Modelling mobility and capacity in wireless networks*

Richard J. Boucherie
University of Twente
r.j.boucherie@math.utwente.nl

1 Introduction

Mobile communications is a rapidly growing service in the field of telecommunications. However, the capacity (frequency spectrum) required to accommodate the increasing number of subscribers to mobile services is severely restricted due to government regulations and other users of the available frequency spectrum (e.g., radio, airline and military communications). Therefore, efficient use has to be made of that part of the frequency spectrum allocated to providers of mobile telecommunications services.

GSM, and its updates GSM – II, IS2000 and GSM – GPRS employed by European service providers, partitions the frequency band into separate frequencies (FDMA) that in turn are divided into time slots (TDMA). Each transmission via F/TDMA obtained channel can be used to establish a transmission between a base station and a mobile terminal. In contrast, third generation UMTS that will be enrolled within a few years uses orthogonal coding (CDMA) to separate transmissions that use the entire frequency band, which also results in a number channels for simultaneous transmission. To further increase capacity, in both standards the area covered by a service provider is divided into cells, each with its own base station. Channels can be re-used in different cells, taking into account interference constraints between transmissions in different cells. This cellular nature substantially increases the capacity of the network, but also imposes considerable problems for the network performance: a call moving from one cell to a neighbouring cell (handover) might find a lack of available channels in that cell, resulting in dropping of the call. Additional capacity can be obtained by restricting the cell size (base stations transmit at lower power) and adding extra (smaller) cells to the network.

There is an obvious trade-off between extra capacity (more smaller cells), and call loss due to handovers, as handovers occur more frequently in a network of smaller cells. Design questions for cellular networks involve such trade-offs. A balanced decision can be made on the basis of the evaluation of loss probabilities and related performance measures. This requires the combined analysis of telecommunications models for the analysis of channel availability, and of traffic models for studying the mobility of the subscribers. Fortunately, from a mathematical perspective, for the analysis of blocking probabilities, especially in the planning phase, only the effective number of available channels in the cells is of importance, so that the distinction between channels obtained via F/TDMA and CDMA need not be taken into account. Below, we present some ingredients required for the development of a stochastic model for analysing both the mobility of subscribers and the telecommunications network in one unifying framework. For a linear network, representing a road, the model is illustrated via the design of a self optimizing cellular mobile communications system that optimises its channel distribution depending on the load of the cells.

2 Telecommunications model: single cell

The Erlang loss model, developed by A.K. Erlang for a telephone switch, is a basic model for the analysis of call blocking probabilities. The model describes a single cell containing C channels in either an F/TDMA or a CDMA system. In this cell, subscribers randomly and independently generate calls of mean duration $\tau$, during which a call occupies a single channel. As the number of subscribers is large compared to the capacity of the cell, the arrival process of calls can conveniently be modelled as a Poisson process with rate $\lambda$, say. For this cell, the fraction, $E(\lambda \tau; C)$, of generated calls that is rejected as a consequence of lack of free channels is...
described by the Erlang loss formula
\[ E(\lambda \tau, C) = \frac{(\lambda \tau)^C}{C!} \left[ \sum_{k=0}^{C} \frac{(\lambda \tau)^k}{k!} \right]^{-1}. \]

The robustness of Erlang’s loss formula forms an important ingredient for the success of stochastic operations research methods in the analysis of telecommunications systems. In particular, the fraction of rejected calls depends only on the mean of the random cell duration. Additional characteristics of the distribution of the call length are not required.

\[ E(\lambda \tau, C) \]

\[ \lambda, \tau \rightarrow \ \text{C channels} \]

Figure 1: Erlang loss system.

It is important to observe that the Erlang loss formula for a cell with finite capacity \( C \) can also be obtained from the equilibrium distribution of the number of calls in a single cell with unrestricted capacity. For that cell, let \( N^\infty \) denote the random variable recording the number of calls in the cell, again assuming calls of mean duration \( \tau \) arriving according to a Poisson process with rate \( \lambda \), the equilibrium probability that \( n \) calls are present at time \( t \) is \( P(N^\infty(t) = n) = \frac{(\lambda \tau)^n}{n!} e^{-\lambda \tau} \). The Erlang loss formula is then obtained truncating \( N^\infty \) at level \( C \). To see this, let \( N^C \) denote the random variable recording the number of calls in the cell with capacity \( C \), then, for \( n = 0, \ldots, C \),
\[
\lim_{t \to \infty} P(N^\infty(t) = n) = \lim_{t \to \infty} P(N^\infty(t) = n|N^\infty(t) \leq C) = P(N^C = n) = \frac{(\lambda \tau)^n}{n!} e^{-\lambda \tau} \left[ \sum_{k=0}^{C} \frac{(\lambda \tau)^k}{k!} \right]^{-1},
\]
from which the Erlang loss formula immediately follows for \( n = C \).

The Poisson form of \( P(N^\infty(t) = n) \) remains valid when the cell with unrestricted capacity is not in equilibrium. Then, \( P(N^\infty(t) = n) = \frac{(\lambda \tau)^n}{n!} e^{-\rho(t)} \), with \( \rho(t) = E \int_0^t \lambda(u) du \), the load offered to the cell during a service time, where \( S \) is the random variable for the service requirement, and \( \lambda(u) \) the arrival rate at time \( u \). Based on the observation that the Erlang loss formula depends only on the load of the cell, the equivalence – in equilibrium – of the distribution of \( N^C \) and the truncation of the distribution of \( N^\infty \), and the observation that also the time-dependent distribution of \( N^\infty \) is Poisson, the truncation of that distribution to the set \( \{0, \ldots, C\} \) is frequently used to approximate the Erlang loss formula for cells with time-dependent load. This approximation is referred to as Modified Offered Load (MOL) approximation [5], and will be invoked below to approximate blocking probabilities for cellular networks with time-varying load.

### 3 Telecommunications network model

Erlang’s loss formula can conveniently be generalised for application to larger (and more realistic) networks containing multiple cells. Within the theory of telecommunications systems, loss networks have been analysed and applied with great success [4]. Such networks can be viewed as mobile networks consisting of several cells, say cells 1, \( \ldots \), \( D \). Cell \( i \) has \( C_i \) channels, offered load \( \lambda_i \tau_i \), where \( \tau_i \) is the mean call length, and \( \lambda_i \) is the arrival rate of fresh calls, \( i = 1, \ldots, D \). Ignoring mobility of the subscribers, an exact expression for the fresh call blocking probability can be obtained as a multi-dimensional extension of Erlang’s loss formula. Moreover, such extensions remain valid for systems with more complicated capacity restrictions involving multiple cells, i.e., restrictions due to sharing of channels over regions containing several cells, which is a common procedure that increases flexibility of the channel allocation, or to networks that also allow for different types of calls that might require multiple channels (e.g., data or video transmissions). The call loss probabilities are obtained from the equilibrium distribution of calls over the cells of the network. Let \( N = (N_i : i = 1, \ldots, D) \) be the random variable recording the number of calls present in cell \( i = 1, \ldots, D \). The joint equilibrium distribution, \( P(N = (n_1, \ldots, n_D)) \), for feasible \((n_1, \ldots, n_D)\), takes the form
\[
\tau(n_1, \ldots, n_D) = \lim_{t \to \infty} P(N(t) = (n_1, \ldots, n_D)) = \prod_{i=1}^D \frac{(\lambda_i \tau_i)^{n_i}}{n_i!} \left[ \sum_{(n_1, \ldots, n_D) \in F} \frac{(\lambda_i \tau_i)^{n_i}}{n_i!} \right]^{-1},
\]
where \( F \) is the feasible region containing all configurations of calls over the cells. The probability that a call is blocked in cell \( i \) is \( B_i = \sum_{(n_1, \ldots, n_D) \in F} P(n_1, \ldots, n_D) \| (n_1, \ldots, n_i + 1, \ldots, n_D) \notin F \).
the probability of states in which an extra call cannot be admitted in cell \( i \).

By analogy with (1), the distribution (2), and the blocking probability \( E_i \) are also obtained as a truncation of the distribution of the network with unrestricted capacity to the set \( F \). Furthermore, observe that the load \( \rho_i = \lambda_i \tau_i \) completely characterises the distribution, which also suggests that MOL with \( \rho_i(u) = \mathbb{E} \int_u^{\infty} \lambda_i(y) dy \) can be used to approximate the distribution of networks with time-varying load.

Mobility of subscribers can be taken into account when considering the fraction of subscribers moving from cell \( i \) to cell \( j \) while making a call. Let \( p_{ij} \) denote this fraction, \( i, j = 1, \ldots, D \), and let \( p_{ii} = 1 - \sum_{j=1}^{D} p_{ij} \) denote the fraction of subscribers that completes its call in cell \( i \), \( i = 1, \ldots, D \). The traffic equations determine the probability that a call is present in cell \( i \), say \( \lambda_i \), and balance for each cell the probability that a call departs from cell \( i \) (due to call termination, \( c_i p_{i0} \), or due to handover \( c_i p_{ij} \)) with the probability that a call arrives as fresh call, \( \lambda_i \), or as handover, \( c_i p_{ji} \). The traffic equations are also closely related to flow equations describing the average flow of calls through the cells. These equations read, for \( i = 1, \ldots, D \),

\[
\begin{align*}
\lambda_i &= \frac{c_i p_{i0} + \sum_{j=1}^{D} c_j p_{ij} \lambda_j}{\mu_j p_{ji} - \mu_j p_{j0}}.
\end{align*}
\]  

where \( \tau_i \) is the holding time in cell \( i \), and \( c_i, i = 1, \ldots, D \), solves (3). The equilibrium distribution of the network with capacity restrictions as expressed by a region \( F \) of feasible states can be approximated via a truncation of the distribution of \( N^\infty \) to \( F \) [3]:

\[
\pi(n_1, \ldots, n_D) \approx \lim_{t \to \infty} P(N^\infty(t) = (n_1, \ldots, n_D)|N^\infty(t) \in F) = \prod_{i=1}^{D} \left( \frac{\lambda_i \tau_i}{n_i!} \right)^{n_i} e^{-\lambda_i \tau_i}.
\]

the distribution (2) with \( c_i \) replacing \( \lambda_i \).

The distributions (2), (4), and (5), based on handover fractions \( p_{ij} \), do not take into account the exact location of the subscribers in the cells. For a F/TDMA-based networks this does not introduce a substantial error. For CDMA-based networks the feasible region \( F \) is based on the interference relation between subscribers in neighbouring cells, which is based on the exact location of subscribers in the cells. However, in the planning phase of the network, capacity predictions are provided based on the distribution of subscribers over the cells, i.e., based on the average location. Results such as the distribution (4) can be sharpened dividing the cells into small regions, which also leads to a multidimensional Poisson distribution [2]. It can be shown that for F/TDMA networks with fixed channel allocation, where cells do not share capacity, and for CDMA networks the approximation of \( \pi(n_1, \ldots, n_D) \) via truncation of the distribution of \( N^\infty \) to \( F \) yields an upper bound on the blocking probabilities \( E_i \) [1], which makes the truncated Poisson distribution (5) amenable for dimensioning of cellular networks.

For networks with time-varying rates, the MOL approximation requires the offered load for the network with unrestricted capacity. For a network with (possibly time-dependent) negative exponential holding times with rate \( \mu_i(t) \) in cell \( i \), the load \( \rho_i(t) \) offered to cell \( i \) can readily be obtained from the time-dependent version of the traffic equations (3):

\[
\frac{d\rho_i(t)}{dt} = \lambda_i(t) + \sum_{j=1}^{D} \mu_j(t)\rho_j(t)p_{ij} - \mu_i(t)\rho_i(t).
\]

The distribution of \( N^\infty(t) \) remains multidimensional Poisson

\[
P(N^\infty(t) = (n_1, \ldots, n_D)) = \prod_{i=1}^{D} \frac{\rho_i(t)^{n_i}}{n_i!} e^{-\rho_i(t)},
\]

and according to the MOL approximation, by analogy with the approximation for the single cell, and based
on the observation that (5) often yields a good approximation for the equilibrium blocking probabilities [1],
\[ P(N(t) = (n_1, \ldots, n_D)) \approx P(N^\infty(t) = (n_1, \ldots, n_D)|N^\infty(t) \in F) \]  
(8)
For cellular networks, often the load \( \rho_i(t) \) varies considerably, for example due to varying mobility (rush hours). An adequate prediction of the offered load is possible based on road traffic measurements and vehicular traffic theory. This is considered next.

4 Highway traffic model

Highway traffic models may be classified as microscopic or macroscopic models. In microscopic models the behaviour of individual drivers is explicitly described. Drivers are assumed to react to the differences in speed of the surrounding drivers by accelerating or decelerating. Due to the large number of drivers, analysis of the resulting model requires extensive computational effort. More attractive from a computational viewpoint are macroscopic models capturing the aggregate behaviour of traffic. Such models cannot describe detailed information such as speed of specific drivers, but often provide a good prediction of the average speed of traffic along the road.

![Traffic jam travelling along road](image)

Figure 3: Traffic jam travelling along road.

Consider a unidirectional road. The macroscopic model consists of two partial differential equations for the density, \( \rho \), and speed, \( v \), of cars. For example [6],
\[
\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial \rho(x,t)v(x,t)}{\partial x} = 0,
\]
\[
\frac{\partial v(x,t)}{\partial t} + v(x,t) \frac{\partial v(x,t)}{\partial x} = -\frac{1}{\tau}(v(x,t) - v^e(\rho)) + \frac{\tilde{v}^e}{2\rho T} \frac{\partial \rho(x,t)}{\partial x},
\]
where \( x \) denotes the location on the highway. These equations can be interpreted as fluid flow equations. The first equation represents conservation of vehicles, the second equation describes the evolution of speed in time. The right hand side of that equation indicates that speed is influenced by relaxation towards the equilibrium speed \( v^e \) (first term), and by anticipation to adjust the speed according to a change in density of cars along the road (second term).

Macroscopic traffic models are amenable to analysis, and serve as input for the stochastic model of the previous section. In particular, the traffic equations (9) describing the macroscopic behaviour of the number of cells in the cells can be replaced by macroscopic equations for the density of cars along highways. Assuming that subscribers generate calls independently, with \( \lambda \) the fresh call arrival rate per unit traffic mass, it can be shown that [7]
\[
\rho_i(t) = \lambda \tau \int_{x \in \text{cell} \ i} \rho(x,t) dx,
\]
where \( \tau \) is the mean call length, that might extend over multiple cells. Thus, the distribution of \( N^\infty(t) \) is given in (7), and an approximation of the blocking probabilities for the network with capacity restrictions is obtained from (8). Moreover, for F/TDMA networks with fixed channel allocation, and for CDMA networks, the fresh call blocking probability \( E_i(t) \) in cell \( i \) is approximated using the Erlang loss formula:
\[
E_i(t) = E(\rho_i(t), C_i),
\]
(9)
with \( C_i \) the capacity of cell \( i \).

For a given capacity allocation to the cells of the wireless network, based on the telecommunications and traffic models, call loss probabilities can be obtained. An alternative, and equally important, question is the optimal capacity allocation to the cells of the network, that is, we might aim for deriving \( C_i(t) \) such that \( E(\rho_i(t), C_i(t)) \leq 1\% \) for all \( t \). In the setting of a GSM network, this is considered below.

5 Self optimizing network

Consider a traffic jam moving along a road passing through several cells a GSM network. The speed and location of the traffic jam can be described using the model of section 4, and the blocking probabilities resulting from the load offered to the cells when the traffic jam passes a cell can be computed from (9). As the load offered by the users caught in the traffic jam may considerably exceed the average load offered to a cell, the capacity of the cell may be insufficient to accommodate that load. Therefore, it is natural to borrow frequencies from neighbouring cells that do not suffer
from overload so as to temporarily increase the capacity of the overloaded cell. Obviously, moving frequencies among cells requires the interference constraints to be satisfied. For simplicity, assume that initially all cells have equal capacity $C$.

![Capacity Requirement Diagram]

Figure 4: Capacity requirement at time $t$.

Consider the situation at time $t$ as depicted in Figure 4, which corresponds to the traffic jam at the location of Figure 3. The capacity required in cell 1 to achieve 15% blocking exceeds the initial capacity $C$. Capacity can be borrowed either from cell 0 or cell 2 as both cells have spare capacity since $C_0(t) < C$ and $C_2(t) < C$. Borrowing capacity must obviously take into account that the traffic jam is moving along the road, and that in the near future the peak of the traffic jam will reach cell 2 so that extra capacity is required in that cell. The choice for either borrowing from the left (from cell 0) or borrowing from the right (cell 2) can be made on the basis of the shape of the traffic density $\rho(x,t)$. In particular, for optimal network performance from the point of view of blocking probabilities, if the traffic density $\rho(x,t)$ is steeper to the left (as is the case in Figures 3 and 4), then borrowing should be from the left, whereas for a traffic density $\rho(x,t)$ that is steeper to the right, borrowing should be from the right [8]. This provides a simple rule of thumb that enables a mobile operator to decide upon a channel borrowing strategy on the basis of monitoring of road traffic information that is available on many of the Dutch highways.

6 Conclusion

Analysis and prediction of performance of cellular mobile networks requires the combined analysis of telecommunications and mobility models. This paper has described a mathematical framework for obtaining call blocking probabilities, and has demonstrated its use for on-line dimensioning of cellular networks.

References


