University of Würzburg
Institute of Computer Science
Research Report Series

Efficient call handling procedures in cellular mobile networks

Michel Mandjes[*†] **Kurt Tutschku**[‡]

Report No. 144 July 1996

[*]*Department of Econometrics,*
Vrije Universiteit Amsterdam,
De Boelelaan 1105,
1081 HV Amsterdam, The Netherlands
mmandjes@econ.vu.nl

[‡]*Institute of Computer Science,*
University of Würzburg
Am Hubland,
97074 Würzburg, Germany
tutschku@informatik.uni-wuerzburg.de

The performance of nowadays cellular mobile networks can be enhanced considerably by using more efficient call handling mechanisms. This paper discusses two advanced schemes, namely guard channels and the retry of blocked handovers. Furthermore, a more detailed customer model is presented in this study in order to obtain a more acccurate cell configuration: instead of the usual Poisson input, we allow the blocked fresh calls to be repeated. We show how this model can mathematically be analyzed and evaluate the efficiency of the call handling mechanisms.

Keywords: Mobile cellular networks, Performance evaluation, Repeated attempt, Guard channels, Handover retry.

[†]Work done while first author was visiting University of Würzburg during May, June, and July 1996. This material is based on work supported by the Dutch Science Foundation NWO. The current address of the first author is KPN Research, PO Box 421, 2260 AK Leidschendam, The Netherlands, E-mail: `m.r.h.mandjes@research.kpn.com`

1 Introduction

The dimensioning of nowadays public cellular communication networks is based rather on rough estimates than on accurate analytical models of the applied protocols and the expected user behavior. Especially the models for the performance analysis of call handling mechanisms are not very accurate and lead to a inefficient system configuration. It is expected that the use of efficient call handling mechanisms can enhance the efficiency of cellular mobile networks. Apart from this, the application of more detailed traffic models shows that the actual load is considerably higher than estimated by the usual traffic models. Thus, the real performance of common systems is lower and a larger efficiency requires better call handling mechanisms.

The characteristic of cellular mobile communication systems is the partitioning of its service area into radio cells, cf. Figure 1. A cell is defined as the supplying area of a transmitter, and its boundary is given by the attenuation of the radio signal due to the wave propagation laws. The *cellular concept* achieves a larger system

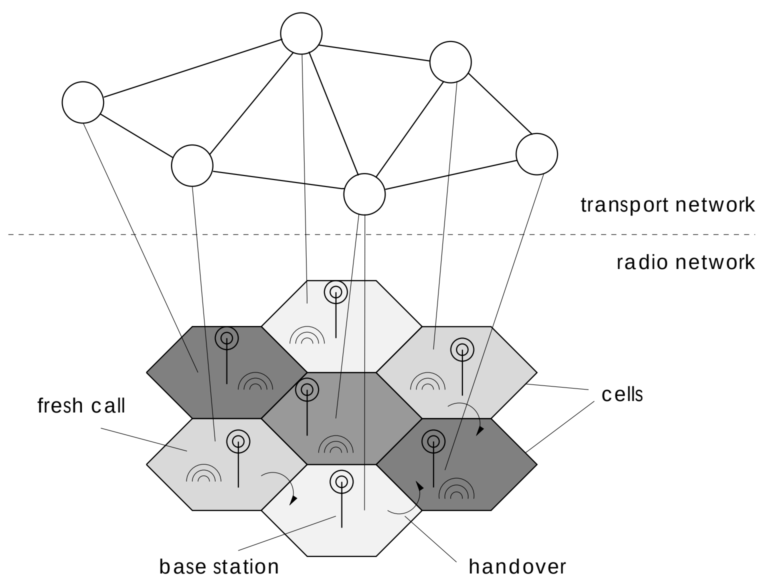

Figure 1: Cellular mobile communication network

capacity by the reuse of the same frequencies in several cells, which are separated in space. This permits the transmission of more than one call on the same physical channel at the same time. The drawback, however, of the cellular concept is the increased system complexity. The system must be able to handle 'handover' events: on-going calls which move across cell boundaries have to be transfered into the adjacent cell without interruption. The event of a new call in a cell, either originating or terminating at the mobile unit, is called a 'fresh call'.

Call handling mechanisms have to perform two tasks. First, they have to manage reliably the admission of fresh calls and handovers to a cell. Second, call handling mechanisms have to ensure that the system operates at high traffic load, i.e., a good handling mechanism has to admit as many calls as possible. However, the disadvantage of running a system close to its limits is the higher vulnerability to a service increase in a very short time scale, e.g. mass calling. Such systems degrade quickly in overload situations. In order to use efficient call handling mechanisms in cellular networks, the configuration of them has to be done very careful and accurate.

The performance of call handling mechanisms in cellular mobile networks depends mainly on three factors. *(i)* First, the performance is influenced by the capability of the system architecture, e.g., by the complexity of the handover protocol and by the performance of the involved transport network elements, cf. Figure 1. *(ii)* In the second place, it is affected by the behavior of the customers within the service area. In order to get a good insight into the performance of the system, we should take into account the distribution of interarrival times and service times, the so-called 'traffic mix', (i.e. the fraction of call requests being handover and fresh calls, respectively), the user mobility, etc. *(iii)* And third, the performance depends on the actual configuration of the network. Loosely speaking, if the architecture has been determined, the parameters of the system have to be chosen, e.g. the proper dimensioning of the number of channels available in a cell. Furthermore, the strength of the influence of these factors depends not only on their direct effect. The factors interact and amplify each other, e.g. the service availability will increase the service demand. Therefore, an efficient network design has to address the factors not in an isolated, but in a comprehensive way.

The contribution of this paper is that it integrates the three aspects mentioned above. We will *(i)* study two handling mechanisms, 'guard channel' and 'handover retry', that prioritize handovers, in order to enhance the system performance. However, as opposed to earlier studies on this subject, we use in our analysis a *(ii)* customer behavior model that enables blocked calls to redial, which is for that reason more realistic. We show how to evaluate the Quality of Service (QoS), under these handling mechanisms and user model. Being able to calculate this QoS, we can choose *(iii)* the configuration, e.g., the number of channels, the number of guard channels, and the parameters of the handover retry procedure. In this text we denote the user initiated repeated attempts to occupy a channel as 'redialing' and the system triggered attempts as 'retrying'.

The organization of the paper is as follows: We first give in Section 2 a short introduction to call handling mechanism of cellular mobile systems, and introduce the schemes that are analyzed in this paper. Section 3 is devoted to the traffic model that we will use throughout this paper. Section 4 is about cell configuration procedures; we explain how this is usually done in cellular mobile networks. Then we argue that this procedure has some deficiencies. Section 5 treats the analysis of our proposal for an advanced call handling mechanism, and shows the efficiency of the suggested mechanisms under our user model. We give some guidelines for selection of the parameters for the new handover mechanism. The last section gives conclusions and remarks.

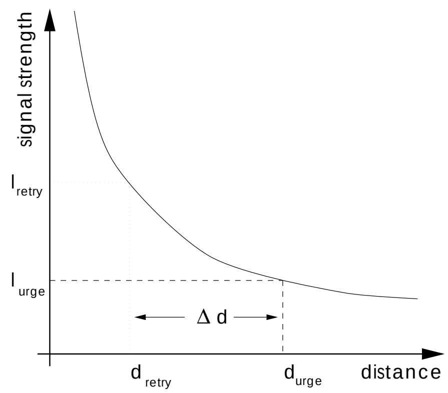

Figure 2: Signal levels for handover commitment

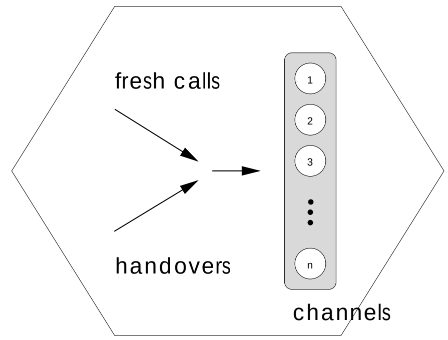

Figure 3: Contenting fresh calls and handover

2 Call handling mechanisms

2.1 Conventional handover mechanisms

In most cellular networks handover events are managed in the following way: the base station and the mobile unit measures regularly the radio signal strength and the mobile unit transmits its measurements to the base station. If the base station detects a decrease in radio signal under a minimal secure level d_{urge}, cf. Figure 2, it initiates a handover request. Usually this type of call transfer is denoted by the term 'rescue handover'. The base station informs the *base station controller* about the request, which than verifies if it is possible to transfer the call into a new cell. Therefore, the controller checks whether there is a free channel available in the new cell, or not. Usually the base station controller does not distinguish between channel requests for fresh calls or handovers. If a handover request can be satisfied, the controller informs the mobile unit to switch to the new cell. If no channel is free, the call is interrupted and is lost.

The drawback of this handover procedure is the fact that the handover requests content for the same channels which are used for fresh calls, cf. Figure 3. Thus, the reliability of this mechanism depends on the load of the new cell. Moreover, since handovers are more valuable than fresh calls, cf. Section 2.2.1, a lost handover decreases the quality of service of the cellular system much stronger than a blocked fresh call.

2.2 Advanced handover procedures

The objective of *advanced* handover procedures is to ensure on the one hand a high traffic load of the cellular system, while one the other hand maintaining a certain quality of service. Therefore we give in this subsection first a definition what we consider as the quality of service in regard to call handling and call admission and present then two advanced call handling mechanisms which are expected to meet these objectives.

2.2.1 Quality of service

In cellular mobile networks, the Quality of Service (QoS) comprises the speech quality as well as the availability of the service within the supplying area. Both factors are mainly determined by the quality of the radio transmission. However, due to the increased traffic in mobile communication networks, the second factor depends more and more on availability of free channels and thus on the proper teletraffic configuration of the system.

From the teletraffic point of view the Quality of Service is determined by the probability of the two events which occur due to the occupancy of all available channels:

a) the fresh call blocking probability P_{BF} and,

b) the handover dropping probability P_{BH}.

Since there is a tradeoff between these two performance measures and the configuration parameter, an overall *cost function* C can be defined as the weighted sum of the blocking probabilities:

$$C = \alpha P_{\text{BF}} + (1-\alpha) P_{\text{BH}}, \tag{1}$$

where $\alpha \in [0;\ 1]$. The value of α indicates the priority of handovers relative to fresh calls. This performance measure was first used by Chang et al. (1994).

One aspect of the cost function measure defined in equation (1) is to focus on blocked handovers. For a subscriber an interrupted call is much more fretful, than a blocked fresh call. Thus, the handover blocking probability should stay one magnitude below the value of the fresh call blocking probability, which is typically in the order of one percent. To equally include the two probabilities into one cost function, we therefore chose $\alpha = \frac{1}{11}$ throughout our study, such that $\alpha/(1-\alpha) = \frac{1}{10}$.

2.2.2 Guard channels

The dependency of the handover procedure on the traffic load of the cell can be reduced by applying channel allocation mechanisms which favor handovers in overload situations. An effective mechanism is the use of *guard channels*. Guard channels are established only when the number of free channels is equal to or less than a predefined threshold g, cf. Figure 4. In this case, fresh calls are rejected and only

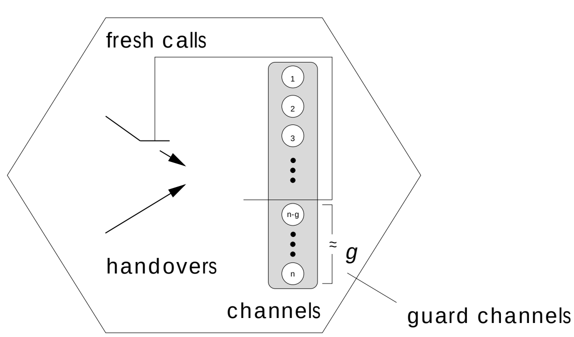

Figure 4: Guard channels for handover request

handover request are served by the cell until all channels are occupied. As soon as the number of unused channels exceeds the threshold, the cell starts again accepting fresh calls.

The guard channel mechanism reduces directly the probability of dropping a handover. Since the guard channel can only be used by handovers, their occupancy depends only on the number of handover requests. However, the reservation of channels for handovers restricts fresh calls from being served and increases their blocking rate.

2.2.3 Handover retry

Handover retry is an other promising mechanism to decrease the blocking probability of handovers. Once a handover has failed, the request enters a retry group, where it attempts a number of retrials before the handover is definitively lost, cf. Figure 5. In a real implementation, the number of handovers in the retry group should be lim-

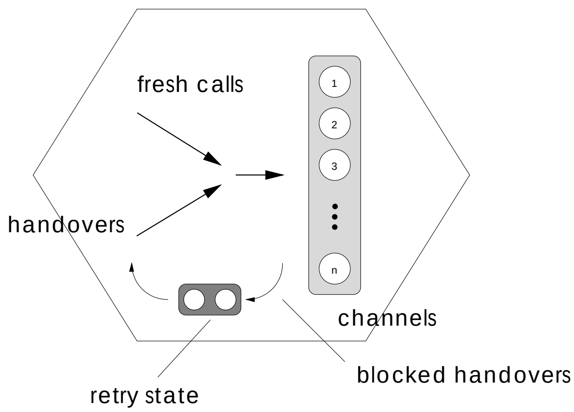

Figure 5: Handover retry mechanism

ited. If many handovers are waiting in the retry group, it is not very probable that a new handover can be served within the time limit a handover can be postponed.

So far the handover retry mechanism has not been implemented in nowadays networks. However upcoming generations of cellular systems will have two technical features which are using this mechanism. The first feature is the so called 'two-level handover'. A handover is triggered already when the strength of radio signal falls below a first threshold l_{retry}, even if the signal strength is still sufficient for radio transmission. The *urgent* handover must be committed, if the radio signal falls under a second threshold l_{urge}, cf. Figure 2. The time, a moving mobile unit needs to cover the distance between the threshold points, can be used for processing the request.
The second technical feature is the 'call re-establishment'. In a radio mobile environment, a call can be always interrupted because of suddenly severe loss of the radio signal due to obstacles. Often, another cell could be used to continue the transmission and a handover request can be initiated to transfer the call. A similar feature already exists in GSM network specifications, Mouly and Pautet (1992) p. 412ff. However, it is not used due to the high complexity which is required in the

controller by the specifications.

Handover retry can be triggered either by a 'centralized' controller or by the handset itself, without user interaction. A centralized controller which manages the retry state is expected to be very efficient. He can occupy resources instantaneously and schedule the requests in an optimal order. However, the mechanism increases the complexity of the controller and is therefore very expensive to implement. A *distributed mechanism* is constituted if the handsets trigger themselves the handover retry. For instance, the handset repeats its request every k time units, with a maximum of K times. Now, the complexity of the call handling mechanism is partly moved out of the controller and distributed into the handset. The efficiency of such a mechanism is not expected to be optimal, but as the performance analysis in Section 5 will show, the mechanism improves the Quality of Service quite well. Furthermore, this type of mechanism very cheap to implement in the handset.

2.3 Overview on analytical models of call handling mechanisms

An early paper on guard channels was published by Guérin in 1988. The model consists of guard channels and a queue of blocked fresh calls. It does not consider any retry state space for handovers. The service of the queue is managed by a central controller. This model was also approached, but from a more computational point of view, by Keilson and Ibe (1995) and Daigle and Jain (1992). Both used a matrix-geometric approach for the analysis.

Chang, Su, and Chiang (1994) considered a more advanced guard channel model. It comprises the guard channel mechanism and two finite queues for waiting fresh calls and handovers. The model allows both types of calls to withdraw from the queues, due to customer impatience or early leaving the handover area, respectively. The queues are managed by a centralized scheduling scheme, i.e. a waiting request is served as soon as a channel becomes idle. A similar model was investigated by Zeng, Mukumoto, and Fukuda (1994). However the authors do not consider the reneging of calls due to impatience.

In contrast to the above mentioned authors, Yoon and Un (1993) investigate two call call handling mechanism which prioritize handovers but do not use guard channels. Both models consists of a single queue for fresh calls and handovers. However, handovers are allowed, if they can not be served immediately, to drive out fresh calls from the head of the waiting line. The two models differ in the handling of fresh calls. The first model considers a last-in,first-out (LIFO) strategy for the service of fresh calls and the second model applies a first-in,first-out (FIFO) scheme. Yoon and Un (1993) compare both models with a guard channel model without a waiting room. This is in fact the classical 'trunk reservation' model.

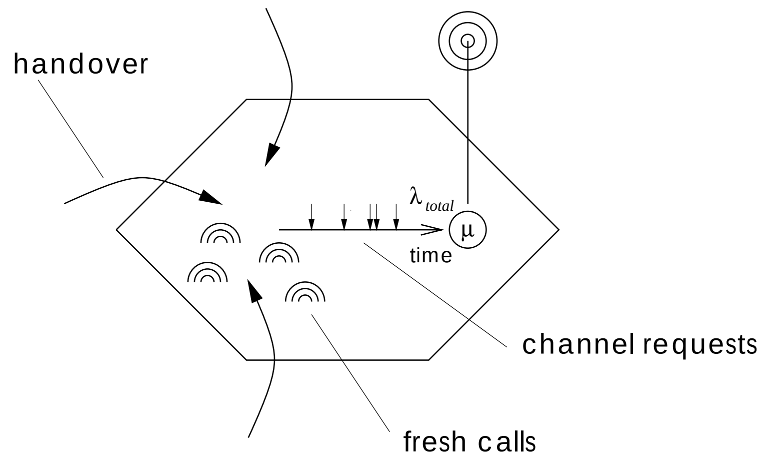

Figure 6: Conventional cellular mobile network traffic model

3 Traffic models for cell dimensioning

3.1 Single request stream model

The commonly used traffic model for a single cell in mobile communication networks is the 'single arrival stream' model. Handovers of on-going calls from adjacent cells and fresh calls are assumed to arrive according to a Poisson process with rates λ_{H} and λ_{F}. Both streams are aggregated in a single sequence of channel requests, cf. Figure 6. Again, the aggregated stream is a Poisson process with λ_{total}:

$$\lambda_{\mathrm{total}} = \lambda_{\mathrm{F}} + \lambda_{\mathrm{H}}. \tag{2}$$

In this model, the the call duration distribution is assumed to be exponential with mean μ^{-1} for both types of traffic.

The major disadvantage of this traffic model is the aggregation of all channel request into one single stream. The model can not be used for the analysis of advanced call handling mechanism, which distinguish between handovers and fresh calls.

Moreover, the 'single stream' model does not consider any temporal user behavior. In overload situations, it is very likely that blocked customer repeat their call attempt after a few seconds. Especially, since modern handsets have the capability to do the redial by just one push of a button. The redial rate is usually one or two magnitudes higher than the normal call arrival rate and thus the offered traffic increases extremely in a very short time scale. It is known from literature on classical switching systems, e.g. Jonin and Sedol (1976) and Macfadyen (1979), that this redial phenomenon degrades the system performance dramatically. An accurate traffic model has to consider this redial behavior in order to avoid the consequences of the phenomenon during system operation.

3.2 Repeated attempts

An important feature of our study is a new user model which considers the impatience of customer and models the redial phenomenon. Therefore we adopt the customer behavior model which was introduced by Tran-Gia (1982). The model assumes, that every time a customer is blocked, he waits for a next attempt with probability θ_0 and will give up the request with the probability $1-\theta_0$. If the customer decides to reattempt, he will try this reattempt after an exponentially distributed time with mean α_0^{-1}. The redial probability θ_0 is not affected by the number of redial already done. In Tran-Gia (1982) an efficient recursive algorithm is given to analyze a loss model used for a finite number of customers. A comprehensive surveys on reattempt queues can be found in Falin (1992) and Yang and Templeton (1987). More recent articles, that specifically study the effect of repeated attempts in cellular mobile networks, are Tran-Gia and Mandjes (1996) and Choi et al. (1995). In contrast to our research, these studies focus more the system degeneration effect of the redial phenomenon.

4 Cell configuration

Conventional cell configuration

The widely used cell configuration method in cellular mobile networks is the application of the Erlang-B-formula, e.g. Mouly and Pautet (1992):

$$P_{\mathrm{B}} = \frac{\frac{(\frac{\lambda_{\mathrm{total}}}{\mu})^n}{n!}}{\sum_{k=0}^{n} \frac{(\frac{\lambda_{\mathrm{total}}}{\mu})^k}{k!}}. \tag{3}$$

The formula relates the offered traffic load $\lambda_{\mathrm{total}}/\mu$ with the number of channels n and the blocking probability P_{B} for all channels requests, under the assumption that the instants of call follow a Poisson process and the call duration is exponentially distributed. The formula does not distinguish between fresh calls and handovers. For cell configuration, it is required that the blocking probability stays below a certain value, in GSM networks typically two percent. For the initial configuration of a cell, where measurements are not available, the offered load λ_{F}/μ is estimated by using the average traffic per subscriber, measured in Erlang, the expected number of subscribers in the cell and the mean call duration. In an operating mobile communication system like GSM, the values of λ_{total} represents the number of call attempts during *busy hour* per time unit. The busy hour is defined as the four consecutive 15min time intervals with highest number of requests. The mean call duration is also measured during the busy hour.

The major drawback of this configuration method is, that it treats fresh calls and handovers in same way. Therefore the method cannot be used for advanced call handling mechanisms, like the two schemes presented in Section 2. Furthermore, since the underlying traffic model does not consider temporal user behavior, the Erlang-B-formula is not even very appropriate in even determing the actual number of required channels for conventional call handling mechanisms. To compensate the insufficiencies, correction factors on the offered traffic λ_{total} can be applied. However, the determination of these factor values is still done by ‘thumb rules’ and therefore the design remains very inaccurate.

Advanced cell configuration

Advanced cell configuration methods have to come up with the right combination of the parameters n, g, k and K of the advanced call handling mechanisms. Furthermore the configuration method has to consider the different objectives, i.e. the both blocking probabilities. Therefore the configuration has to make use *(i)* of an accurate traffic model, like the new one present in Section 3.2, and *(ii)* of a detailed analytical model for the mechanisms. We present and evaluate such a model in the next section.

5 Model and performance evaluation

In this section we will first describe the handling schemes that were presented in Section 2, guard channels and handover retry, in mathematical terms, using the customer model introduced in Section 3. In the second subsection we show that the network can be modeled as a continuous-time Markov chain. We conclude this section with a graphically evaluation of the efficiency of the advanced call handling mechanisms.

5.1 Model

The analytical model of the investigated advanced call handling mechanism is shown in Figure 7. The model comprises a guard channel mechanism as well as a handover retry. The arrival processes are assumed to be Poissonian: the interarrival times of fresh calls and handover calls are exponentially distributed with mean $\lambda_{\mathrm{F}}^{-1}$ and $\lambda_{\mathrm{H}}^{-1}$ and the call duration , i.e. the time before a call terminated or a handover to another cell is attempted, is exponentially distributed with mean μ^{-1}. The fresh call redial behavior is modeled in a way as proposed in Section 3.2. The handover retry mechanism is similar the one presented in Section 2.2.3. Since this kind of handover retry mechanism is notoriously difficult to analyze, we approximate the mechanism by one which a repeats the requests after exponentially distributed time with mean α_1^{-1}. Blocked handovers enter a retry group with probability θ_1, where $\alpha_1 = k^{-1}$ and $\theta_1 = 1 - K^{-1}$. This retry mechanism is parallel to fresh call redial model. The number of handovers in the retry state B is limited, as explained in Section 2.2.3.

5.2 Markov chain

Our model for performance analysis is the following. The state of the system is described by the triple (X, Y, Z). The first coordinate X is the number of channels occupied. Obviously, this number varies between 0 and the number of channels available, say n. Y reflects the number of blocked fresh calls that are redialing. Finally, Z denotes the number of handover calls in their retry queue.

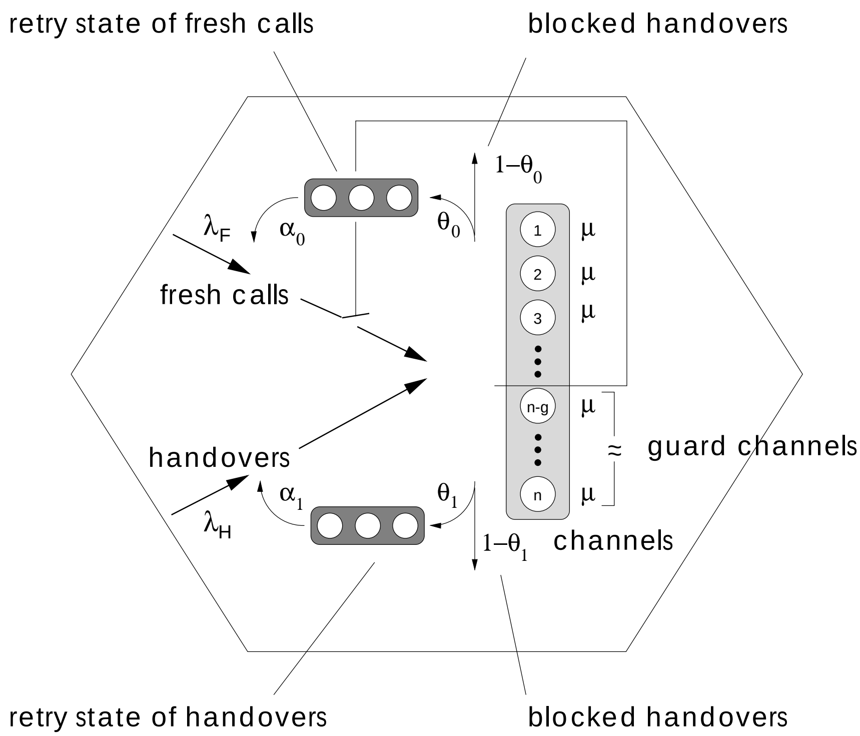

Figure 7: Analytical model of a cell with user redial behavior, guard channels and handover retry

It is standard to derive the state equations, which give the equilibrium distribution of the above Markov chain. Let $x(i,j,k)$ be this long-run distribution of this Markov chain, say $P(X=i, Y=j, Z=k)$, where

$$(i,j,k) \in S := \{0,\ldots,n\} \times \{0,1,\ldots\} \times \{0,\ldots,B\}.$$

Let $x(i,j,k) \equiv 0$ when $(i,j,k) \notin S$. Then the state equations are for $i = 0,\ldots,n-g-1:$

$$x(i,j,k)(\lambda_{\text{F}} + \lambda_{\text{H}} + i\mu + j\alpha_0 + k\alpha_1) =$$

$$x(i-1,j,k)(\lambda_{\text{F}} + \lambda_{\text{H}}) + x(i+1,j,k)(i+1)\mu +$$
$$x(i-1,j+1,k)(j+1)\alpha_0 + x(i-1,j,k+1)(k+1)\alpha_1,$$

i.e., this is the case in which every new request, a fresh call as well as a handover is accepted. For $i \in \{n-g,\ldots,n-1\}$ we have that only handovers are accepted immediately for service. For $i = n-g$ we get:

$$x(i,j,k)(\lambda_{\text{F}}\theta_0 + \lambda_{\text{H}} + i\mu + j(1-\theta_0)\alpha_0 + k\alpha_1) =$$

$$x(i-1,j,k)(\lambda_{\text{F}} + \lambda_{\text{H}}) + x(i+1,j,k)(i+1)\mu +$$
$$+x(i,j-1,k)\lambda_{\text{F}}\theta_0 + x(i-1,j+1,k)(j+1)\alpha_0 +$$
$$x(i,j+1,k)(j+1)(1-\theta_0)\alpha_0 + x(i-1,j,k+1)(k+1)\alpha_1.$$

For $i = n-g+1,\ldots,n-1$:

$$x(i,j,k)(\lambda_{\text{F}}\theta_0 + \lambda_{\text{H}} + i\mu + j(1-\theta_0)\alpha_0 + k\alpha_1) =$$

$$x(i-1,j,k)\lambda_{\rm H} + x(i+1,j,k)(i+1)\mu +$$
$$x(i,j-1,k)\lambda_{\rm F}\theta_0 + x(i,j+1,k)(j+1)(1-\theta_0)\alpha_0 +$$
$$x(i-1,j,k+1)(k+1)\alpha_1.$$

For $i = n$, we have that arriving calls of both types cannot be accepted immediately.

$$x(i,j,k)(\lambda_{\rm F}\theta_0 + \lambda_{\rm H}\theta_1 + i\mu + j(1-\theta_0)\alpha_0 + k(1-\theta_1)\alpha_1 =$$

$$x(i-1,j,k)\lambda_{\rm H} + x(i,j-1,k)\lambda_{\rm F}\theta_0 +$$
$$x(i,j,k-1)\lambda_{\rm H}\theta_1 + x(i,j+1,k)(j+1)(1-\theta_0)\alpha_0 +$$
$$x(i,j,k+1)(k+1)(1-\theta_1)\alpha_1 + x(i-1,j,k+1)(k+1)\alpha_1.$$

Analogously to the results in Tran-Gia and Mandjes (1996), the fresh call blocking probability and handover blocking probability can be given in terms of the equilibrium distribution $x(i,j,k)$. The fresh call blocking probability equals the mean number of blocked fresh calls, i.e. the fresh calls that leave the system before getting a connection per unit time, divided by the mean number of arriving fresh calls per unit time:

$$P_{\rm BF} = \frac{\sum_{j,k}((1-\theta_0)\lambda_{\rm F} + j(1-\theta_0)\alpha_0)x(n,j,k)}{\lambda_{\rm F}}.$$

In the same manner, we get for the handover blocking probability:

$$P_{\rm BH} = \frac{\sum_{j,k}((1-\theta_1)\lambda_{\rm H} + k(1-\theta_1)\alpha_1)(x(n-1,j,k) + x(n,j,k))}{\lambda_{\rm H}}.$$

To evaluate these performance measures, of course the distribution $x(i,j,k)$ must be known. Some comments on how to find this distribution, are provided in the next subsection.

5.3 Evaluation

In this section we will graphically evaluate the mechanisms described in the previous sections. As parameters we chose: the call termination rate $\mu = 1/120$ sec^{-1} and the number of channels $n = 15$. The 'traffic mix', that is the ratio of the mean number of fresh call requests per unit time and the mean number of handover attempts, is $\lambda_{\rm F}/\lambda_{\rm H} = 24$. We consider the performance for different values of the offered load $\rho = (\lambda_{\rm F} + \lambda_{\rm H})/n\mu$, by varying $\lambda_{\rm F}$ and keeping the traffic mix $\lambda_{\rm F}/\lambda_{\rm H}$ constant.

For the evaluation of call handling mechanism, we focus on the three performance measures, as defined in Section 2.2.1: the fresh call blocking probability $P_{\rm BF}$, cf. Figure 8, the handover blocking probability $P_{\rm BH}$, cf. Figure 9, and the cost function C, cf. Figure 10, with $\alpha = 1/11$:

$$C = \frac{P_{\rm BF} + 10P_{\rm BH}}{11}. \tag{4}$$

We will subsequently consider four analytical models of call handling mechanisms, coming closer and closer to the model proposed in this study.

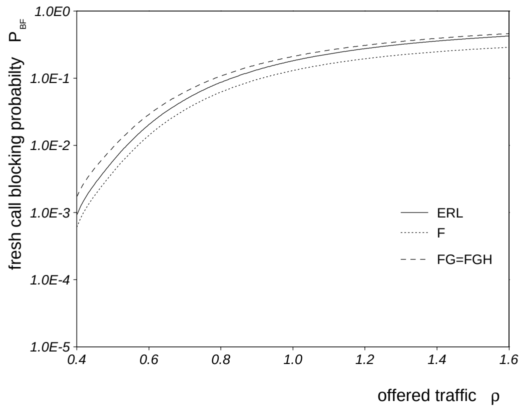

Figure 8: Fresh call blocking propability

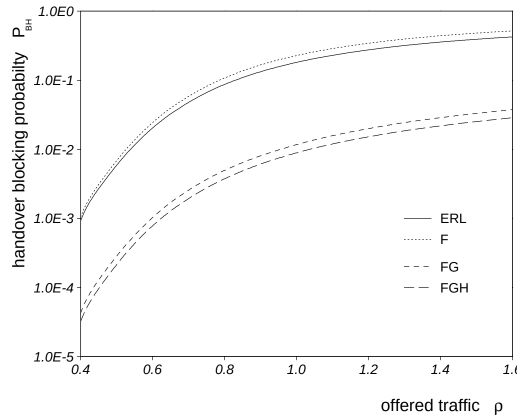

Figure 9: Handover blocking propability

ERL This is the model without guard channels and handover retry, in which the blocked fresh calls are not supposed to redial. This model is a special case of the model presented in this section. Here the parameters are: $g = 0, \theta_0 = \theta_1 = 0$. The probabilities are calculated by the well-known explicit Erlang-B formula, see Section 4.

F The second model is the model without any advanced call handling mechanism, neither guard channels nor handover retry, but with redialing of blocked fresh calls. We take $\alpha_0 = 0.4$ and $\theta_0 = 0.75$. That means that on average four fresh call redials are done with 2.5 seconds in between. The curve with redialing of fresh calls (F) shows that it is very dangerous to neglect the effect of redialing blocked fresh calls: the handover blocking probabilities are considerably larger than on (ERL). On the other hand, of course, fresh call blocking will occur more rarely. However, since the first effect has more impact on the cost function than the second one: the cost function increases increases due to the redialing. The calculations of (F) are done with a recursive algorithm similar to the one presented in Tran-Gia (1982).

FG Two mechanisms were designed to prioritize handovers, as explained in section 2. The first is to add guard channels (FG) to our model with redialing blocked fresh calls (F). We took $g = 1$. The corresponding calculations are done with an algorithm described by Tran-Gia and Mandjes (1996). We see that indeed handover blocking rates, and the cost function as well, decrease.

FGH The second call handling mechanism is the mobile unit triggered retry of blocked handovers. The model that we get is the one shown in Figure 7. Here, we choose as the number of guard channels $g = 1$. Furthermore, we took $\alpha_1 = 1$ and $\theta_1 = 0.75$. This selection means, that on average a handover repeats its request four times, each after one second. The maximum allowed number of calls being in the retry state, B, is two. The curves (FGH) show the effect of the implementation of this mechanism, in addition to the guard channel, in the model with redialing blocked fresh calls. These results were found by solving the balance equations of the Markov chain of the comprehensive model described in Subsection 5.1. We see that the fresh call blocking probability stays on the same level, whereas the handover blocking rate and cost

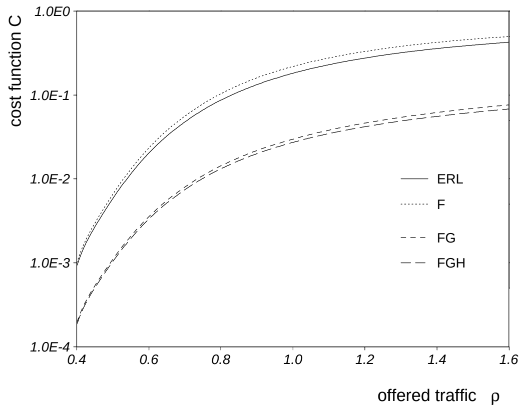

Figure 10: Cost function of the different handling schemes

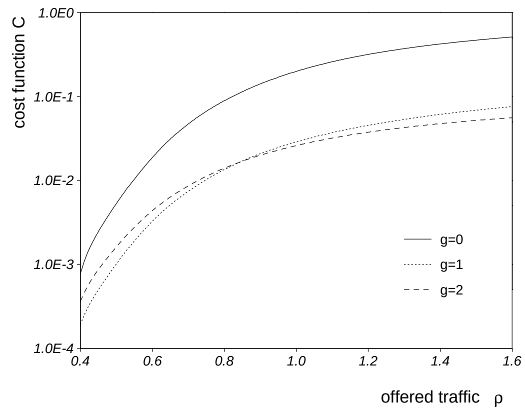

Figure 11: Effect of multiple guard channels

function C decreases, but not as much as by introducing the guard channel.

The number of guard channels

Up till now, we considered only models with a single guard channel. Of course, it is interesting to see the effect is of multiple guard channels. It is clear that, by increasing the number of guard channels g, the handover blocking probability P_{BF} will increase, whereas the fresh call blocking probability P_{BH} decreases. Since the cost function C incorporates both, it is not obvious what is the optimal number of guard channels in order to minimize C. Therefore, in Figure 11 we consider the performance of the (FGH) model for different numbers of guard channels. We see that the largest gain is achieved by introducing one guard channel. Using two guard channels instead of one means, for small values of the offered traffic, even a cost function degradation. Of course, this picture depends heavily on the choice of α in the definition of the cost function, but as a general guideline our analysis shows that mostly one or at most two guard channels are optimal.

6 Conclusion

The performance of call handling mechanisms in cellular networks depends mainly on three factors: *(i)* the call handling scheme, *(ii)* the user behavior and *(iii)* the proper traffic configuration of the cell. To address the first factor, we investigated two advanced call handling schemes, that prioritize handover calls in order to enhance the Quality of Service: guard channels and handover retry. As expected and already discussed in the literature, cf. Guérin (1988), the guard channel mechanism improves strongly the performance of the system. Furthermore our research indicate that it is mostly not necessary to take more than one guard channel. An additional performance improvement can be obtained by using a handover retry mechanisms. The attractive feature of the mechanism, we proposed in our study, is that it is organized in a distributed manner, and therefore does not increase the system complexity. In order to implement this mechanism, only a timer and a counter have to be added in the handsets. The mechanism does not require the controller to record a centralized queue.

To address the second factor, the temporal user behavior, we presented a simple but efficient model that can be used to capture the phenomenon of repeated attempts of blocked calls. As discussed in the evaluation of the mechanisms, in order to guarantee the desired blocking rate for handovers, it is dangerous to neglect the redial behavior of mobile subscriber, as seen from the figures in Section 5.

Both adaptations, advanced call handling schemes and improved user modeling, can be used to obtain a more accurate configuration method of the cells in mobile networks. We showed that such a system can be modeled as a continuous-time Markov chain and that the solution of its balance equations is mathematically tractable. For different scenarios of the parameters, i.e. number of channels n, number of guard channels g, the probability of a next retry of a blocked handover θ_1 and the mean time until the next retry α_1^{-1}, the relevant performance measures can be easily calculated. A network designer can thus select them such that all service criteria, i.e. the blocking probabilities of fresh calls and handovers as well as the cost function, are kept below a certain predetermined value. The new configuration method increases the efficiency of the cellular system even in high load situations.

References

Chang, C.-J., T.-T. Su, and Y.-Y. Chiang (1994). Analysis of a cutoff priority cellular radio system with finite queueing and reneging/dropping. *IEEE/ACM Transactions on Networking 2*(2), 166–175.

Choi, B. D., K. B. Choi, and Y. W. Lee (1995). M/G/1 Retrial queueing systems with two types of calls and finite capacity. *Queueing Systems 19*, 219–229.

Daigle, J. N. and N. Jain (1992). A queueing system with two arrival streams and reserved servers with application to cellular telephone. In *Proceedings of the IEEE INFOCOM '92*, pp. 9C.2.1 – 9C.2.7.

Falin, G. (1992). A survey of retrial queues. *Queueing Systems 7*, 127–168.

Guérin, R. (1988). Queueing-blocking system with two arrival streams and guard channels. *IEEE Transactions on Communications 36*(2), 153–163.

Jonin, G. L. and J. J. Sedol (1976). Telephone systems with repeated calls. In *Proceedings of the 8th Int. Teletr. Congr.*, Munich.

Keilson, J. and O. C. Ibe (1995). Cutoff priority scheduling in mobile cellular communication systems. *IEEE Transactions on Communications 43*(2/3/4), 1038–1045.

Macfadyen, N. W. (1979). Statistical observation of repeated attempts in the arrival process. In *Proceedings of the 9th Int. Teletr. Congr.*, Torremolinos.

Mouly, M. and M.-B. Pautet (1992). *The GSM System for Mobile Communications*. 4, rue Elisée Reclus, F-91120 Palaiseau, France: published by the authors, ISBN: 2-9507190-0-7.

Tran-Gia, P. (1982). Überlastprobleme in rechnergesteuerten fernsprechvermittlungssystemen - modellbildung und analyse. Bericht 36, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart. In German.

Tran-Gia, P. and M. Mandjes (1996). Modeling of customer retrial phenomenon in cellular mobile systems. Forschungsbericht, Preprint-Reihe Nr. 142, Universität Würzburg, Institut für Informatik.

Yang, T. and J. G. C. Templeton (1987). A survey on retrial queues. *Queueing Systems 2*, 203–233.

Yoon, C. H. and C. K. Un (1993). Performance of personal portable radio telephone systems with and without guard channels. *IEEE Journal on Selected Areas in Communications 11*(6), 911–917.

Zeng, Q.-A., K. Mukumoto, and A. Fukuda (1994). Performance analysis of mobile cellular radio system with priority reservation handoff procedures. In *Proc. VTC 44*, Stockholm, Sweden, pp. 1829–1833.

Preprint-Reihe
Institut für Informatik
Universität Würzburg

Verantwortlich: Die Vorstände des Institutes für Informatik.

[90] U. Hertrampf. On Simple Closure Properties of #P. Oktober 1994.

[91] H. Vollmer und K. W. Wagner. Recursion Theoretic Characterizations of Complexity Classes of Counting Functions. November 1994.

[92] U. Hinsberger und R. Kolla. Optimal Technology Mapping for Single Output Cells. November 1994.

[93] W. Nöth und R. Kolla. Optimal Synthesis of Fanoutfree Functions. November 1994.

[94] M. Mittler und R. Müller. Sojourn Time Distribution of the Asymmetric $M/M/1//N$ – System with LCFS-PR Service. November 1994.

[95] M. Ritter. Performance Analysis of the Dual Cell Spacer in ATM Systems. November 1994.

[96] M. Beaudry. Recognition of Nonregular Languages by Finite Groupoids. Dezember 1994.

[97] O. Rose und M. Ritter. A New Approach for the Dimensioning of Policing Functions for MPEG-Video Sources in ATM-Systems. Januar 1995.

[98] T. Dabs und J. Schoof. A Graphical User Interface For Genetic Algorithms. Februar 1995.

[99] M.R. Frater und O. Rose. Cell Loss Analysis of Broadband Switching Systems Carrying VBR Video. Februar 1995.

[100] U. Hertrampf, H. Vollmer und K. W. Wagner. On the Power of Number-Theoretic Operations with Respect to Counting. Januar 1995.

[101] O. Rose. Statistical Properties of MPEG Video Traffic and their Impact on Traffic Modeling in ATM Systems. Februar 1995.

[102] M. Mittler und R. Müller. Moment Approximation in Product Form Queueing Networks. Februar 1995.

[103] D. Rooß und K. W. Wagner. On the Power of Bio-Computers. Februar 1995.

[104] N. Gerlich und M. Tangemann. Towards a Channel Allocation Scheme for SDMA-based Mobile Communication Systems. Februar 1995.

[105] A. Schömig und M. Kahnt. ΔVergleich zweier Analysemethoden zur Leistungsbewertung von Kanban Systemen. Februar 1995.

[106] M. Mittler, M. Purm und O. Gihr. Set Management: Synchronization of Prefabricated Parts before Assembly. März 1995.

[107] A. Schömig und M. Mittler. Autocorrelation of Cycle Times in Semiconductor Manufacturing Systems. März 1995.

[108] A. Schömig und M. Kahnt. Performance Modelling of Pull Manufacturing Systems with Batch Servers and Assembly-like Structure. März 1995.

[109] M. Mittler, N. Gerlich und A. Schömig. Reducing the Variance of Cycle Times in Semiconductor Manufacturing Systems. April 1995.

[110] A. Schömig und M. Kahnt. A note on the Application of Marie's Method for Queueing Networks with Batch Servers. April 1995.

[111] F. Puppe, M. Daniel und G. Seidel. ΔQualifizierende Arbeitsgestaltung mit tutoriellen Expertensystemen für technische Diagnoseaufgaben. April 1995.

[112] G. Buntrock, und G. Niemann. Weak Growing Context-Sensitive Grammars. Mai 1995.

[113] J. García and M. Ritter. Determination of Traffic Parameters for VPs Carrying Delay-Sensitive Traffic. Mai 1995.

[114] M. Ritter. Steady-State Analysis of the Rate-Based Congestion Control Mechanism for ABR Services in ATM Networks. Mai 1995.

[115] H. Graefe. ΔKonzepte für ein zuverlässiges Message-Passing-System auf der Basis von UDP. Mai 1995.

[116] A. Schömig und H. Rau. A Petri Net Approach for the Performance Analysis of Business Processes. Mai 1995.

[117] K. Verbarg. Approximate Center Points in Dense Point Sets. Mai 1995.

[118] K. Tutschku. Recurrent Multilayer Perceptrons for Identification and Control: The Road to Applications. Juni 1995.

[119] U. Rhein-Desel. ΔEine "'Übersicht"' über medizinische Informationssysteme: Krankenhausinformationssysteme, Patientenaktensysteme und Kritiksysteme. Juli 1995.

[120] O. Rose. Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic. Juli 1995.

[121] A. Schömig. On Transfer Blocking and Minimal Blocking in Serial Manufacturing Systems — The Impact of Buffer Allocation. Juli 1995.

[122] Th. Fritsch, K. Tutschku und K. Leibnitz. Field Strength Prediction by Ray-Tracing for Adaptive Base Station Positioning in Mobile Communication Networks. August 1995.

[123] R. V. Book, H. Vollmer und K. W. Wagner. On Type-2 Probabilistic Quantifiers. August 1995.

[124] M. Mittler, N. Gerlich, A. Schömig. On Cycle Times and Interdeparture Times in Semiconductor Manufacturing. September 1995.

[125] J. Wolff von Gudenberg. Hardware Support for Interval Arithmetic - Extended Version. Oktober 1995.

[126] M. Mittler, T. Ono-Tesfaye, A. Schömig. On the Approximation of Higher Moments in Open and Closed Fork/Join Primitives with Limited Buffers. November 1995.

[127] M. Mittler, C. Kern. Discrete-Time Approximation of the Machine Repairman Model with Generally Distributed Failure, Repair, and Walking Times. November 1995.

[128] N. Gerlich. A Toolkit of Octave Functions for Discrete-Time Analysis of Queuing Systems. Dezember 1995.

[129] M. Ritter. Network Buffer Requirements of the Rate-Based Control Mechanism for ABR Services. Dezember 1995.

[130] M. Wolfrath. Results on Fat Objects with a Low Intersection Proportion. Dezember 1995.

[131] S. O. Krumke and J. Valenta. Finding Tree–2–Spanners. Dezember 1995.

[132] U. Hafner. Asymmetric Coding in (m)-WFA Image Compression. Dezember 1995.

[133] M. Ritter. Analysis of a Rate-Based Control Policy with Delayed Feedback and Variable Bandwidth Availability. January 1996.

[134] K. Tutschku and K. Leibnitz. Fast Ray-Tracing for Field Strength Prediction in Cellular Mobile Network Planning. January 1996.

[135] K. Verbarg and A. Hensel. Hierarchical Motion Planning Using a Spatial Index. January 1996.

[136] Y. Luo. Distributed Implementation of PROLOG on Workstation Clusters. February 1996.

[137] O. Rose. Estimation of the Hurst Parameter of Long-Range Dependent Time Series. February 1996.

[138] J. Albert, F. R"ather, K. Patzner, J. Schoof, J. Zimmer. Concepts For Optimizing Sinter Processes Using Evolutionary Algorithms. February 1996.

[139] O. Karch. A Sharper Complexity Bound for the Robot Localization Problem. June 1996.

[140] H. Vollmer. A Note on the Power of Quasipolynomial Size Circuits. June 1996.

[141] M. Mittler. Two-Moment Analysis of Alternative Tool Models with Random Breakdowns. July 1996.

[142] P. Tran-Gia, M. Mandjes. Modeling of customer retrial phenomenon in cellular mobile networks. July 1996.

[143] P. Tran-Gia, N. Gerlich. Impact of Customer Clustering on Mobile Network Performance. July 1996.

[144] M. Mandjes, K. Tutschku. Efficient call handling procedures in cellular mobile networks. July 1996.

[145] N. Gerlich, P. Tran-Gia, K. Elsayed. Performance Analysis of Link Carrying Capacity in CDMA Systems. July 1996.