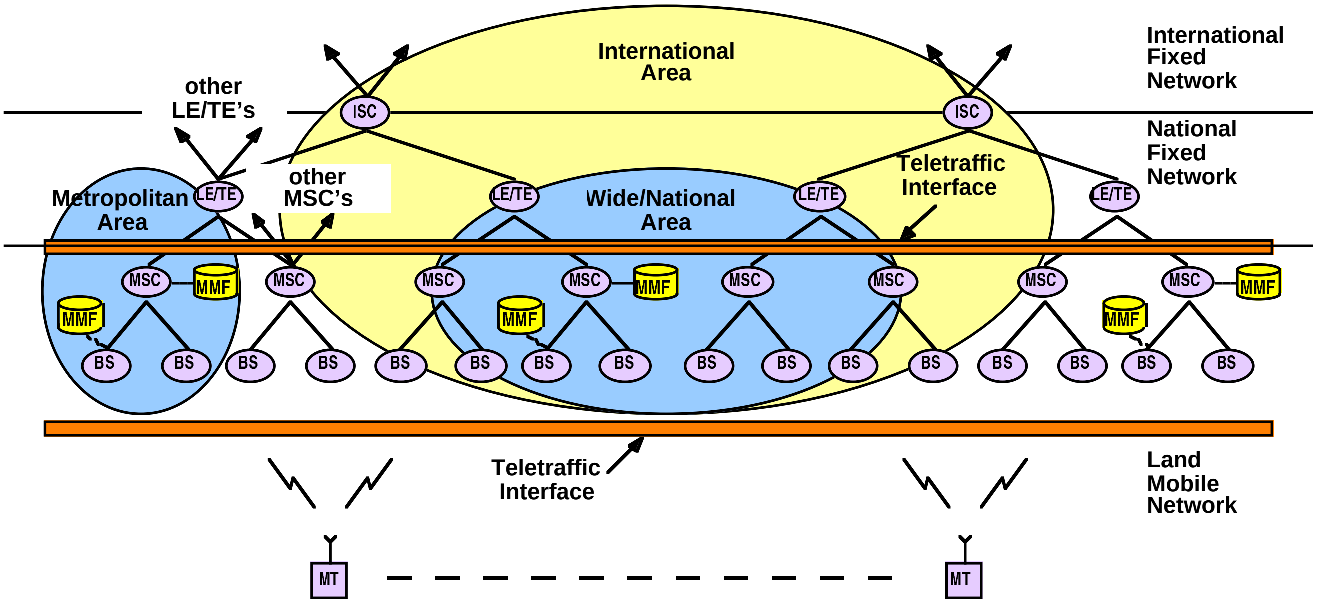
- Dimensioning the mobile infrastructure real-time resources in BS, MSC, and MMF systems to handle the mobility processing needs;
- Developing database architectures and dimensioning database real-time capacity in both the mobile infrastructure and core network to handle mobility functions;
- Dimensioning signaling and trunking capacity in both the mobile infrastructure and core network.

To characterize the traffic demand for these problems, models of user behavior must be developed. These models must characterize user density and mobility, user calling behavior (e.g., arrival, destination and holding time statistics), and user re-attempt behavior. These user behavior models are then coupled with system characteristics and operations to derive traffic load parameters such as rate of handovers, rate of location updates, channel occupancy time, etc. These traffic loads are then used to dimension the system and meet specified GOS targets.

Probing the Literature

Much research efforts have been spent to characterize user density and mobility, calling behavior and their performance impacts on wireless networks. Given the large volume of results in the literature, it is impossible to give a comprehensive review of them here. Rather, the purpose here is to present a brief overview of some of the models and results that, in our view, are representative. Readers can find other work on the subject referenced by the papers cited here. When appropriate, areas that require further study will also be pointed out.

The basic purpose of the mobility and teletraffic models is to capture the movement and calling behavior of subscribers as a means to predict or evaluate capacity and performance of wireless networks. Specifically, user movement in terms of direction and speed affects the time duration in which users stay in a cell or location area. In turn, a short time duration results in a frequent call handover when users are making calls, or a frequent update to location databases for call delivery even when users are not making calls. Highly mobile users may also require



KEY

MT	Mobile Terminal	BS	Base Station
MSC	Mobile Switching Center	LE	Local Exchange
TE	Tandem Exchange	ISC	International Switching Center
MMF	Mobility Management Function		

Figure 8. *Scope of traffic demand characterization for cellular networks (separated mobile and fixed network, mobile-to-mobile communication).*

more network resources for paging and other signaling functions for call delivery than is required for slower moving users. Furthermore, as discussed earlier, user density and service penetration play an important role in network planning and engineering. An area with a high user density is likely to yield high traffic demands, for which sufficient equipment has to be deployed to handle the projected traffic load. Last but not least, these models should capture the calling behavior in terms of call arrival rate, re-attempt, and call holding time distribution for proper system dimensioning. As traffic load of new and handover calls, location updates, call delivery and other signaling load have significant impacts on service quality (e.g., call blocking and dropping probability), there is always a strong demand for mobility and teletraffic models for analyzing and engineering cellular networks.

Radio Interface

Mobility and teletraffic models related to the radio interface typically address traffic load characterization and handover performance. By assuming uniform user density, and randomly chosen fixed movement direction and speed, the model proposed by [35] expresses the cell-crossing (i.e., handover) rate as a function of mobile density, mean velocity, and cell perimeter. The model is often referred to as the fluid-flow model. Although it neglects many aspects of practical situations, it appears to be the simplest mobility model with closed-form formulas. In fact, the assumption of uniform user density and movement direction or its variant are commonly used to study network performance; for example, see [36], [37], [38], and [39]. As pointed out below, [40] shows that the fluid model can closely approximate certain practical scenarios, but fails to do so in others.

In [38], the authors prove that the uniform assumption leads to a biased sampling condition. That is, the speeds of cell-crossing terminals are statistically different from those that remain within a cell. The probability distribution function for the speed of cell-crossing terminals are derived for use in performance analysis and simulation to obtain consistent comparisons among different design alternatives. Using the same assumption, [39] shows that (a) the handover rate (i.e., the mean number of handovers per call) increases as the square root of the increase in the cells per unit area; and (b) the handover rate is given by the ratio of mean call duration to the mean cell sojourn time. These results offer an understanding of design trade-offs and sizing issues in evolving wireless networks.

Although [41] also assumes uniform user density, the author proposes that the speed and direction of a mobile terminal are regenerated randomly and independently after an exponentially distributed random time (i.e., a random traveling distance). The mobility model is motivated by the need of reducing the complexity of simulation models for studying channel occupancy time, but [41] also presents arguments to justify the mobility assumption in certain settings. It is found that the channel occupancy time can be closely approximated by an exponential distribution in practical situations.

Recently, [42] develops a mathematical formulation for tracking movement of mobile terminals, which move randomly with degrees of freedom matching the mobility conditions in practical networks. The model can be used to characterize cell residence time, channel holding time, and the average number of handovers per call. It is found that cell residence time can be described by generalized gamma distributions, while channel holding time can be approximated by exponential distributions. The latter is consistent with the findings in [41], although the formulation in [42] provides additional modeling flexibility and capability.

Instead of tracking user movement, [43] studies various channel assignment strategies for handover calls by assuming that the residence times of a mobile terminal staying in different cells are independent and identically distributed. Such an assumption is made for tractability reasons. If it can be validated by actual field measurements, the mobility model will be useful in system design and engineering.

It is evident that the assumption of random user movement is not appropriate for cases such as users placing phone calls while driving. For this reason, researchers have observed the need of combining teletraffic theory and vehicular traffic theory [44] to estimate call and signaling traffic load. For example, [40] proposes a simulation method, which keeps track of terminal locations in cellular networks by using a vehicular traffic model based on realistic relationships among vehicle speed, density and volume. For uniform street layouts such as the Manhattan street patterns, it is found that the fluid-flow model [35] yields accurate cell/area boundary crossing rate when compared with the detailed simulation. However, the fluid-flow model can over or under-estimate the crossing rate in non-uniform street patterns. Furthermore, [45] proposes a simulation and analytic model to consider teletraffic load and vehicle movement. The authors also use the models to study the impacts on call blocking probability in case of sudden change of vehicular density in a ring-shaped service area.

Another mobility model used to capture time and space dynamics in cellular networks along a highway is presented in [46]. Assuming that each vehicle alternates between calling and non-calling state randomly, the model uses differential equations to describe the movement of both vehicle types. As a result, call traffic load for each cell along the highway and handover rates at cell boundaries can be obtained. With vehicles arriving to the system according to time-dependent Poisson processes, new call load and handover also form Poisson processes. The model can be viewed as a traffic demand and useful in studying the time and spatial dynamics in mobile networks. For instance, their analysis shows that a significant increase of offered load results if users initiate calls with a rate inversely proportional to their speed (e.g., in a traffic jam). In fact, this coupling effect between movement and calling behavior has not been fully understood. Additional study will be worthwhile on the subject as well as on re-attempt behavior, and two-dimensional traveling space. These aspects are addressed in other studies, although not in the framework of one comprehensive model. As examples of recent work: The impact of the user re-attempt behavior on protecting handovers against fresh calls is addressed in [47] (although disregarding the spatial dynamics); a two-dimensional space is addressed in [48] by reducing the problem to a combination of one-dimensional geometries (but assuming, among others, that the distribution of users on a segment is uniform); “spatial point patterns” modeling (two-dimensional) space, time-frozen distributions of mobile users while accommodating in a flexible and computation-friendly way a range of possible geometrical constraints are considered in [49] as a multi-dimensional extension of the Markovian Arrival Process (MAP), [50], based on arrival rates dependent on the “environment state” of the system.

It is worth noting the potential importance of the concept of different realms of traffic and mobility models with different levels of details for street, region, metropolitan and national areas, as reported in [51] and [52]. This is so because different amounts of details simplify the models, while adequately capturing the major essence of mobility and traffic situations in question. As a result, the simple models may be proven to be sufficient and useful in planning and engineering practical networks. Towards this goal, additional research work will be highly desirable. Recently, [52] introduces a set of mobility models with scope ranging from city to street level. Three models that cover city, zone and street, respectively, are intended to capture mobility at different scales as a means to estimate mobility and traffic parameters for engineering procedures. (As discussed later, [51] also makes a similar observation that mobility models with different levels of details are needed.)

Before leaving the discussion on the radio interface, it must be pointed out that mobility and traffic models for overlaid, micro/macrocell architecture [53], [54] are even more complicated than those for the single layer architecture. Due to the overlapping coverage of micro and macrocell, a call can be served either by micro or macrocell. This brings about new approaches of dynamic channel assignment according to the mobility of users; see e.g., [19], [20], [55], [56] and [21]. Further, efficient reuse of radio spectrum in the overlaid networks also becomes an issue [57], [18]. In terms of teletraffic issues, when a call is blocked by a microcell, it can be “overflowed” to the associated macrocell to see if the latter has spare radio channels. Existing methods such as [58], [59] and [60] may be useful for analyzing the call overflow, but additional factors such as mobility need also be considered to balance performance for new and handover calls [61]. A survey of some of these issues for micro/macrocell overlays can be found in [62].

It is clear that additional research will be needed to fully address the mobility and teletraffic issues for the overlaid networks.

Fixed Infrastructure

In the study of teletraffic issues for the signaling functions and the fixed infrastructure, terminal or personal mobility need not, in general, be considered as detailed as in most of the models discussed above. Rather, models with large scales will be adequate.

Adding new base stations and reducing cell size are a common approach to meet increased traffic demand. Small cell and registration areas however tend to increase signaling traffic. Using the fluid-flow model [35], [63] reveals a potentially significant increase of traffic load on the signaling links due to a combination of high terminal density and mobility, and small location area in PCS networks. Using the number of location updates between two calls, [64] proposes a framework for estimating the signaling load. Similarly, [65] and [66] predict a large increase of workload for the network databases to support mobility when compared with IN network services.

To avoid the performance impacts of signaling and database load due to mobility, many new mobility management or location tracking algorithms have been devised and analyzed. The common goal of these new methods is to reduce the network signaling (including paging over the radio channels [67], [68], [69]) and database load, thus improving call-setup delay, network capacity and perceived service quality, while efficiently delivering calls to mobile users. Since the details of the algorithms lie beyond the main scope of this paper, instead of discussing them here, readers are referred to the papers and their cited work on the subject that are published recently in two special issues of IEEE JSAC [70] and [71].

Last but not least, it is worth noting that [51] proposes a realistic teletraffic modeling framework, which consists of topology, call and mobility model. The call model is characterized by actual call data in an existing telephone network. The mobility model considers user movement at three different scales, resulting in metropolitan, national, and international submodel. The mobility parameters in the first submodel is estimated from personal transportation surveys, while those for the latter two are approximated from the airplane passenger traffic data. Using this framework, workload for the location database can be studied by simulation. Potentially, it can also be useful for evaluating various mobility management algorithms and network topology design.

	Radio Interface	**Fixed Infrastructure**
Traffic/mobility model	Fluid **[35]**; random change of direction and speed **[41]**, **[42]**; i.i.d. cell residence time **[43]**; vehicular traffic movement **[40]**, **[45]**, **[46]**, **[52]**	Fluid **[63]**, **[65]**, **[66]**, **[64]**; city/area model **[52]**; metro/national gravity model **[51]**
Call process	New and handover calls **[36]**, **[37]**, **[43]**; new and re-attempts **[47]**; call overflow in overlaid networks **[61]**	G/G/1 model **[72]** for signaling links and nodes
Typical traffic issues	Handover traffic rate **[41]**, **[39]**; handover prioritization **[36]**, **[73]**, **[37]**, **[43]**, **[21]**; paging **[67]**, **[68]**, **[69]**; dynamic channel assignment **[74]**, **[19]**, **[20]**, **[55]**, **[56]**; bandwidth reuse in overlaid networks **[18]**, **[57]**	Signaling link load **[63]**, **[64]**; location database load and processing capacity **[65]**, **[66]**, **[51]**; mobility management **[70]**, **[71]**

Table 7. *A classification and typical use of some traffic demand models.*

A summary of these mobility and traffic models, and related teletraffic issues for mobile networks is given in Table 7. The purpose there is to show the potential of the models in studying network performance and design issues.

Where Do We Stand?

Despite a large volume of (at times) quite sophisticated mobility, traffic demand and dimensioning models, theorists, for the sake of tractability, often make simplifying assumptions about user density, and assume certain statistical properties of channel holding time, cell residence time and other mobility related parameters, when modeling mobile communication networks. As the mobile and UPT services will undoubtedly provide users with rich features and multimedia capability, mobility and teletraffic issues for the future networks will become more complicated than those in the current second generation networks. It will be a challenge to the teletraffic community to provide engineering tools for different system generations meeting the robustness and simplicity requirements demanded for a smooth system operation.

Some Popular Assumptions: Traffic Engineering "Myths"?

In the area of terminal mobility and cellular systems, there has been increasing consensus in the open literature on a series of working assumptions which have led to mathematically tractable problems. Given the complexity of mobile system operation and the need for traffic engineering procedures with ITU-T significance, it is important to revive considerations of how well the models being used represent "real world" systems. This is by no means meant to undermine the value of traffic models, but rather to stress the need of validating with field data the indications from the models and determine that the models are accurate enough to justify their adoption in a sound teletraffic engineering practice. A point to note is that traffic statistics, and in general information on network operation, for a specific cellular radio network are highly sensitive proprietary information for an operator due to the competitive nature of the industry. This information is rarely published in the public domain nor shared outside the companies. (Infrastructure suppliers may occasionally have access to a limited amount of this information, but this could mostly be restricted to the start-up phase of a network). Since access to the wealth of information on real life network operation is effectively very restricted, contributions on traffic engineering to the open literature usually have to work from abstractions whose rationale and impact may not be backed in all cases by deep knowledge about real system operation and needs. These factors contribute to some of the disconnects between theory of traffic engineering and the real world.

Some of the most popular assumptions related to traffic and mobility modeling have included:

- **Radio cells have regular (hexagonal) shape.**
  Cells are defined in terms of radio coverage as provided by the power emitted by the base transceiver station/radio port antenna around which a cell is constituted. As such, the cell "boundary" is associated with the limiting distance from the antenna site beyond which communication with a mobile terminal becomes troublesome. Due to terrain characteristics and existence of obstacles of various nature interfering with propagation of radio waves, the boundary of a cell is normally fuzzy and a cell coverage may even be jeopardized over an area.
- **Handovers (in FDMA/TDMA systems) are accomplished as soon as a user crosses the boundary between adjacent cells).**
  Since cells are usually not regularly shaped, their radio coverage must overlap to some extent. This overlap provides a window during which the handover should be completed without substantially affecting the quality of the connection in a pre-determined time interval. How effective and rapid the handovers are depends on the efficiency and sophistication of the (proprietary) handover algorithms and the processing power.

  Moreover, although distance from the antenna site (i.e. the emitted power) plays an important role, (co-channel) interference is another major aspect impacting channel quality and hence the handover process. Although energy-saving and interference-reducing techniques based on activity periods of traffic sources (discontinuous transmission) are normally employed, co-channel interference may well override the effect of distance and become a major driver in the handover process. This is particularly true for CDMA systems where soft handover is a key system function.

- **The user flow rate out of a cell equals the flow rate into the cell contributed by users from neighboring cells.**
  Uniform distribution of users over a certain area for a specific time is a condition hard to be met in practice. Instead, user densities tend to exhibit accumulation points and considerable dynamics, especially in metropolitan areas and for a high penetration service such as the telephone service. In these conditions the accumulation points in the density move across cells, with incurred consequences on the balance of out- and in-flow rate of users even over relatively short periods.
- **The handover rate is an input parameter to the traffic engineering process similarly as the call arrival rate.**
  The call arrival rate is associated with the user propensity to request a service. The call arrival process characterizes the traffic demand and, unlike the mobility behavior, has some dependence on the communication infrastructure provided by the network operator and/or the service provider. Instead, the handover process is totally "transparent" to the user in that it is out of his or her control and is conditioned by the actual arrangements provided to ensure coverage and call carrying capacity while accommodating user mobility. For this reason, the E.750-series envisages that the characterization of the handover process is part of the engineering methods as opposed to be handled under terminal mobility traffic modeling proper.

To summarize, frequently used assumptions have included: Pure geometrical considerations for (radio) channel quality degradation; uniform and/or time-invariant distribution of the user base over the studied area; random or unidirectional motion of users, etc. Moreover, in the case of handover handling, the handover rate is often postulated as opposed to being derived based on such aspects as base station deployment, user mobility behavior, channel quality degradation due to uneven radio coverage and/or user activity, and system operation. The models based on the above assumptions have been typically used to estimate the transition rate of moving users in (cellular) wireless communication systems when crossing the "boundaries" between adjacent "regions", be these latter radio cells or location areas. The purpose of the models has been to devise strategies for maintaining target GOS and QOS objectives to established user traffic connections - e.g. handling handovers - and/or to estimate the signaling load involved in supporting both in-call and in-between-call mobility (messages for handover control, registration/authentication and location updating).

As already explained, given the safety margins with which cellular networks have been dimensioned, the assumptions have not diminished the value of (first order) indications drawn from the models. In particular, always more sophisticated computer aided tools are being developed to address the various stages of network planning, including estimation of the radio area coverage with a resolution depending on the granularity provided by the terrain data base. In actual situations, these tools may help in providing estimates of the transition rates - although normally at the cost of massive computation effort. As for the time-space dependence of traffic demand, the common operators' attitude has been to dimension for the "worst case" and basing on fixed channel allocation. This has meant that what has really mattered has been the identification of the envelope of the population maxima over the considered service area across the 24-hour period. With such an approach, studies on population density and related dynamics have been interesting for cellular operators to the extent that they have been able to provide indication on maximum resource allocation. Finally, the handover phenomenon is considered an extremely complex issue, impacted by a multitude of operation factors. As a rule, a substantial deviation from an "endemic" amount of handovers is taken as a flag for incorrect system operation (e.g. non-optimal handover hysteresis or wrong choice of antenna beam-width) leading to corrective actions based on switch statistics on a case-by-case basis. In this respect, studies providing an insight on the basic mechanisms underlying the handover process have been useful for relating key operation parameters.

In conclusion, models based on common, simplifying assumptions have had merits in helping to understand basic relationships between operation parameters. The extent to which these models have been used as an integral part of the traffic engineering practice is, however, quite limited. This is due to both the complexity of cellular networks planning and optimization and to the safety and plan-ahead margins in network dimensioning which have been affordable until recently for the network deployment. Increased competition in service offerings and higher expectations on service quality will drive for higher efficiency in operating network resources and for adequate, more comprehensive traffic engineering models.

The Challenges

Some of the challenges theorists have to face for supporting traffic engineering activity in a fast evolving design environment such as that relating to mobile systems are as follows:

- While second generation (digital) mobile systems are flanking and replacing first generation (analog) systems, third generation systems with even more advanced capabilities are being studied and standardized. The performance objectives of these new systems and the techniques on which they are based are different. The resulting GOS targets and the control and dimensioning methods have a dependence on the type of system being considered, and this widens and specializes the range of issues that need to be considered for traffic engineering purposes, thus making the traffic engineering task more arduous;

- One of the major thrusts in developing advanced, third generation mobile systems is to enable flexibility and efficiency in accommodating a variety of operation and traffic demand scenarios. This should be made possible by the adaptive capabilities underlying the system design. The wide-range that these adaptive capabilities will take (from adaptive source/channel coding to dynamic allocation of radio resources) has not only a profound impact on traffic engineering, but also makes the relationship between traffic engineering and network planning extremely intricate. Since the operation details of advanced mobile systems are still the object of ongoing standardization activity, relatively few firm results are available on which traffic engineering for advanced systems can be based;
- Traffic engineering studies for mobile systems have many objectives and use quite different modeling techniques. A common ingredient to these studies is the modeling of mobility which is very often tailored to the particular objective or tuned to the actual modeling technique chosen for the sake of problem tractability. This has not led to the development of mobility models general or flexible enough to be universally used for testing the effectiveness of control methods and/or the efficiency of dimensioning procedures.

Non-Voice Traffic

Although the problems addressed in this note are in principle related to the characterization of any kind of teletraffic demand, the main focus of the existing literature and ITU-T studies has been to concentrate on characterizing voice traffic demand in a PCS environment. However, with the increasing penetration of circuit switched and packet data and video services, the steadily growing popularity of interactive communication modes, and the prospects for advanced, post-IN architectures for supporting “agent based” execution of communication and processing tasks, the dimensions of traffic engineering for PCS (and personal communications) necessarily increase. As a matter of fact, not only does traffic engineering for fixed networks need to be revisited to account for the mentioned aspects, but a range of issues involving mobility, service accessibility and customization of user service environment have to be addressed to take care of the parallel, associated evolution of the PCS landscape.

The study of problems related to the above aspects are also in the scope of traffic engineering for PCS in ITU-T. In the current ITU-T work-program it is envisaged that a satisfactory description, for ITU purposes, of the traffic demand related to more traditional services, such as voice services, will be instructive and beneficial for paving the way to traffic demand description of other non-voice services.

Status and Prospects of ITU-T Work

The ITU-T activity on traffic engineering for Personal Communications is set in a framework where: i) mobile personal communications services continue to have exceptionally high penetration growth; ii) awareness among operators of the need for more efficient use of resources that were previously over dimensioned; iii) users have high expectations on both new mobile services and service quality; and, iv) a new generation of mobile systems with advanced features and self-organization capabilities is being standardized which will allow a truly seamless communication environment, shielding the user from many of the technicalities associated with using his or her terminal and accessing services across different operator domains.

Against this background, ITU-T is progressing the E.750 series trying to strike a balance between usefulness to operators, soundness of theoretical approach, and timeliness in covering advances in system design. To meet these objectives, the needs of ITU-T should first be expressed, and then a strategy for developing the E.750 series should be identified.

The Needs of ITU-T

Currently, the needs of ITU-T in terms of terminal mobility traffic demand modeling could be formulated as follows:

- Fundamental parameters associated with in-call and in-between-call user behavior as well as customer base mobility characteristics should be investigated, also considering the dependence on the variety of service types and offerings;
- The scope of traffic engineering encompasses metropolitan, national and international areas. This translates into the need for models with different level of detail;
- The teletraffic interfaces for which the traffic demand should be characterized divide into two broad categories: The radio interface(s) and the interfaces related to boundaries within the fixed network. While the embodiment of the mobility aspects are key for the radio interface category, the second type of characterization requires that the

nature of the traffic streams resulting from the composition of mobile related and fixed network related traffic demand be understood;

- Radio coverage aspects and user re-attempt behavior have a definite bearing on traffic demand and should become essential ingredients of the related characterization. Moreover, the interrelation between user density, speed of motion and radio coverage should be understood and expressed.

The needs of ITU-T in terms of traffic engineering and dimensioning methods can be characterized as follows:

- Models and methods need to be described for translating forecasts of traffic demand into capacity expansion of the available radio spectrum. For example, methods need to be described for cellular systems on how to modify the cellular layout/design and frequency re-use patterns to increase the overall channel capacity;
- Methods need to be provided to translate models of user behavior (e.g., user density, mobility, and calling behavior) into traffic demand (i.e., new and repeated call attempt rates) and then determine the GOS for a proposed use of radio spectrum. In addition, this traffic demand needs to be used to determine real-time processing load in the fixed infrastructure of the mobile network (e.g., in the MSC and MMF systems), and standardized methods are needed to dimension those resources;
- User density, call behavior and mobility models need to be used to standardize how to characterize and determine signaling loads in both the radio link and fixed network related to location updating, paging and location tracking. Standard methods need to be developed for dimensioning the needed resources in the radio link and the fixed network to handle these loads by evaluating and trading-off such aspects as: Paging and location update signaling costs; designing location areas to achieve a desired level of location update load; designing alternative database layouts (e.g., centralized vs. distributed HLR environments) in developing an overall system design.

Finally, the proposed methodology for standardization is to decouple the traffic demand characterization from the dimensioning and control procedures; this would then harmonize with the methodology used for traffic engineering of fixed networks and services. To accomplish this, the traffic demand processes have to be kept distinct from processes induced by network operation, such as channel quality maintenance and handover handling.

A Way Forward for Progressing the E.750 Series

Given the many dimensions of the traffic engineering tasks on one hand and the evolving technological scenario for mobile systems and services

	TERMINAL MOBILITY TRAFFIC MODELING (Recommendation E.760)	
Topic	**Comments**	**Priority**
Mapping population density into traffic demand associated with radio resources for metropolitan areas (user plane)	Account should be taken of re-attempt behavior; inclusion of short-term (daily) and longer-term (seasonal) population density shift over the investigated area; in-call mobility behavior highly desirable	High
Cell crossing rate for metropolitan areas	Geometrical considerations, different classes of users	High
Characterization of traffic demand at the interface between the radio access network and the fixed infrastructure of the mobile network (user plane)		Medium
Location area crossing rate	Geometrical consideration, different classes of users	Medium
	DIMENSIONING METHODS FOR LAND MOBILE SYSTEMS (Recommendation E.780)	
Topic	**Comments**	**Priority**
Dimensioning methods for cellular systems (phase 1)	One layer cellular lay-out, circuit switching (FDMA/TDMA), handover protection against fresh calls, repeated call attempts; no co-channel interference considerations	High
Dimensioning methods for cellular systems (phase 2a)	One layer cellular lay-out, circuit switching (FDMA/TDMA), handover protection against fresh calls, co-channel interference modeling	Medium
Dimensioning methods for cellular systems (phase 2b)	Two- (multi-) layer cell architecture, circuit switching (FDMA/TDMA), handover protection against fresh calls, repeated call attempts; no co-channel interference considerations	Medium
Dimensioning methods for cellular systems (phase 3)	Two- (multi-) layer cell architecture, circuit switching (FDMA/TDMA), handover protection against fresh calls, repeated call attempts; co-channel interference considerations	Low

Note - Inclusion of CDMA radio transmission technique and packet-switching in the scope of Draft Recommendation E.780 is pending, at the time of writing, due to the definition process on related matters going on in ITU-R TG 8/1 (IMT-2000).

Table 8. *Short-term work-plan and allocation of priorities for the progress of Draft Recommendation E.760 and E.780 on terminal mobility and land mobile, terrestrial-based systems (example).*

on the other hand, it is clear that a development path has to be identified for producing Recommendations in the E.750 series. Restricting the attention to terminal mobility aspects, this path should enable the coverage of all important aspects of traffic engineering in the complex relationship between radio coverage and frequency planning, and radio network dimensioning - while allowing for a progressive extension of the subjects covered so as to build on consolidated material.

In particular, from the preceding sections it seems that a sensible approach should be based on the following principles:

- Traffic engineering for mobile networks is part of a complex cycle whose aim is to optimize the dimensioning of the radio network. Traffic engineering should help in rationalizing some key phases in the cycle and contribute to the identification of operation regions which are both stable and cost-effective from a resource usage point of view. The actual operating region shall be identified in the field through fine-tuning of the operation parameters, with radio transmission and coverage constraints playing a decisive role.

  An effective way for traffic engineering to feed in the cycle would be to make available a series of "modules" for addressing mobility, traffic demand and dimensioning issues. These modules should be scaleable and accommodate both the geographical scope of the network and the fundamental differences in the architecture of mobile systems and the technique (Radio Transmission Technique) used in the radio interface.
- The scope of traffic engineering for networks supporting terminal mobility depends on a multitude of factors affecting, among others, network architecture, network planning principles and network operation. The combination of the possible options related to these factors identifies scenarios characterized by different frequency of actual system deployment and operation complexity - and consequently different degree of complexity from a traffic engineering point of view. It is suggested that the work to be progressed in the E.750 series relating to terminal mobility modeling and dimensioning methods for land mobile systems be organized in the framework of scenarios of increasing complexity, with less complex system being addressed first. This "phased approach" (exemplified in Table 8) would enable to address a large base of actual systems and capitalize on stable results.

Conclusions

At the cross-point between fundamental changes in the communications paradigms, coexistence between different generations of mobile systems, intensive competition between operators and need to optimize network resource usage, and user's aspiration to access, interact with and manipulate a personalized set of services in a genuine seamless communication environment, ITU-T traffic engineering for networks supporting terminal mobility has to develop an adequate path for producing related Recommendations in the E.750 series. This path should reconcile such issues as usefulness to operators, soundness of theoretical approach, and timeliness in covering advances in system design, among others. Clearly, this path will reflect both a priority choice in progressively advancing the coverage of the Recommendations and a strategy in building on consolidated material along the dimensions of geographical scope, functionality of the spanned network segments, architecture of the mobile and fixed network, supported services and transport modes, etc.

Key to the process of progressing the E.750 series is the understanding of the role of traffic engineering in the radio planning and capacity dimensioning cycle. This role identifies traffic engineering as an enabler for helping in locating regions for optimal resource usage and stable operation, with final setting of operation parameters decided in the field and dictated by radio transmission and coverage constraints. As a consequence of the complex relationship between traffic engineering and radio coverage planning for different mobile systems on one hand, and the ever increasing support of computer aided tools for addressing the stages involved in coverage and capacity planning on the other hand, a modular structure of the traffic engineering procedures seems appropriate. This structure should help in matching the traffic models to the characteristics of the network architecture and design of the radio interface.

Finally, the need for increased efficiency in network resource usage calls for mobility, traffic demand and dimensioning models which are comprehensive (e.g. reflect the self-organizing capabilities characterizing the design of advanced mobile systems) and incorporate a sufficient degree of realism - while remaining robust and manageable. Being able to respond to these requirements will continue to be a challenge for the teletraffic community.

Acknowledgments

The E.750-series is being developed in the Traffic Engineering working party of ITU-T SG 2. The series has also benefited from the cooperation with other groups in ITU (notably TG 8/1 of ITU-R) included in inter-sector activities on ITU IMT-2000. The opinions expressed in this article are the authors' view and do not necessarily reflect the consensus view of ITU-T.

The authors would like to acknowledge the constructive comments of the anonymous reviewers which improved the paper readability.

References

[1] P. E. Wirth, "The Role of Teletraffic Modeling in the New Communications Paradigms," IEEE Communications Magazine, vol. 35, no. 8, August 1997, pp. 86-92.

[2] J Karlson and B Eklundh, "A Cellular Mobile Telephone System with Load Sharing and Enhancement of Directed Retry," IEEE Trans Commun, vol COM-37, pp.530-535, May 1989.

[3] ITU-T Recommendation I.114, "Vocabulary of terms for Universal Personal Telecommunication," March 1993.

[4] Electronic Industries Association, 'Mobile Station - Base Station Compatibility Standard for dual mode wideband spread spectrum cellular system," Interim Standard IS-95.

[5] D. Grillo, "Personal Communications and Traffic Engineering in ITU-T -- The Developing E.750-Series of Recommendations," IEEE Personal Communications, vol. 3, no. 6, December 1996, pp. 16-28.

[6] ITU-T Recommendation E.600, "Terms and definitions for traffic engineering," March 1993.

[7] M. Callendar, "Future Public Land Mobile Telecommunication Systems," IEEE Personal Communications, vol. 1, no. 4, Fourth Quarter 1994, pp. 18-22.

[8] IEEE Personal Communications, special issue on "IMT-2000: Standards Efforts of the ITU," (M. H. Callendar Ed.), vol. 4, no. 4, August 1997.

[9] IEEE Personal Communications, special issue on "The European Path Towards Advanced Mobile Systems," (D. Grillo, S. T. S. Chia, and N. Rouelle Eds.), vol. 2, no. 1, February 1995.

[10] J.Schwarz da Silva, B. Arroyo-Fernandez, B. Barani, J. Pereira, and D. Ikonomou, "Mobile and Personal Communications: ACTS and Beyond," in "Wireless Communications: TDMA versus CDMA," Kluwer Academic Publishers, pp. 379-414.

[11] V. A. Bolotin, "Telephone Circuit Holding Time Distributions," Proc. of 14th International Teletraffic Congress (J. Labetoulle, and J. W. Roberts Eds.), Antibes Juan-les-Pins, June 6-10, 1994, pp. 125-134.

[12] A. Parviala, "Observed Traffic Variations and Their Influence in the Choice of Intensity Measurement Routine," Telektronikk, vol. 91, no. 2/3, 1995, pp. 69-78.

[13] V.A. Bolotin, "New Subscriber Traffic Variability Patterns for Network Traffic Engineering," Proc. of Int'l Teletraffic Congress (ITC'15), Washington D.C., June 1997, pp.867-866.

[14] E. Chlebus, "Empirical Validation of Call Holding Time Distribution in Cellular Communications Systems," Proc. of 15th International Teletraffic Congress (V. Ramaswami, and P. E. Wirth Eds.), Washington, 23-27 June, 1997, pp. 1179-1188.

[15] S. T. S. Chia, "Global System for Mobile Communications - Radio Coverage Design and Planning," BT Engineering, Vol 14, Oct 1995, pp. 216-224.

[16] S. T. S. Chia, "Providing Ubiquitous Cellular Coverage to a Dense Urban Environment," Proc. of IEEE VTC '94, Stockholm, June 1994.

[17] M. Mouly, and M.-B. Pautet, "The GSM System for Mobile Communications," published by the authors, Palaiseau, France, 1992.

[18] S. T. S. Chia, "Radio and System Design for a Dense Urban Personal Communication Network," IEE Electronics and Communication Engineering Journal, vol. 7 no., August 1995, pp. 177-184.

[19] V. Brass, and W. F. Fuhrmann, "Traffic Engineering Experience from Operating Cellular Networks," IEEE Communications Magazine, vol. 35, no. 8, August 1997, pp. 66-71.

[20] ITU-T Recommendation M.60, "Maintenance terminology and definitions," March 1993.

[21] B. Eklundh, "Channel Utilisation And Blocking Probability in a Cellular Mobile Telephone System with Directed Retry," IEEE Trans on Comm., vol COM-34, no.4, April 1986, pp.329-337.

[22] L. Wang, G.L. Stuber and C. Lea, "Architecture Design, Frequency Planning, and Performance Analysis for a Microcell/Macrocell Overlaying System," IEEE Trans. Veh. Tech., vol.46, no.4, Nov. 1997, pp.836-848.

[23] M. D. Austin, and G. L. Stuber, "Velocity Adaptive Handoff Algorithms for Microcellular Systems" IEEE Trans. Veh. Technol., vol . 43, no. 3, pp.549-561, 994.

[24] C.W. Sung and, and W.S. Wong, "User Speed Estimation and Dynamic Channel Allocation in Hierarchical Cellular System," Proc. of IEEE Veh. Tech. Conf., Stockholm, Sweden, June 1994, pp.91-95.

[25] B. Jabbari and W.F. Fuhrmann, "Teletraffic Modeling and Analysis of Flexible Hierarchical Cellular Networks with Speed-Sensitive Handoff Strategy," IEEE J. Sel. Areas on Commun., vol.15, no.8, Oct. 1997, pp.1539-1548.

[26] F. Box, "A Heuristic Technique for Assigning Frequencies to Mobile Radio Network," IEEE VT 27 (2), May 1978, pp.57-64.

[27] A. Gamst, "Some Lower Bounds for a Class of Frequency Assignment Problems," IEEE VT 35 (1), Feb 1986, pp.8-14.

[28] R. Mathar, and J. Mattfeldt, "Channel Assignment in Cellular Radio Networks," IEEE VT 42 (4), Nov 1993, pp.647-656.

[29] ITU-R Recommendation M.1034-1, "Requirements for the radio interface(s) for International Mobile Telecommunications-2000 (IMT-2000)".

[30] ITU-T Recommendation E.220, "Interconnection of Public Land Mobile Networks (PLMN)," February 1996.

[31] ITU-T Recommendation E.751, "Reference connections for traffic engineering of land mobile networks," February 1996.

[32] ITU-T Recommendation E.771, "Network Grade of Service parameters and target values for circuit-switched public land mobile services," October 1996.

[33] ITU-T Recommendation E.773, "Maritime and aeronautical mobile GOS concept," October 1996.

[34] D. Everitt, "Traffic Engineering of the Radio Interface for Cellular Mobile Networks," Proceedings of the IEEE, Special Issue on Wireless Networks, September 1994, pp. 1371-1382.

[35] R. Thomas, H. Gilbert and G. Mazziotto, "Influence of the Movement of the Mobile Station on the Performance of a Radio Mobile Cellular Network," Proc. of 3rd Nordic Seminar, Paper 9.4, Copenhagen, Sept. 1988.

[36] D. Hong and S.S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Handoff Procedures," IEEE Trans. Veh. Tech., vol.VT-35, Aug. 1986, pp.77-92.

[37] H. Xie and S. Kuek, "Priority Handoff Analysis," Proc. of IEEE 42nd Veh. Tech. Conf., NJ, May 1993, pp.855-858.

[38] H. Xie and D.J. Goodman, "Mobility Models and Biased Sampling Problem," Proc. of IEEE 2nd Int. Conf. Universal Personal Commun. (ICUPC'93), Ottawa, Canada, Oct. 1993, pp.804-807.

[39] S. Nanda, "Teletraffic Models for Urban and Suburban Microcells: Cell Sizes and Handoff Rates," IEEE Trans. on Veh. Tech., vol.42, no.4, Nov. 1994, pp.673-682.

[40] I. Seskar, S.V. Maric, J.M. Holtzman and J. Wasserman, "Rate of Location Area Updates in Cellular Systems," Proc. of IEEE 42nd Veh. Tech. Conf., May 1992, Coloardo, pp.694-697.

[41] R.A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System," IEEE Trans. on Veh. Tech., vol.35, no.3, Aug. 1987, pp.89-99.

[42] M.M. Zonoozi and P. Dassanayake, "User Mobility Modeling and Characterization of Mobility Patterns," IEEE J. on Sel. Areas in Communications, vol.15, no.7, Sept. 1997, pp.1239-1252.

[43] Y. Lin, S. Mohan and A. Noerpel, "Queueing Priority Channel Assignment Strategies for PCS Hand-Off and Initial Access," IEEE Trans. Veh. Tech., vol.43, Aug. 1994, pp.704-712.

[44] A.D. May, Traffic Flow Fundamentals, Prentice Hall, 1990.

[45] G. Montenegro, M. Sengoku, Y. Yamaguchi, and T. Abe, "Time- Dependent Analysis of Mobile Communication Traffic in a Ring-Shaped Service Area with Nonuniform Vehicle Distribution," IEEE Trans. on Veh. Tech., vol.41, no.3, Aug. 1992, pp.243-254.

[46] K.K. Leung, W.A. Massey and W. Whitt, "Traffic Models for Wireless Communication Networks," IEEE J. on Sel. Areas in Communications, vol.12, no.8, Oct. 1994, pp.1353-1364.

[47] P. Tran-Gia, and M. Mandjes, "Modeling of Customer Retrial Phenomenon in Cellular Mobile Networks," IEEE JSAC special issue on "Personal Communications - Services, Architecture and Performance Issues," (D. Grillo, A. Sasaki, R. A. Skoog, and R. Warfield Eds.), vol. 15, no. 8, Oct. 1997, pp. 1406-1414.

[48] K. Nakano, K. Saita, M. Sengoku, Y. Yamada, and S. Shinoda, "Mobile Communication Traffic Analysis on a

Road System Model," Proc. of Int'l Conference on the Performance and Management of Complex Communication Networks, Tsukuba, Nov. 17-21, 1997, pp. 1-16.

[49] G. Latouche, and V. Ramaswami, "Spatial Point Patterns of Phase Type," Proc. of 15th International Teletraffic Congress (V. Ramaswami, and P. E. Wirth Eds.), Washington, 23-27 June, 1997, pp. 381-390.

[50] D. M. Lucantoni, K.S. Meier-Hellstern, and M. F. Neuts, "A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes," Advances in Applied Probability, vol. 22, no. 1, 1990, pp. 676-705.

[51] D. Lam, D.C. Cox, and J. Widom, "Teletraffic Modeling for Personal Communications Services," IEEE Commun. Mag., Feb. 1997, pp.79-87.

[52] J.G. Markoulidakis, G.L. Lyeropoulus, D.F. Tsirkas and E.D. Sykas, "Mobility Modeling in Third-Generation Mobile Telecommunication Systems," IEEE Personal Commun., vol.4, no.4, Aug. 1997, pp.41-56.

[53] C.L. I, L.J. Greenstein and R. Gitlin, "A Microcell/Macrocell Cellular Architecture for Low- and High-Mobility Wireless Users," IEEE J. Select. Areas on Commun., vol.11, Aug. 1993, pp.885-891.

[54] S. Rappaport and L.R. Hu, "Microcellular Communication Systems with Hierarchical Macrocell Overlays: Traffic Performance Models and Analysis," Proc. of IEEE, vol.82, Sept. 1994, pp.1383-1397.

[55] K.L. Yeung and S. Nanda, "Channel Management in Microcell/Macrocell Cellular Radio Systems," IEEE Trans. Veh. Tech., vol.45, no.4, Nov. 1996, 601-611.

[56] K.K. Leung, S. Nanda and K.L. Yeung, "Operations of Cellular Communications Systems Based on Mobility and Teletraffic Statistics Collected by Mobile Units," U.S. Patent no. 5,623,535, April 22, 1997.

[57] A. Ganz, C.M. Krishna, D. Tang and Z.J. Haas, "On Optimal Design of Multilayer Wireless Cellular Systems," IEEE Commun. Mag., vol.35, no.2, Feb. 1997, pp.88-93.

[58] A. Kuczura, "The Interrupted Poisson Process as an Overflow Process," Bell Sys. Tech. J., vol.52, 1973, pp.437-438.

[59] H. Heffes, "A Class of Data Traffic Processes -- Covariance Function Characterization and Related Queuing Results," Bell Sys. Tech. J., vol.59, no.6, 1980, pp.897-929.

[60] K.S. Meier-Hellstern, "The Analysis of a Queue Arising in Overflow Models," IEEE Trans. on Commun., vol.37, no.4, April 1989, pp.367-372.

[61] P. Fitzpatrick, C.S. Lee and B. Warfield, "Teletraffic Performance of Mobile Radio Networks with Hierarchical Cells and Overflow," IEEE J. Sel. Areas on Commun., vol.15, no.8, Oct. 1997, pp.1548-1557.

[62] G.P. Pollini, "Trends in Handover Design," IEEE Communications Magazine, March 1996, pp.82-90.

[63] K.S. Meier-Hellstern, E. Alonso and D.R. O'Neil, "The Use of SS7 and GSM to Support High Density Personal Communications," Proc. of IEEE ICC'92, Chicago, IL, June 1992, pp.1689-1702.

[64] G.P. Pollini, K.S. Meier-Hellstern and D.J. Goodman, "Signalling Traffic Volume Generated by Mobile and Personal Communications," IEEE Commun. Mag., June 1995, pp.60-65.

[65] C.N. Lo, R.S. Wolff and R.C. Bernhardt, "An Estimate of Network Database Transaction Volume to Support Universal Personal Communication Services," Proc. of IEEE 1st Int'l Conf. on Universal Personal Commun. (ICUPC'92), Dallas, TX, Sept. 1992, pp.236-241.

[66] C.N. Lo and R.S. Wolff, "Estimated Network Database Transaction Volume to Support Wireless Personal Communications Services," Proc. of IEEE ICC'93, Geneva, Switzerland, May 1993, pp.1257-1263.

[67] S.T.S. Chia, "Location Registration and Paging in a Third Generation Mobile System," BT Tech. J., vol.9, no.4, Oct. 1991.

[68] S. Madhavapeddy, "Optimal Paging in Cellular Moible Telephone Systems," Proc. of Int'l Teletraffic Congress (ITC'14), 1994, pp.493-502.

[69] C. Rose, "Minimizing the Average Cost of Paging and Registration: A Timer-Based Method," ACM Wireless Networks, vol.2, no.2, June 1996, pp.109-116.

[70] IEEE Journal on Selected Areas in Communications, Special Issue on "Networking and Performance Issues of Personal Mobile Communications," H. Ahmadi, et al. (Ed.), Sept. 1997.

[71] IEEE Journal on Selected Areas in Communications, Special Issue on "Personal Communications -- Services, Architecture, and Performance Issues," D. Grillo, et al. (Ed.), Oct. 1997.

[72] L. Kleinrock, Queueing Systems, vol.I: Theory, Wiley, New York, 1975.

[73] S. Tekinay and B. Jabbari, "A Measurement-Based Prioritization Scheme for Handovers in Mobile Cellular Networks," IEEE J. Sel. Areas on Commun., vol.10, Oct. 1992, pp.1343-50.

[74] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," IEEE Personal Communications, pp.10-31, June 1996.

Biographies

Davide Grillo (M' 84) received his Doctorate in Statistics from the University of Rome, Italy, in 1965. In the same year he joined the Fondazione Ugo Bordoni, Rome, where he is currently Manager, Personal Communications. He was a Visiting Scientist at the Siemens Central Laboratory, Munich, Germany, the Department of Informatics at the University of Dortmund, Dortmund, Germany, and the IBM Research Laboratory, Zurich, Switzerland. He has been involved in several research areas, including: Telephone network operation and control; switching exchange architecture; packet switching networks; LAN/MAN architecture, interconnection and control; and resource allocation strategies in the radio subsystem of mobile networks. He has extensively published in those fields. He was a Guest Editor of an issue on LAN interconnection of the IEEE Journal on Selected Areas in Communications, a special issue of PERFORMANCE EVALUATION dedicated to performance analysis of high speed telecommunication systems, two special issues of IEEE Personal Communications on the European research and standardization activities for advanced mobile systems, and an issue of IEEE Journal on Selected Areas in Communications on Personal Communications. He has participated in the activity of the International Teletraffic Congress, and he was the promoter and Technical Chairman of the ITC Seminar on Personal Communications in 1992.

He has been an active member of Study Group 2 of ITU-T for questions on traffic engineering of telephone networks and ISDN, where from 1989 to 1996 he lead a group concerned with traffic engineering of networks supporting mobile and UPT services. In this capacity, he was the initiator of a dedicated series of Recommendations, the E.750 series. In ITU-T Study Group 2, he is currently Rapporteur for Performance Objectives, coordinator for the development of the E.750 series and Chairman of an ITU-T Focus Group dedicated to the progress of the series. He is also actively involved in TG 8/1 (FPLMTS/IMT-2000) of ITU-R, where he acts as Liaison Rapporteur to Study Group 2 of ITU-T. Dr. Grillo has been involved in the RACE projects on mobile communications (RACE I 1043, and RACE II MoNet and ATDMA) with responsibility in the co-ordination of traffic performance modeling activities. Dr. Grillo is a Technical Editor of IEEE Personal Communications.

Ronald A. Skoog (M' 89) received the B.S. degree in Electrical Engineering from Oregon State University in 1964, and the M.S. and Ph.D. degrees in Electrical Engineering from M.I.T. in 1965 and 1969, respectively.

From 1969 to 1971 he taught in the Department of Electrical Engineering and Computer Science at the University of California, Berkeley. He joined AT&T Bell Laboratories in 1971, and worked in the area of transmission facility network planning and optimization. In 1981 he began working in the areas of common channel signaling and switched network performance where he has worked on network architecture studies, signaling network protocols system, performance analysis and simulation studies, network survivability and reliability studies, and network traffic engineering and dimensioning methodologies. He currently manages a group responsible for switched network performance and reliability analysis.

Dr. Skoog has served as an associate editor for the IEEE Transactions on Automatic Control, as a Guest Editor of a special issue of the Communications Magazine (July 1990) on the role of SS7 in the evolving telecommunications networks, as a Guest Editor of the April, 1994 issue of the IEEE Journal on Selected Areas of Communications covering the topic "Common Channel Signaling Networks: Performance, Engineering, Protocols, and Capacity Management," as a Guest Editor for the June 1995 IEEE Communications Magazine Feature Topic on "Mobility and Intelligent Networks" and as a Guest Editor of an issue of IEEE Journal on Selected Areas in Communications on Personal Communications (October 1997). He is an Associate Rapporteur for signaling issues in the ITU-Telecommunications Standardization Sector Study Group 2; he is a member of IEEE and Sigma Xi.

STANLEY CHIA (M'87, SM'96) graduated from the University of Warwick with a B.Sc. in Engineering Electronics in 1983 and gained a Ph.D. in Microcellular Radio Propagation from the University of Southampton in 1987. He joined BT Laboratories in 1988 and has worked extensively on both radio and network aspects of digital cellular radio systems. In 1989, he became the Cellular Coverage Core Task leader of the European RACE (R1043) Mobile Project, leading an international team on third-generation mobile system research. During 1992, he was seconded to head the Radio Engineering Department of SmarTone Mobile Communications Ltd., Hong Kong, providing the first GSM service in the Asia Pacific region in record time. In 1994, he was appointed Technical Director of AirTel, the BT Spanish GSM Consortium. He joined Cellnet, a national cellular operator in UK, in 1995 leading a multi-discipline team responsible for developing the transmission network evolution strategy. In 1996, Dr. Chia joined AirTouch Communications Inc. first as a Principal Engineer in the International Technology Development Group responsible for providing technical consultancy to AirTouch's joint venture operations worldwide. Subsequently in 1997, he took up an appointment in AirTouch's Strategic Technology Department as Technology Director leading the Radio Science and Technology Group. Dr. Chia has published numerous technical memoranda and professional papers. He was awarded three times the IEE Younger Member Paper Premium in 1987, 1988 and 1990 and holds two patents on mobile radio technologies. He is a Member of the IEE, a Senior Member of the IEEE and is a Charter Engineer. He was a member of the IEE Professional Group E8 and is a member of the editorial board of the IEEE Personal Communications Magazine. He was a Guest Editor of a special issue of IEEE Personal Communications on the European research and standardization activities for advanced mobile systems.

KIN K. LEUNG (S'78-M'86-SM'93) received his B.S. degree with first class honors in electronics from the Chinese University of Hong Kong, Hong Kong, in 1980, and his M.S. and Ph.D. degrees in computer science from University of California, Los Angeles, in 1982 and 1985, respectively.

He attended UCLA under an exchange program between the Chinese University of Hong Kong and University of California. In 1986, he joined AT&T Bell Laboratories in Holmdel, New Jersey. Currently, he is a Technology Consultant at Broadband Wireless Systems Research Department of AT&T Laboratories, working on radio resource allocation, power control, MAC protocols, and mobility management in broadband wireless networks. He has made contributions and published in the areas of wireless and computer communication networks, performance modeling, and distributed processing and databases. He also holds a number of issued and pending patents in these areas.

He received the Distinguished Member of Technical Staff Award from AT&T Bell Laboratories in 1994 for his research work on performance analysis methodologies and their applications to enhance AT&T switching products and communication services. He served on the technical program committee for a number of conferences and as a Guest Editor of the October 1997 issue of the IEEE Journal on Selected Areas of Communications on "Networking and Performance Issues of Personal Mobile Communications." Currently, he is an editor for the IEEE Transactions on Communications in the area of wireless network access and performance.