Call Admission Control in Cellular Networks

Schneps-Schneppe, Manfred ; Iversen, Villy Bæk

Published in:
Mobile Networks

Link to article, DOI:
10.5772/35866

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
1. Introduction

The service area of a cellular network is divided into cells. Users are connected to base stations in the cells via radio links. Channel frequencies are reused in cells that are sufficiently separated in distance so that mutual interference is below tolerable levels. When a new call is originated in a cell, one of the channels assigned to the base station of the cell is used for communication between the mobile user and the base station (if any channel is available for the call). If all the channels assigned to this base station are in use, the call attempt is assumed to be blocked and cleared from the system (blocked calls cleared). When a new call gets a channel, it keeps the channel until either the call is completed inside the cell or the mobile station (user) moves out of the cell. When the call is completed, the channel is released and becomes available to serve another call.

When a mobile station moves across the cell boundary and enters a new cell, a handover is required. Handover is also named handoff. If an idle channel is available in the destination cell, a channel is assigned to it and the call stays on; otherwise the call is dropped. Two commonly used performance measures for cellular networks are dropping probability of handover calls and blocking probability of new calls. The dropping probability of handover calls represents the probability that a handover call is dropped during handover. The blocking probability of new calls represents the probability that a new call is denied access to the network.

Call admission control (CAC) algorithms are used in order to keep control on dropping probability of handover calls and blocking probability of new calls. They determine whether a call should be accepted or rejected at the base station. Both the blocking probability of new calls and the dropping probability of handover calls are affected by the call admission algorithm used. The call admission algorithms must give priority to handover calls as compared to new calls.

Dedicated to the memory of Janis Sedols, Doctor of Mathematics (Dr.sc.comp.), 24.03.1939–11.08.2011. Dr. Sedols was active in writing this chapter.

Paper financed from EDRF’s project SATTEH (No. 2010/0189/2DP/2.1.1.2.0./10/APIA/VIAA/019) being implemented in Engineering Research Institute “Ventspils International Radio Astronomy Centre” of Ventspils University College (VIRAC).
Various priority based call admission algorithms have been reported in the literature, as for example (Beigy & Meybodi, 2003); (Ahmed, 2005); (Ghaderi & Boutaba, 2006). They can be classified into two basic categories: (1) reservation based, and (2) call thinning schemes.

1. Reservation based schemes:
   In these schemes, a subset of channels is reserved for exclusive use by handover calls. Whenever the number of calls (new calls) exceeds a certain threshold, these schemes reject new calls until the number of simultaneous calls (new calls) decreases below the threshold. These schemes accept handover calls as long as the cell has idle channels. When the number of calls is compared with the given threshold, this scheme is called call bounding (Ramjee et al., 1997); (Hong & Rappaport, 1986); (Oh & Tcha, 1992); (Haring et al., 2001); (Beigy & Meybodi, 2005). When the current number of new calls is compared with the given threshold, the scheme is called new call bounding scheme (Fang & Zhang, 2002). Equal Access sharing with reservation schemes reserve an integral number of channels or a fractional number (Ramjee et al., 1997) of channels for exclusive use by handover calls. Schemes with fractional number of guard channels have better control of the blocking probability of the new calls and the dropping probability of the handover calls than schemes with integral number of guard channels.

2. Call thinning schemes:
   These schemes accept new calls with a certain probability that depends on the number of ongoing calls in the cell (Ramjee et al., 1997); (Beigy & Meybodi, 2004). New call thinning schemes accept new calls with a probability that depends on the number of ongoing new calls in the cell (Fang & Zhang, 2002); (Cruz-Pérez et al., 2011). Both schemes accept handover calls whenever the cell has free channels.

In Sections 2 and 3, we compare four basic CAC strategies by examining new call and handover call blocking probabilities for the following schemes:

1. Dynamic reservation,
2. Fractional dynamic reservation,
3. Static (fixed) reservation,
4. New call bounding scheme.

Section 4 deals with Dynamic reservation and Static reservation in two-tier networks. To a great extent, our purpose is a tutorial one because there are many papers on CAC schemes, but they usually contain incomparable numerical results developed by computer simulations. Similar research as our is done by (Ramjee et al., 1997). They show that the guard channel scheme is optimal for minimizing a linear objective function of call blocking and dropping probabilities. The scheme studied below appeals also to network providers in terms of maximizing the revenue obtained by simple mathematical means.

1.1 Dimensioning of multi-tier networks

We have many types of multi-tier cellular networks all around us.

1.1.1 Mobile networks

A multi-tier cellular network is a network that has different types of cells overlaying each other. Each type of cell differs from others by the size. The smaller the size of the cells in a
certain area is, the more channels are available for users (since the number of channels per cell is fixed). We may consider four types of cells:

1. Pico-cells (range 10 – 50 m) are used inside buildings and lifts. The cell antennae are placed in corners of a room or in hallways. Pico-cells are used when the number of users in a building is large and signals from the outside cells cannot penetrate the building. A new type having almost the same features as pico-cells are called femto-cells.

2. Micro-cells (range 50 m – 1 km) are cells used mainly in cities where there are a lot of users.

3. Macro-cells (range 1 km – 20 km) are used in rural areas since the number of users is small, and in populated areas where micro-cells are too small to handle frequent handovers of users that are moving fast while making calls. For example, if you are in a high speed car and connected to a micro-cell, and the car moves too fast for the call to be handed over from one base station (cell antenna) to another, then the call will be dropped.

4. Satellites (world wide coverage).

Having multi-tier cellular networks increases the number of cells, which means that more users are able to use the network without being blocked, and that users in cars or any high speed vehicles are able to talk without worrying about their calls being disconnected.

1.1.2 Hybrid networks

The proliferation of computer laptops, personal digital assistants (PDA), and mobile phones, coupled with the nearly universal availability of wireless communication services is enabling the goal of ubiquitous wireless communications (Beigy & Meybodi, 2003); (Ramjee et al., 1997); (Leong & Zhuang, 2002); (Guerin, 1988). Unfortunately, to realize the benefits of omni-present connectivity, users must contend with the challenges of a confusing array of incompatible services, devices, and wireless technologies. Rice University, USA, is developing RENÉ (Rice Everywhere NEtwork) (Aazhang & Cavallaro, 2001), a system that enables ubiquitous and seamless communication services. The key innovations are a first-of-its-kind multi-tier network interface card, intelligent proxies that enable a new level of graceful adaptation in unmodified applications, and a novel approach to hierarchical and coarse-grained quality of service provisioning. The design of RENÉ requires a coordinated, collaborative effort across traditional layers and across different time scales of the system to maintain uninterrupted user connectivity.

1.1.3 Battlefield networks

Future battlefield networks will consist of various heterogeneous networking systems and tiers with disparate capabilities and characteristics, ranging from ground ad hoc mobile, sensor networks, and airborne-rich sky networks to satellite networks. It is an enormous challenge to create a suite of novel networking technologies that efficiently integrate these disparate systems (Ryu et al., 2003).

The key result of this chapter is the application part (Section 5) with the extension of the Equivalent Random Traffic method for estimation of throughput for networks with traffic splitting and correlated streams. The excellent accuracy (relative error less than 1%) is shown by numerical examples. The ERT-method has been developed for planning of alternate routing in telephone systems by many authors: (Wilkinson, 1956); (Bretschneider, 1973); (Fredericks, 1980) and others. In this chapter we propose an extension of the ERT-method
to take account of correlated streams. Sections 5.7 and 5.8 contain next step in ERT-method extension, namely the application of Neal’s theory (Neal, 1971), and his formulas for covariance of correlated streams.

2. Analysis of CAC strategies in single tier networks: Two channel case

We compare four basic CAC schemes by examining new call and handover call blocking probabilities:

- Strategy 1 - Dynamic reservation: the cutoff priority scheme is to reserve some channels for handover calls. Whenever a channel is released, it is returned to the common pool of channels.
- Strategy 2 - Fractional dynamic reservation: the fractional guard channel scheme (the new call thinning scheme) is to admit a new call with a certain probability which depends on the number of busy channels.
- Strategy 3 - Static (fixed) reservation: all channels allocated to a cell are divided into two groups: one to be used by all calls and the other for handover calls only (the rigid division-based CAC scheme).
- Strategy 4 - New call bounding scheme: limitation of the number of simultaneous new calls admitted to the network.

We consider a $N$-channel cell without waiting positions and two Poisson call flows: handover call flow of intensity $A$ and new call flow of intensity $B$. The holding times are exponentially distributed with mean value equal one. The exponential distribution simplifies the formulae and the fact that handover calls already have been served for some time before entering the cell considered, does not influence the remaining service time, as the exponential distribution is without memory. Our optimization criteria is the same for all schemes: to get the maximum revenue if each served $A$-call costs $K$ units ($K > 1$) and each served $B$-call costs one unit.

2.1 Dynamic reservation strategy is better than fractional dynamic reservation strategy

Let us start from the 2nd strategy: fractional dynamic reservation. This strategy seems to be more general than the dynamic reservation strategy, but we shall prove that such statement is not true. The system is modeled by a three-state Markov process (Fig. 1) having the following parameters: Number of channels $N = 2$, $A$ and $B$ = call flow intensities, $p_i$ = the probability of accepting $B$-calls for service in state $i$.

![Fractional dynamic reservation: two-channel state transition diagram.](image)

The stationary state probabilities $P_i$ are defined by equations (up to a normalization factor):

$$P_0 = 1$$

$$P_1 = A + Bp_0$$

$$P_2 = \frac{(A + Bp_0)(A + Bp_1)}{2}$$
The average revenue $H$ equals:

$$H = \frac{A(P_0 + P_1) \cdot K + B(p_0P_0 + p_1P_1) \cdot 1}{P_0 + P_1 + P_2}$$

$$= \frac{2(AK + Bp_0 + (A + Bp_0)(AK + Bp_1))}{2 + (A + Bp_0)(2 + A + Bp_1)}$$

(1)

This expression is the ratio of two polynomials, each one with probabilities $p_0$ and $p_1$ included in the first power. Consider the expression (1) as a function of one of the probabilities $p$. After multiplying by the relevant constant we can get it into the form $(p + a)/(bp + c)$. The derivative of this expression has the form $(ac)/(bp + c)^2$. Consequently, the expression (1) has invariant sign in the range of values $(0, 1)$, and its extreme values are located at the ends of the interval, i.e. probabilities $p_0$ and $p_1$ can only take values 0 or 1. Therefore, the dynamic reservation strategy has advantage over fractional dynamic reservation strategy.

2.2 Dynamic reservation strategy is better than static reservation strategy

This model (strategy 1) is a special case of the previous one: you can reserve 0, 1 or 2 channels, corresponding to choice of probability $(p_0, p_1)$ in the form of (1,1) (1,0) or (0,0), which, in own order, corresponds to the values of $R$ of 0, 1 or 2. Accordingly, the revenue from formula (1) takes the form:

$$H_0 = \frac{2(AK + B)(1 + A + B)}{2(1 + A + B) + (A + B)^2}$$

(2)

$$H_1 = \frac{2(B + AK(1 + A + B))}{2 + (A + B)(2 + A)}$$

(3)

$$H_2 = \frac{2AK(1 + A)}{2 + 2A + A^2}$$

How many channels should be reserved? This depends on the parameters $K$, $A$ and $B$. To find the optimal value of $R$, one should solve two equations pointing to the boundary of $K$:

$$H_0 = H_1 \quad \text{and} \quad H_1 = H_2.$$ 

We get:

$$K_1 = 1 + \frac{2 + A + B}{A(1 + A + B)}$$

(4)

$$K_2 = 1 + \frac{2(A + 1)}{A^2}$$

It is easy to verify that for any values of $A$ and $B$, the inequality $K_1 < K_2$ is true since:

$$K_2 - K_1 = \frac{2 + 2A + 2B + A^2 + AB}{A^2(1 + A + B)}.$$ 

Hence we have the following solution for optimal reservation $R$ at $N = 2$ channels:

$$R = 0 \quad \text{if} \quad K < K_1$$

$$R = 1 \quad \text{if} \quad K_1 < K < K_2$$

$$R = 2 \quad \text{if} \quad K_2 < K$$
2.3 Static reservation

In the cases \( R = 0 \) and \( R = 2 \), this strategy does not differ from the strategy \( 1 \) above. Therefore, it remains to consider the case \( R = 1 \). This Markov model has four states (Fig. 2):

- \((00)\) – both channels are free,
- \((01)\) – the common channel is engaged,
- \((10)\) – the guard channel is engaged,
- \((11)\) – both channels are engaged.

![State transition diagram](image)

Fig. 2. Two channels, static reservation: the model and the state transition diagram.

From linear balance equations under the assumption of statistical equilibrium we obtain the state probabilities (up to a normalizing factor):

\[
\begin{align*}
P_{00} &= 2 + 2A + B \\
P_{01} &= (A + B)^2 + 2B \\
P_{10} &= A(2 + A + B) \\
P_{11} &= A((A + B)^2 + A + 2B)
\end{align*}
\]

The average revenue takes the form:

\[
H_4 = \frac{A(2 + 4A + 3B + 2A^2 + 3AB + B^2) \cdot K + B(2 + 4A + B + A^2 + AB) \cdot 1}{2 + 4A + 3B + 3A^2 + 5AB + B^2 + A^3 + 2A^2B + AB^2}
\]

Our aim is to prove that the static reservation strategy cannot be more profitable than the dynamic reservation strategy. That is, we must prove that for any values of \( A, B \) and \( K \) at least one of the values of \( H_0 \) and \( H_1 \) are not smaller than \( H_4 \).

It is easy to verify that by replacing \( K \) in this formula by \( K_1 \) from expression (4) we obtain \( H_4 = 2 \). We get the same result by substituting this value of \( K_1 \) for \( K \) in expressions (2) and (3), i.e. the equalities \( H_0 = H_1 = H_4 = 2 \) are true for any \( A, B \) and given \( K = K_1 \). Furthermore, we note that \( H_0, H_1 \) and \( H_4 \) are linear functions of \( K \), i.e. straight lines. Note that for \( K < K_1 \) the inequalities \( H_0 < H_4 < H_1 \) are true. Hence these three straight lines intersect at one point, and the straight line \( H_4 \) is located between the two others. This means that for any values of \( A, B \) and \( K \), at least one of the values of \( H_0 \) and \( H_1 \) is not less than the value \( H_4 \), q.e.d.

2.4 New call bounding scheme (strategy 4)

In case of a 2-channel system, the only nontrivial variant of strategy 4 (Restriction on number of B-calls admitted) is: no more than one B-call. The state transition diagram of this model
looks similar to that in Fig. 2, and the expression for average revenue is:

\[
H_5 = \frac{2(A(A+B+1) \cdot K + B(A+1))}{2(A+B+1) + A(A+2B)}
\]

This strategy is similar to strategy 3 (Static reservation). As shown in Fig. 3, the straight lines \(H_4\) and \(H_5\) are close: up to point \(K < 1.25\) strategy 4 is a little more profitable, and from \(1.25 < K\) strategy 3 is more profitable. But for any \(K\), strategy 4 is worse than the optimal strategy 1, since \(H_5 < H_0\) up to \(K < 1.25\) and \(H_5 < H_2\) from \(1.25 < K\).

Fig. 3. Dependence of the revenue from handover call cost \(K\) for three models: \(D\) – dynamic reservation, \(F\) – static reservation, and \(L\) – restriction on number of admitted \(B\)-calls.

**Conclusion.** Hence it has been proved mathematically that the optimal service strategy in a two-channel system is dynamic reservation. Graphically, this fact is illustrated in Fig. 3:

- reservation \(R = 0\), optimal for values of \(K \leq 1.25\), reservation line \(D(R = 0)\),
- reservation \(R = 1\), optimal for values of \(K \geq 1.25\), reservation line \(D(R = 1)\).

**Problem 1.** It is noteworthy that all four straights in Fig. 2 intersect at one point. This experimental fact deserves further study for systems with more than two channels.

### 3. Comparison of four strategies in single tier network: Common case

#### 3.1 Fractional dynamic reservation

**Theorem.** For a \(N\)-channel loss system where \(B\)-calls are accepted with probability \(p_i\), depending on the number of busy channels \(i (i = 0, 1, \ldots, N)\), the optimal fractional dynamic reservation is limited to probabilities \(p_i\) equal to 0 or 1.
Fig. 4. Fractional dynamic reservation: the common case.

The stationary state probabilities $F_i$ are defined by equations (up to normalization factor):

$$F_0 = 1,$$
$$F_1 = A + B p_0,$$
$$F_2 = (A + B p_0)(A + B p_1)/2,$$
$$\ldots$$

$$F_N = (A + B p_0)(A + B p_1)\ldots(A + B p_{N-1})/N!$$

The lost revenue is equal to:

$$C = \frac{B \cdot (F_0 p_0 + F_1 p_1 + \ldots + F_{N-1} p_{N-1}) + (A K + B) \cdot F_N}{F_0 + F_1 + \ldots + F_N}$$

We should maximize the average revenue:

$$H = A \cdot K + B \cdot 1 - C.$$

Note that this expression is the general case of formula (1). As above, the expression (5) is the ratio of two polynomials, each of which includes the probability $p_i$ in first power. Consider the expression (5) as a function of the probability $p_i$ for any $i$. By the same argument as above we prove the Theorem.

*Problem 2.* The previous theorem does not imply which of the probabilities $p_i$ are equal to 0 or 1. Common sense is that $p_i = 1$ for $i = 1, \ldots, R$, and $p_i = 0$ for $i = R + 1, \ldots, N$. How to prove this mathematically?

### 3.2 Dynamic reservation as maximum revenue strategy

We compare two strategies: dynamic and static reservation. On the basis of numerical results we have shown that with the optimal reservation $R$ the expected revenue is always higher for the model with the dynamic reservation (Fig. 5). Naturally, when the handover call cost increases, then the number of reserved channels will increase. For $K = 2$ the optimal dynamic reservation is $R = 2$, and for $K = 4$ it is $R = 4$. The curves, of course. coincide at the ends of the definition interval when $R$ is equal to 0 or $N$.

Numerical calculations (Fig. 6) show that strategy 4, restriction of number of B-calls admitted, is similar to strategy 3. For large values of $R$ these strategies are almost identical, but even with the optimal value of $R$, strategy 4 has only a slight advantage over strategy 3.

**Conclusion.** The results of numerical analysis confirm that the optimal strategy is dynamic reservation. This statement is strictly proved in the case of a two-channel system.
Fig. 5. Dependence of revenue on the size of the reservation $R$ for the dynamic reservation $D$ and fixed reservation $F$. The cost of handover call is $K = 2$.

Fig. 6. Dependence of revenue on the size of the reservation for three strategies: $D =$ dynamic reservation, $F =$ static (fixed) reservation, and $L =$ restriction on the number of admitted $B$-calls.
3.3 On queueing effect

In queuing priority schemes, new calls and handover calls are all accepted whenever there are idle channels for that type of calls. When no idle channels are accessible, calls may be queued or blocked (i.e. cleared from the system). Queueing priority schemes can be divided into three groups: new call queuing schemes (Chang et al., 1999), handover call queuing schemes (Yoon & Kwan, 1993); (Tian & Ji, 2001); (Agrawal et al., 1996), and all calls queuing schemes (Yoon & Kwan, 1993); (Chang et al., 1999).

Computational analysis has shown that waiting positions do not change the advantage of dynamic reservation strategy. Fig. 7 displays two pairs of curves: one pair is the same as in Fig. 6, the second one relates to a case with three waiting positions. Of course, the revenue is growing, but the preference of dynamic reservation keeps the place.

[Diagram showing waiting positions and revenue curves]

Fig. 7. Waiting positions do not change the advantage of dynamic reservation strategy. One pair of curves is the same as in Figure 6, the other pair relates to the case with three waiting positions.

4. Two-tier network

4.1 Dynamic reservation versus static reservation

There are several reasons for designing multi-tier cellular networks. One is to provide services for mobile terminals with different mobility and traffic patterns. The required performance measures can be met if the traffic and mobility patterns can be classified into more homogeneous parts and treated separately. Consider a system where there are two mobility classes. If the cell radii are optimized for low-mobility terminals, then the high-mobility terminals will have to make a lot of handovers during a communication session. On the other hand, if the optimization is made regarding the handover performance of high-mobility terminals, then the traffic load in each cell may exceed acceptable limits.

In multi-tier cellular networks, different layers offer the designer the opportunity of class based optimization. In case of a two-tier network we consider three or seven micro-cells overlaid by one large macro-cell offered calls from two mobility classes (Fig. 8). High-mobility
Fig. 8. A macro-cell covering (a) three and (b) seven micro-cells.

calls of intensity \( A \) are served by macro-cell only. Low-mobility calls of intensity \( B \) are served by the micro-cells as first choice and, if the reservation strategy admits it, by the macro-cell as second choice. Arriving calls are served as follows. The mobility class of the call is identified. High-mobility calls of intensity \( A \) are served by macro-cell only. Low-mobility calls of intensity \( B \) are served by the appropriate micro-cell as first choice, and if reservation strategy allows it by macro-cell as second choice. In both cases, our optimization criteria is the same: to maximize the revenue when each served A-call costs \( K \) units and each served B-call costs one unit \( (K > 1) \). When calls reach the macro-cell level, they are no longer

Fig. 9. A two-tier cellular network call flow model.

differentiated according to their mobility classes. Therefore, the calls of the high mobility class terminals and the overflowed handover calls from micro-cells are treated identically. New calls from micro-cells may not use the guard (reserved) channels upon their arrival. If no non-guard channel is available, then new calls are blocked. High mobility calls are blocked if all macro-cell channels are busy. Fig. 9 shows schematically how calls are served and what order is followed when serving them. As above in the case of single-tier network we compare two reservation strategies:

a) Dynamic reservation: The cutoff priority scheme is to reserve a number of channel for high-mobility calls in the macro-cell. Whenever a channel is released, it is returned to the common pool of channels.
b) Static reservation: Divide all macro-cell channels allocated to a cell into two groups: one for the common use by all calls and the other for high-mobility calls only (the rigid division-based CAC scheme).

Fig. 10. Dependence of the revenue on the reserved number of channels for two-tier networks: (a) Three micro-cells and one macro-cell, (b) Seven micro-cells and one macro-cell.

Numerical results for two two-tier network examples are obtained. They are not qualitatively different from the results of the one-tier model discussed above. Fig. 10.a (three micro-cells and one macro-cell) and Fig. 10.b (seven micro-cells and one macro-cell) show that the dynamic reservation strategy gives the higher maximum revenue in both cases if the reserved number of channels \(R\) is properly chosen. The parameters are as follows:

- \(A\) = high-mobility call flow
- \(B\) = low-mobility call flow
- \(N_1\) = number of micro-cell channels for each cell
- \(N_2\) = number of micro-cell channels.

**Conclusion:** In case of two-tier network, the results of numerical analysis confirm that the optimal strategy is dynamic reservation.

### 4.2 Channel rearrangement effect

In hierarchical overlaying cellular networks, traffic overflow between the overlaying tiers is used to increase the utilization of the available capacity. The arrival process of overflow traffic has been verified to be correlated and bursty. This characteristic has brought great challenges to performance evaluation of hierarchical networks. In most published works, the discussion is focussed on traffic loss analysis in homogenous hierarchical networks, e.g. micro/macro cellular phone systems as in the numerical analysis below. In the paper (Huang et al., 2008), the authors address the problems of performance evaluation in more complicated scenarios by taking account of heterogeneity and user mobility in hierarchical networks. They present an approximate analytical loss model. The loss performance obtained by our approximated analytical model is validated by simulation in a heterogeneous multi-tier overlaying system.

Fig. 12 shows the dependence of revenue on channel rearrangement from macro-cell to micro-cell. We are looking for maximum revenue when low-mobility calls cost one unit and high-mobility calls cost \(K = 3\) units.

### 4.3 On optimal channel distribution (future study)

Fig. 13 shows two arrangements each of 18 radio channels for use by 4 call streams. Fig. 13.a shows a two-tier network with 3 individual channels per stream in the first tier and 6 common
channels in the second tier. In Fig. 13.b some kind of a homogeneous single tier network is depicted: each call has access to 9 channels equally distributed between streams. Such kind of arrangement could be implemented by modern DSP techniques.

Fig. 14 depicts the loss probability curves for these two schemes. Case (a) relates to pure loss system, case (b) relates to scheme with one waiting position per stream. What is surprising? In case (a), beginning with a loss probability as low as 0.56% (less than 1%), it is advantageous to use the equally distributed scheme. Therefore, the traditional two-tier network could be recommended here at a very low call rates. Table 1 contains more detailed data on loss probability. When a single waiting position is added, the advantage of the equally distributed scheme increases even more and the cross point of curves occurs at the loss probability equal to 0.025%. 

Fig. 11. An illustration of channel rearrangement.

Fig. 12. Dependence of revenue on channel rearrangement from macro-cell to micro-cell.
Fig. 13. (a) Grading type two-tier network, and (b) single-tier network with equally distributed channels.

![Diagram](image)

(a) Grading  
(b) Regular scheme

Fig. 14. (a) Comparison of channel arrangement between cells for loss system: grading type two-tier network (Fig. 14.a) is preferable for low rates only, but as load grows the scheme (Fig. 14.b) becomes preferable; (b) the same for the case of one waiting position per stream.

A historical remark regarding this strange phenomenon. It goes almost 50 years back, to the results of V.E. Beneš, a distinguished mathematician from Bell Laboratories, who worked on modeling so-called crossbar telephone switches. In (Beneš, 1966) he writes: "The question has arisen, whether there are examples of pairs of networks, with the same number of cross-points, the first of which is better than the second at one value of \( L \), while the second
is better than the first at another value of $L"$. The same type of problem was a goal of the studies (Schneps-Schneppe, 1963), but only for the earlier telephone exchange generation with step-by-step switches. In any case, the question remains related to preference of multi-tier networks in comparison with equally distributed schemes, taking modern digital signal processing (DSP) techniques into account.

5. Multi-tier network dimensioning by Equivalent Random Traffic Method

5.1 Introduction

Let us recall some ITU-T documents relating to the dimensioning of circuit groups in traditional telephone networks. These documents deal with dimensioning and service protection methods taking traffic routing methods into account. Recommendations E.520, E.521, E.522 and E.524 deal with the dimensioning of circuit groups with high-usage or final group arrangements.

Recommendation E.520 deals with methods for dimensioning of single-path circuit groups based on the use of Erlang’s formula (Fig. 15.a).

Recommendations E.521 and E.522 provide methods for the dimensioning of simple alternative routing arrangements as the one shown in Fig. 15.b, where only first and second-choice routes exist, and where all the traffic overflowing from a circuit group is offered to the same circuit group. Recommendation E.521 provides methods for dimensioning the final group satisfying GoS (Grade-of-Service) requirements for a given capacity of the high-usage circuit groups. Recommendation E.522 advises on how to dimension high-usage groups to minimize the cost of the whole arrangement. Fig. 15.b shows a two-tier network. For the dimensioning of such simple alternative routing arrangements the ERT-method is applicable.

Recommendation E.524 provides overflows approximations for non-random traffic inputs which allows for the dimensioning of more complex arrangements, e.g. three-tier network with correlated streams as shown in Fig. 15.c). The extended ERT-method described below relates to this case and could serve as a basis for a revision of Recommendation E.524.

Fig. 15. Examples of three network arrangements: single-tier, two-tier, and three-tier with correlated streams.
5.2 Erlang formula and its generalization for non-integer number of channels

We consider a system of $N$ identical fully accessible channels (servers, trunks, slots, call center agents, pool of wavelengths in the optical network, etc) offered Poisson traffic $L$ and operating as a loss system (blocked calls cleared). The probability that all $N$ channels are busy at a random point of time is equal to:

$$B = E(N, L) = \frac{L^N}{\sum_{i=0}^{N} \frac{L^i}{i!}}$$

(6)

This is the famous Erlang-B formula (1917) (Iversen, 2011). For numerical analysis of (6) we use the well-known recurrent formula:

$$E(N + 1, L) = \frac{L \cdot E(N, L)}{N + 1 + E(N, L)}$$

(7)

with initial value $E(0, L) = 1$. For the ERT-method we need Erlang-B formula for non-integral number of channels. How to get solution for a non-integral $N$? The traditional approach is based on the incomplete gamma function using:

$$E(N, L) = \frac{L^N \cdot e^{-L}}{\Gamma(N + 1, L)}$$

(8)

where

$$\Gamma(N + 1, L) = \int_{L}^{\infty} t^N e^{-t} dt$$

We propose a new approach for engineering applications. Let the value $N$ be from the interval $(0,1)$. We introduce a parabolic approximation for $\ln R = \ln E(N, L)$ at points $N = 0$, $N = 1$, and $N = 2$:

$$\ln E(0, L) = \ln(1) = 0,$$

$$B = \ln E(1, L) = \ln \frac{L}{L+1},$$

$$C = \ln E(2, L) = \ln \frac{L^2}{L^2 + 2L + 2}.$$  

Then we get the requested approximation:

$$\ln E(N, L) = \left( \frac{C}{2} - B \right) N^2 + \left( 2B - \frac{C}{2} \right) N$$

or in a more convenient form

$$E(N, L) = L^N (N+1)^{N^2-2N} (L^2 + 2L + 2)^{N^2/2}$$

(9)

Thus we have an initial value of $E(N, L)$ for $N$ inside the interval $(0,1)$ and we may calculate $E(N, L)$ at any $N$ by means of recurrent formula (7). In (Schneps-Schneppe & Sedols, 2010) the proposed approximation (6) is compared numerically with earlier known Erlang-B formula approximations (Rapp, 1964); (Szybicki, 1967); (Hedberg, 1981) and it is shown to be more accurate.
5.3 Kosten’s model and its new interpretation

The basic idea of the Equivalent Random Traffic (ERT) method is to use Erlang-B formula for overflow traffic offered to a secondary channel group of infinite capacity (Fig. 16), the so-called Kosten model (Kosten, 1937). Kosten’s paper contains formulae for all binomial moments of number of busy channels in secondary overflow group. In practice, only the two first moments are used for characterization of the overflow traffic: mean traffic intensity $M$ and variance $V$ (reference is usually made to Riordan’s paper (Riordan, 1956)):

$$M = L \cdot E(N, L)$$  \hspace{1cm} (10)

$$V = M \cdot \left(1 - M + \frac{L}{N + 1 - L + M}\right)$$  \hspace{1cm} (11)

From these two parameters one introduce a new parameter $Z$, the so-called peakedness:

$$Z = \frac{V}{M} = \left(1 - M + \frac{L}{N + 1 - L + M}\right) \geq 1$$ \hspace{1cm} (12)

Experience shows that peakedness $Z$ is a very good measure for the relative blocking probability a traffic stream with given mean value and variance is subject to.

We offer a new interpretation of Kosten’s results. We consider the scheme in Fig. 17. There are $N$ common channels and one separate channel. From (10 and (11 we get a new formula for the variance $V$ when both mean $M$ and mean $M_1 = L \cdot E(N + 1, L)$ are known. From recurrence formula (7) follows:

$$M_1 = \frac{LM}{N + 1 + M}$$

or

$$M + N + 1 = \frac{LM}{M_1}$$

After substitution of $(M + N + 1)$ into (11) we get:

$$V = M \cdot \left(1 - M + \frac{L}{M_1 - L}\right) = M \left(1 - M + \frac{M_1}{M - M_1}\right)$$ \hspace{1cm} (13)

We can reduce this expression to a simpler form:

$$V = M^2 \left(\frac{1}{M - M_1} - 1\right) = M \left(\frac{M}{M - M_1} - M\right)$$  \hspace{1cm} (14)

Note that $M - M_1$ is the load carried by a single channel and therefore it is always less than one. Formula (14) is useful for applications of the ERT-method in case of traffic splitting.
5.4 ERT-method

The ERT-method has been developed for planning of alternate routing in telephone networks by several authors: (Wilkinson, 1956); (Bretschneider, 1973); (Fredericks, 1980); and others. Fig. 18 explains the essence of the method. In (Fig. 18.a) \( g \) traffic streams which may for example be overflow traffic from other exchanges are offered to a transit exchange. As it is non-Poisson traffic, it cannot be described by classical traffic models. We do not know the distributions (state probabilities) of the traffic streams, but we are satisfied (most often the case in applications of statistics) by describing the \( i \)-th traffic stream by its mean value \( M_i \) and variance \( V_i \). The aggregated overflow process of the \( g \) traffic streams is said to be equivalent to

\[
L_1 \quad \text{N}_1 \quad \frac{M_1}{V_1} \quad L_2 \quad \text{N}_2 \quad \frac{M_2}{V_2} \quad \cdots \quad L_g \quad \text{N}_g \quad \frac{M_g}{V_g} \quad K
\]

\[
L \quad \text{N} \quad \frac{M}{V} \quad K
\]

\[
L \quad N + K \quad L \cdot E(L, N + K)
\]

Fig. 18. Application of the ERT-method: (a) \( g \) independent traffic streams offered to a common group of \( K \) channels, (b) equivalent group, (c) Erlang-B formula applied to a common group with \( N + K \) channels.

the traffic overflowing from a single full accessible group with the same mean and variance as the total overflow traffic. The total traffic offered to the group with \( K \) channels has the mean value:

\[
M = \sum_{i=1}^{g} M_i
\]

We assume that the traffic streams are independent (non-correlated), and thus the variance of the total traffic stream becomes:

\[
V = \sum_{i=1}^{g} V_i
\]

Therefore, the total traffic is described by \( M \) and \( V \). We now consider this traffic to be equivalent to a traffic flow which is lost from a full accessible group and has same mean value \( M \) and variance \( V \) (Fig. 18.b). For given values of \( M \) and \( V \), we therefore solve equations (10) and (11) with respect to \( N \) and \( L \). Then it is replaced by the equivalent system (Fig. 18.c) which is a full accessible system with \( N + K \) channels offered the traffic \( L \).

5.5 On accuracy of the ERT-method

Let us give a computational analysis of the classical ERT-method by a three-tier network shown in Fig. 19. There are four streams each offering a traffic equal to 5 erlang traffic. On first tier there are two servers per stream, on second tier there are three servers, and on third tier two servers. Application of the ERT-method to dimension the alternate routing networks consists of three steps.
Fig. 19. Three-tier network.

1. First step is a direct application of formulae (10) (11) for two streams and two individual lines (parameters: \( L = 5, N = 2 \)).

2. The second step, we apply the formulae (10) (11) to a three-channel group. We get the equivalent parameters \( L = 9.265, N = 5.83071 \) and the lost traffic \( M = 4.3224 \). The exact value given in brackets is obtained by solving the system of equations of the Markov process, and is equal to 4.30349, i.e. the relative error is less than 1%.

3. Third step: The two overflow streams are fed into the two lines. We get the equivalent parameters: \( L = 16.2076, N = 10.3188, \) and the lost traffic \( M = 6.97707 \). The exact value (in brackets) is 6.91011, i.e. the relative error again is less than 1%.

The results of calculations show the excellent accuracy of the method. However, such accuracy is not preserved when the number of channels in the third step increases. If instead of two channels we have \( (2 + g) \) channels in the common group, then Table 2 shows values of the loss for different \( g \) values. It is obvious the accuracy drops when increasing \( g \). For a value of \( g = 10 \), the relative error is bigger than 3%, but always on the safe side, and the absolute loss probability is very small.

<table>
<thead>
<tr>
<th>( g )</th>
<th>Loss probability</th>
<th>Loss probability</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact</td>
<td>ERT-method</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.4083</td>
<td>0.4105</td>
<td>0.539</td>
</tr>
<tr>
<td>2</td>
<td>0.3239</td>
<td>0.3277</td>
<td>1.173</td>
</tr>
<tr>
<td>4</td>
<td>0.2459</td>
<td>0.2506</td>
<td>1.911</td>
</tr>
<tr>
<td>6</td>
<td>0.1765</td>
<td>0.1812</td>
<td>2.663</td>
</tr>
<tr>
<td>8</td>
<td>0.1181</td>
<td>0.1218</td>
<td>3.133</td>
</tr>
<tr>
<td>10</td>
<td>0.0724</td>
<td>0.0747</td>
<td>3.177</td>
</tr>
</tbody>
</table>

Table 2. On accuracy of the classical ERT-method.

5.6 Fredericks & Hayward’s ERT-method

In (Fredericks, 1980) an equivalence method is proposed which is simpler to use than Wilkinson-Bretscheider’s method. The motivation for the method was first put forward by W.S. Hayward. For given values of \( (M, V) \) of a non-Poisson flow, Frederick & Hayward’s approach implies direct use of Erlang’s formula \( E(N, M) \), but with scaling of its parameters as \( E(N/Z, M/Z) \). The scaling parameter \( Z = V/M \) is the peakedness (12) (Fig. 20).
Fig. 20. An illustration of Fredericks & Hayward’s approach.

The accuracy of Fredericks & Hayward approach is numerically compared with ERT and with exact values. The calculations were performed for different variants of the scheme shown in Fig. 21. In general, its accuracy is comparable to that of the Wilkinson approach (Table 3). However, in our opinion Wilkinson’s approach is more reliable and always yields worst-case values.

Fig. 21. On accuracy of Fredericks & Hayward’s approach.

Table 3. On accuracy of Fredericks & Hayward’s approach.

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$N_1$</th>
<th>$L_2$</th>
<th>$N_2$</th>
<th>$K$</th>
<th>$M$ (Wilkinson)</th>
<th>$M$ (Hayward)</th>
<th>$M$ (exact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2.5</td>
<td>0</td>
<td>1</td>
<td>1.9740</td>
<td>1.9723</td>
<td>1.9721</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2.5</td>
<td>0</td>
<td>1</td>
<td>5.9635</td>
<td>5.9620</td>
<td>5.9589</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2.8717</td>
<td>2.8514</td>
<td>2.8498</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>0.2614</td>
<td>0.2268</td>
<td>0.2608</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>0.1122</td>
<td>0.0859</td>
<td>0.1140</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>0.1636</td>
<td>0.1386</td>
<td>0.1640</td>
</tr>
<tr>
<td>2.5</td>
<td>7</td>
<td>8.25</td>
<td>4</td>
<td>9</td>
<td>0.30268</td>
<td>0.2884</td>
<td>0.30273</td>
</tr>
</tbody>
</table>

5.7 Correlation of overflow streams

As it is, the ERT-method is not applicable to the analysis of multi-tier networks with correlated streams as shown in Fig. 15.c. In 1960’s, this type of problem appeared when dimensioning so-called gradings, the basic structural block in step-by-step exchanges. An important result was developed independently by (Descloux, 1962) and (Lotze, 1964). They determined mean $M$ and variance $V$ of the overflow stream components when split up after a first choice group as shown in Fig. 22.b. On the basis of Kosten’s model with two parameters $M$ and $V$, they developed a 5-parameter model for the two stream case: mean values $M_1$ and $M_2$, variances $V_1$ and $V_2$, and covariance Cov.

Fig. 22. (a) Kosten’s model, (b) two-stream model.
If \( A = p_1 L \) and \( B = p_2 L \), where \( p_1 + p_2 = 1 \), then the peakedness of partial stream \( V_i/M_i \) is defined by total peakedness \( V/M \):

\[
\frac{V_i}{M_i} - 1 = p_i \left( \frac{V}{M} - 1 \right),
\]

and covariance

\[
\text{Cov} = p_1 p_2 (V - M).
\]

The covariance formula is proved according to the theorem of variance for mutually dependent variables

\[
V = V_1 + V_2 + 2 \cdot \text{Cov}.
\]

### 5.8 Correlated streams: Neal’s formulae

During early 1970’s, Scotty Neal from Bell Labs studied the covariance of correlated streams in alternative routing networks (Neal, 1971). Below we use results from Neal’s paper to develop some formulae in notations of Fig. 23. We are looking for covariance between two overflow streams after groups with \( D \) and \( F \) channels, respectively. The key to Neal’s solutions is the original work (Kosten, 1937). Neal (Neal, 1971) extended the ERT-method to mutually dependent streams. On basis of the extended Kosten’ model (Fig. 23) he developed a technique for taking correlation into account when combining dependent streams of overflow traffic. More precisely, Neal has considered a 5-parameter Markov model (Fig. 23) with 5 parameters:

1. Number of busy channels in the first choice group (up to \( N \)),
2. Number of busy channels \( i \) in the first alternate group \( (0 \leq i \leq D) \),
3. Number of busy channels \( j \) in the second alternate group \( (0 \leq j \leq F) \),
4. Number of busy channels in the first imaginary infinite channel group,
5. Number of busy channels in the second imaginary infinite channel group.

![Diagram](image-url)

**Fig. 23.** An illustration to Neal’s formulas.

After rather sophistical derivations of two-dimensional binomial moment generating functions and using Kosten’s approach, Neal obtains linear equations for the two-dimensional probabilities \( \beta(i,j), (0 \leq i \leq D, 0 \leq j \leq F) \). The initial value is:

\[
\beta(0,0) = E(N,L) \tag{15}
\]

and other probabilities are defined by linear balance equations:

\[
(i+j)\nu_{i+j}\beta(i,j) = A \beta(i-1,j) + B \cdot \beta(i,j-1) - A \binom{D}{i-1} \beta(D,j) - B \binom{F}{j-1} \beta(i,F),
\]

where \( 0 \leq i \leq D, \ 0 \leq j \leq F, \ i + j > 0 \), \( \nu_{i+j} \) is

\[
\nu_{i+j} = \binom{D}{i} \binom{F}{j} - \binom{D}{i-1} \binom{F}{j} - \binom{D}{i} \binom{F}{j-1} + \binom{D}{i-1} \binom{F}{j-1}, \tag{16}
\]
with the following recurrent formulas for coefficients $v_i$:

$$v_i = \frac{L}{i \cdot v_{i-1}} + 1 + \frac{N - L}{i}, \quad i > 0, \quad \text{where} \quad v_0 = \frac{1}{E(N,L)}, \quad (17)$$

Then

$$M_1 = A \cdot \beta(D,0)$$
$$M_2 = B \cdot \beta(0,F)$$

$$Cov = \frac{A Q_1 + B Q_2}{2} - M_1 M_2 \quad (19)$$

where

$$Q_1 = \sum_{j=0}^{F} \left( \frac{\beta(D,j)}{1 + \sum_{j=1}^{F} \left( \frac{F}{j - 1} \prod_{k=j+1}^{F+1} \frac{B}{k v_k} \right)} \right)$$

$$Q_2 = \sum_{j=0}^{D} \left( \frac{\beta(j,F)}{1 + \sum_{j=1}^{D} \left( \frac{D}{j - 1} \prod_{k=j+1}^{D+1} \frac{A}{k v_k} \right)} \right)$$

Based on Neal’s formulae (15) – (19 we get the lost stream intensities $M_1$ and $M_2$ and the variance $V$. From (18) follows that loss probability of the first stream is $\beta(D,0)$.

### 5.9 Comments on Neal’s results

In the following discuss the applicability of Neal’s results.

#### 5.9.1 Algorithm

We extract the equations which have $j = 0$ from the equation system (16). Eliminating members with zero coefficients and considering that $\beta(0,0)$ is known, we obtain the system with $D$ equations referring to $\beta(i,0)$:

$$i v_i \beta(i,0) = A \beta(i-1,0) - A \left( \frac{D}{i-1} \right) \beta(D,0). \quad (20)$$

By solving this system of linear equations we get expressions for $\beta(i,0)$:

$$\beta(D,0) = \frac{1}{\sum_{i=0}^{D} \frac{D!}{(D-i)!} A^i \prod_{j=0}^{i} v_j}$$

$$\beta(i,0) = \beta(D,0) \left( \frac{D}{i} \right) \left( 1 + \sum_{j=1}^{D-i} \prod_{k=1}^{j} \frac{v_{i+k}(D+1-i-k)}{A} \right), \quad 0 < i < D \quad (22)$$

Values $v_j$ are obtained by formula (17). Using direct test we can ascertain that (21) and (22) together with statement (15) indeed satisfies the system of equations (20).
5.9.2 The modified Erlang formula

Formula (21) in the form

\[ \beta(D,0) = \frac{A^D}{D!} \sum_{i=0}^{D} \frac{A^{D-i}}{(D-i)!} \prod_{j=0}^{i} \nu_j \]  

(23)

is an obvious analogy to the Erlang formula for the scheme shown in Fig. 23. This formula is applicable to the ERT-method. It allows for non-integral number of channels.

From this the mean intensity \( M_1 \) follows:

\[ M_1 = \frac{AD^{+1}}{D!} \sum_{i=0}^{D} \frac{A^{D-i}}{(D-i)!} \prod_{j=0}^{i} \nu_j \]  

(24)

Formula (24) is easily implemented and allows for non integer values of \( N \).

5.9.3 Variance

By analogy of (13) we get the variance:

\[ V_1 = M_1 \left( 1 - M_1 + \frac{M_{1+}}{M_1 - M_{1+}} \right) \]  

(25)

where

\[ M_{1+} = A \beta(D+1,0) \]

Substituting \( A \) by \( B \) and \( D \) by \( F \) in (24) and (25) we get \( M_2 \) and \( V_2 \) in similar way.

5.10 Extended ERT-method. Numerical example

Consider an example where the extended ERT-method can be used and the covariance obtained by using formulas (18) - (19). Let us calculate the mean value \( M \) of the overflow traffic for the following scheme (Fig. 24).

Fig. 24. Scheme with correlated streams.

Using formulas (24) – (25) we find:

\[ M_1 = 0.180018, \quad M_2 = 0.873221, \quad V_1 = 0.252930, \quad V_2 = 1.25284 \]

Using (18) - (19) we calculate the covariance of the two streams:

\[ \text{Cov} = 0.0282596 \]

We now can calculate the intensity of the flow which is overflowing to the group with \( K \).
channels:
\[ M^* = M_1 + M_2 = 1.053239, \quad V^* = V_1 + V_2 + 2 \text{ Cov} = 1.56229. \]

Using the extended ERT-method we get the equivalent group:
\[ L^* = 3.33306, \quad N^* = 3.44900. \]

Therefore, using Erlang-B formula
\[ M = L^* \cdot E(N^* + K, L^*). \]

We can obtain mean intensity \( M \) of overflow stream for various values of \( K \) as shown in Table 4. The results of calculations show the excellent accuracy of the extended ERT-method.

<table>
<thead>
<tr>
<th>( K )</th>
<th>( M ) exact</th>
<th>( M ) by ERT</th>
<th>Rel. error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63322</td>
<td>0.63801</td>
<td>0.756</td>
</tr>
<tr>
<td>2</td>
<td>0.34583</td>
<td>0.34936</td>
<td>1.019</td>
</tr>
<tr>
<td>3</td>
<td>0.17067</td>
<td>0.17128</td>
<td>0.357</td>
</tr>
<tr>
<td>4</td>
<td>0.07616</td>
<td>0.07492</td>
<td>1.634</td>
</tr>
<tr>
<td>5</td>
<td>0.03091</td>
<td>0.02929</td>
<td>5.215</td>
</tr>
</tbody>
</table>

Table 4. Accuracy of the Extended ERT-method for correlated streams.

However, such accuracy is not preserved when the number of channels \( K \) in the final group increases. Table 4 shows values of the loss for different \( K \) values. It is obvious that for increasing \( K \) the accuracy drops. For the value \( K = 5 \) the relative error is greater than 5%. The same effect one observes in Table 1 and Table 2. For decreasing (very small) blocking probabilities the accuracy increases, but the absolute error decreases.

6. Conclusions

There are two parts in this chapter: a theoretical one and an application one. In the theoretical part, we consider four strategies for call admission control (CAC) in single and two-tier cellular networks, which are designed to ensure priority of handover calls, namely: dynamic reservation (cutoff priority scheme), fractional dynamic reservation (fractional guard channel scheme), static reservation (fixed division-based CAC scheme) and restriction on the number of simultaneous new calls admitted (new call bounding scheme). We show the advantage of dynamic reservation by numerical analysis and strictly prove it in the case of two-channel system with blocking.

The application part deals with Equivalent Random Traffic method for multi-tier networks dimensioning. The key result is an extension of the Equivalent Random Traffic method for estimation of the throughput for networks with traffic splitting and correlated streams. The excellent accuracy (relative error less than 1%) is illustrated by numerical examples.

7. References


