Scalable Schemes for Mobile Networks with Multiple Services *

Khalid Begain[1], Gunter Bolch[2], Miklós Telek[3]

[1] Department of Computer Science, Mu'tah University
Mu'tah 61710, Jordan, begain@mutah.edu.jo

[2] Department of Computer Science IV, University of Erlangen-Nürnberg
91058 Erlangen, Germany, bolch@informatik.uni-erlangen.de

[3] Department of Telecommunications, Technical University of Budapest,
H-1521 Budapest, Hungary, telek@hit.bme.hu

Abstract

The introduction of new service categories with different bandwidth requirements, e.g., data and multimedia, to cellular mobile radio networks makes many of the traditional mechanisms unusable or less efficient. The call admission and the handover handling are of the most sensitive issues to this extension. The performance of all services including the traditional voice and the new services can be dramatically affected if no appropriate schemes are used. In this paper, we propose call admission and handover handling schemes for a cellular mobile network that offers two service types: voice and data. The data connections are assumed to transmit at different transmission rates that are integer multiples to that of one radio channel. In the case of congestion, the base station asks the active data connections to reduce their transmission rate in order to provide free channels for the newly arrived request of both service types. This is basically intended for incoming handover requests. The request will be rejected if the transmission rate of the active connections reaches a given minimum rate. Similar mechanism can also be used for new call arrivals, but some priority can be given to handovers by setting a higher transmission rate threshold for the rejection. As an extension to the proposed scalability, a queuing of new calls is also proposed and analyzed. Analytical models were built for the two proposed schemes together with the traditional channel reservation scheme. The effect of different traffic and configuration parameters on the performance measures like the grade of service, blocking probabilities, and utilization, are studied using the proposed technique. Results show that the proposed schemes provide very good performance and more fairness among the different service types.

Keywords: Cellular Mobile Network, Call Admission, Handover Handling, Markov models, MOSEL language

*This work was partially supported by the German-Hungarian R&D program and OTKA F-23971 and partially by Alexander von Humboldt Stiftung, Bonn, Germany

1 Introduction

Mobility is one of the main features of all aspects of todays life. The mobile telephony has introduced a major breakthrough in the communication industry and, therefore, the cellular mobile telecommunication is one of the most developing areas in the field. In the first generations of wireless networks, most efforts were concentrated on the telephony services. Mainly to increase the capacity of the telecommunication system to provide more and better quality voice connections. Recently, the need for mobility has also extended to data communications. More and more applications arise were mobile data connections are required. This appears natural with the rapid growth of the number and the capability of notebook and portable computers and with the newly announced rather small mobile terminals that integrate many types of services like fax, e-mail, and web browsing for the mobile users. In existing networks, the data services are served with telephone connections using modems. Nevertheless, it is clear that the transmission rates of one voice channel will not be sufficient for more sophisticated data services like image transfer or multimedia.

The performance of cellular telecommunication networks offering telephony services was investigated by many authors with different assumptions and operation conditions, but only few results have appeared regarding networks supporting different classes of services. Most of the published papers try to give proposals for the solution of the problems associated with the mobile nature of the network, like the termination of ongoing call due to handover failures and new call blocking. Mainly, the former issue is addressed because the handover failure is considered to be worse than new call rejection because it results in the interruption of an ongoing connection which means in the case of data, for example, the need to repeat the whole transmission again. While in the later case, the user is forced to repeat the attempt some time later.

In [1], techniques for reducing the handover failure probability are proposed and several priority based schemes are defined and evaluated in which channels are reserved for handover. The performance of personal networks based on micro-cells covering city streets was investigated in [2] and many important teletraffic parameters were evaluated. In [3], the performance of hierarchical cellular system based on micro-cells and overlaying micro-cells is analyzed. Handovers between micro-cells belonging to different hierarchical levels are introduced and analyzed.

New access protocols for the integration of different service classes in packet switched environment, mainly based on Packet Reservation Multiple Access (PRMA) are introduced and analyzed in [5] and [6]. A model of circuit switching cellular system with m classes of services is introduced in [4]. Each class is defined by different resource and performance requirements and therefore has different blocking behavior.

In [7], a cellular mobile network offering voice, data, and multimedia services is studied. Here, a new technique for reducing the handover failure probability of multimedia connections is introduced by using the so-called *Decoupling Technique*; the data component of the multimedia connection can be temporarily suspended and only the voice part continues in the case of shortage of capacity at the destination cell of handover. The paper introduces a Markov model and presents results on the carried traffic, handover failure probabilities, and new call blocking probabilities.

The performance of cellular mobile network with multimedia services was addressed in [9] using approximate queuing networks algorithm which allows the study of networks comprises large number of cells.

As a previous work, the decoupling technique was extended in [10] and used for data connections assuming that the data mobile terminals (MT) can adapt their transmission rate to the availability of bandwidth and the congestion status of the cellular network to avoid the interruption of ongoing connections due to handover failures.

In this paper, we introduce schemes for both new call and incoming handover requests based on the scalability of the transmission rate at which the data mobile terminals can transmit and on the queuing scheme of the new call requests.

The paper is organized as follows. Section 2 introduces the investigated system by defining the main features, assumptions, and the proposed schemes. In Section 3, a Markov model is developed for each of the schemes. Section 4 defines the main performance parameters of interest. The numerical results are presented in Section 5 followed by some concluding remarks in the end of the paper.

2 The Studied System

2.1 Basic Assumptions

The main features and assumptions of the system are the following:

- The studied cellular mobile telecommunication system comprises large number of cells and offers two types of services; voice and data connections.
- In each cell, the base station manages N radio channels.
- Each voice connection requires 1 channel
- The network offers data connections with maximum transmission rate C_{dmax} which is assumed to be an integer multiple of the transmission rate of one radio channel. Therefore, we shall express this value in the unit of (transmission rate of) one (radio) channel.
- We assume a finite number, M, of potential mobile terminals in the cell.

2.2 New Call and Handover Handling Schemes

In the followings, we define three schemes to deal with both new calls and incoming handover calls of the service types. The first is based on the channel reservation technique which is widely used in cellular networks with only voice connections. The main reason for including this technique is to provide a reference. The second mechanism is one of the proposed ones which is based on the scalability of the transmission rate of data in the mobile terminals. The third mechanism extends the second one by adding queuing capabilities.

2.2.1 Handover priority based on channel reservation

- In each cell, there are C_h reserved channels for handover requests, so they are not available for the newly arrived calls.
- If, at the instant of new call admission of any type, there is no enough "available" channels in the cell (C_h+1 for voice and $C_h + C_{dmax}$ for data), then the call will be rejected.
- In the case of incoming handover request from a neighboring cell, the request is only rejected if there is no enough "free" channels in the cell (1 channel for voice and C_{dmax} for data).

2.2.2 Scalable Data Connection scheme

In this scheme, the data terminals are assumed to be able to adapt their transmission rate according to control information received from the base station reflecting the congestion status of the network. Nevertheless, for an ongoing connection in a given cell, there is a minimum transmission rate C_{dmin} at which the data MT is always allowed to transmit. This value is set to maintain the connection by many data transmission protocols such as TCP/IP. Again, we assume that the possible transmission rates (including C_{dmin}) are all integer multiples of the transmission rate of one channel. Furthermore, we define the transmission rate threshold C_{dnew}, where $C_{dmin} \leq C_{dnew} \leq C_{dmax}$, for controlling the priority level of handover requests over the new call arrivals. The technique works as follows:

- We assume that all ongoing data connections run on the same transmission rate and that they always up-speed as far as there are enough free channels in the cell but not exceeding C_{dmax}.
- In the case of incoming handover request of both voice and data, if the number of free channels is not enough to serve the request, then the base station will ask all ongoing data connections to reduce their transmission rate in order to provide space for the incoming request but not less than C_{dmin}. The incoming connection transmit at the same transmission rate as the ongoing ones.
- The incoming handover request will be rejected only if, at the instant of request, there are not enough free channels and the actual transmission rate of the active data connections would drop bellow C_{dmin} because of the acceptance of the new request.
- In the case of new call arrival (voice or data), the system behaves in the same way as handover except that the ongoing data connection will be asked to reduce their rate only down to C_{dnew}. Note, that $C_{dnew} = C_{dmax}$ means that no rate reduction is allowed for the advantage of new calls so that they will be rejected immediately if no enough channels are available. On the other hand, if $C_{dnew} = C_{dmin}$, then new calls are treated in the same way as incoming handovers; that is handover has no priority over new calls.

2.2.3 Scalable Data Connection with Queuing

In this proposal, we extend the above mentioned behavior so that we assume small buffers for new call requests of both types of services. The queuing of handover requests is also possible but the queuing of new call arrivals is less sensitive regarding the queuing time than the case of handover. In the later case, the handover requests are queued during the time interval which the mobile terminal spends in the handover area that is in the overlapping area between two cells. During this time, its communication with one of the base stations degrades with a rate depending on various factors, such as its velocity, direction of travel, and so on. The most common queuing scheme for both cases is FIFO. The queuing mechanism was proposed for cellular network with only voice calls in [11] and studied in [8]. In this paper, we restrict the queuing mechanism to the new call requests because, as we shall show with numerical results, that the queuing of handover requests will not add much to the improvement on handover failure probability achieved with the scalability of data connections as described before. Furthermore, we assume that voice have priority over data so that no data request will be served from the data buffer if there is a waiting voice request in queue.

2.3 The Driving Processes

The analysis is focused only on one cell of the cellular mobile system, whose behavior is isolated from those of the other cells, that are collectively described only through the handover requests toward the investigated cell. Of course, this assumption stands for the case when the network is highly symmetric and the traffic is homogeneous and in this case the analysis of one cell may suffice for assessment of the quality of service offered by the whole system. Otherwise, many cells must be jointly or independently studied.

The main processes in the considered cell are:

1. The new call arrival processes for voice and data are assumed to be Poisson processes with arrival rates λ_v and λ_d for voice and data, respectively. In real systems these arrival rates are not constant because of the varying inactive population in the cell and generally computed as the product of the new call rate of one MT (λ_{v0} and λ_{d0}) and the number of MTs in the cell that are not involved in ongoing voice and data connections.

2. The incoming handover request arrival processes from adjacent cells follow Poisson processes with mean arrival rates λ_{hv} for voice and λ_{hd} for data.

3. The time of connections until normal termination, *call duration or holding time*, are random variables which are assumed to be independent and exponentially distributed with mean $1/\mu_v$ in the case of voice and $1/\mu_d$ in the case of data.

4. The amount of time that an ongoing call remain in the area of the investigated cell is called *dwell time*, if the call is still active after this time, the MT requests a handover toward an adjacent cell. The dwell time of connections are assumed to be independent exponentially distributed random variables with mean $1/\mu_{hv}$ and $1/\mu_{hd}$ for voice and data, respectively. This parameters depend on the average velocity of the mobile terminals and on the diameter of the cell.

3 The Analytical Model

The state of the cell, at any time instant, is determined by the number of active calls of each class of traffic, so we define it as the vector:

$$X = (v, d)$$

where v is the number of active voice calls, d is the number of active data connections in the considered cell.

Let $n_{occ}(X)$ denote the number of occupied channels in the cell. It is clear that a permissible system must satisfy the condition $n_{occ}(X) \leq N$. Let K denote the number of feasible states. Then, we can define Ω to denote the state space of the system given that the states are conveniently ordered from $0, \ldots, K-1$. Let $\mathbf{Q}$ denote the infinitesimal generator matrix of the Markov model. The resulting model is thus homogeneous and irreducible on the finite state space Ω and therefore the steady state distribution $\mathbf{p} = \{p_x\}$, $x = 0..K-1$, exists, unique and can be computed through the matrix equation $\mathbf{p} \cdot \mathbf{Q} = \mathbf{0}$, subject to $\sum_{x=0}^{K-1} p_x = 1$.

The transition rates; the elements of matrix $\mathbf{Q}$ can be obtained from the analysis of the driving process in the system. The result of the analysis can be different depending on the scheme used. Therefore, we shall study the case of each scheme giving the rules and rates for the possible state transitions following the style used in the rule definition syntax of

MOSEL language, [12], which is based on a "Which state follows (v, d) if..." logic. In the followings, we briefly summarize the possible cases.

3.1 Channel Reservation

In this case, the number of occupied channels can be computed for each system state $(x = 1, \ldots, K-1)$ as:

$$n_{occ}(x) = v(x) + d(x) * C_{dmax},$$

where $v(x)$ and $d(x)$ are the number of active voice and data connections in state x, respectively. In the rest of the paper, we shall ignore writing the index (x) and we shall indicate it only when necessary.

The size of the state space is given by:

$$K = (N+1) * \left(\left[\frac{N}{C_{dmax}} \right] + 1 \right),$$

where $[y]$ is the integer part of y.

The rules and the rates for the possible transitions are:

- **New call requests:** A new call is accepted if the number of free channels excluding those reserved for handovers, is enough for the new request. The first part of Table 1 shows for both types of new calls the conditions, the successor states, and the transition rates.

- **Incoming handover requests:** An incoming handover request is accepted if the number of free channels is enough for the new request. (See the second part of Table 1.) Note that the level of priority to handover is express with C_h. If $C_h = 0$ then handover requests have the same treatment as new calls.

- **Completion of calls and outgoing handover requests:** Since only one cell is analyzed, the completion of a call and the outgoing handover request (either successful or unsuccessful) have the same effect on the model, that is, some previously occupied channels become free (last part of Table 1).

Table 1: Transition rules and rates for channel reservation scheme

Event	Service	Condition	Successor State	Rate
New Call Arrival	Voice	$n_{occ} \leq N - C_h - 1$	$v+1, d$	λ_v
	Data	$n_{occ} \leq N - C_h - C_{dmax}$	$v, d+1$	λ_d
Incoming Handover	Voice	$n_{occ} \leq N - 1$	$v+1, d$	λ_{hv}
	Data	$n_{occ} \leq N - C_{dmax}$	$v, d+1$	λ_{hd}
Outgoing Ho. and Call Termination	Voice		$v-1, d$	$v \cdot (\mu_v + \mu_{hv})$
	Data		$v, d-1$	$d \cdot (\mu_d + \mu_{hd})$

3.2 Scalable Data Connections

Given the scalability of data connections and the assumption that all active data connections will transmit at the highest possible transmission rate ($\leq C_{dmax}$), the number of occupied channels can be computed for each system state as:

$$n_{occ} = \min(v + d * C_{dmax}, N).$$

Furthermore, we define the following *state dependent* quantities:

- The actual transmission rate of the active data connections is given by:

$$C_{dact} = \min\left(\left\lfloor \frac{N-v}{d} \right\rfloor, C_{dmax}\right).$$

  Obviously, $C_{dmin} \leq C_{dact} \leq C_{dmax}$.

- The threshold for the acceptance of new calls:

$$n_{nacc} = v + d * C_{dnew}.$$

- The threshold for the acceptance of incoming handover requests:

$$n_{hacc} = v + d * C_{dmin}.$$

The size of the state space in this case is given by:

$$K = (N+1) * \left(\left\lfloor \frac{N}{C_{dmin}} \right\rfloor + 1\right).$$

Given the above quantities, the rules and the rates for the possible state transitions are:

- **New call requests:** A new call is accepted if there are enough free channels or if there is still a possibility to reduce the transmission rate of the active data connections ($C_{dact} > C_{dnew}$). The first part of Table 2 shows for both types of new calls the conditions, the successor states, and the transition rates.

Table 2: Transition rules and rates for scalable data connections scheme

Event	Service	Condition	Successor State	Rate
New Call Arrival	Voice	$n_{nacc} \leq N-1$	$v+1, d$	λ_v
	Data	$n_{nacc} \leq N-C_{dnew}$	$v, d+1$	λ_d
Incoming Handover	Voice	$n_{hacc} \leq N-1$	$v+1, d$	λ_{hv}
	Data	$n_{hacc} \leq N-C_{dmin}$	$v, d+1$	λ_{hd}
Outgoing Ho. and Call Termination	Voice		$v-1, d$	$v \cdot (\mu_v + \mu_{hv})$
	Data		$v, d-1$	$d \cdot (\frac{C_{dact}}{C_{dmax}}\mu_d + \mu_{hd})$

- **Incoming handover requests:** A newly arrived incoming handover request is accepted if there are enough free channels or if there is still a possibility to reduce the transmission rate of the active data connections ($C_{dact} > C_{dmin}$) (Second part of Table 2).

- **Completion of calls and outgoing handover requests:** Just like in the previous scheme, the completion of a call and the outgoing handover request have the same effect on the model, that is, some previously occupied channels become free (last part of Table 2). If ($C_{dact} < C_{dmax}$) and the number of free channels after the departure of the leaving call is sufficient ($N - n_{occ} > d$), then all the data connections will up-speed to the highest possible rate.

3.3 Scalable Data Connections with Queuing

In this case, we assume two buffers; one for the new voice call requests with capacity of Q_v requests and another for new data requests with capacity of Q_d requests. We extend the description of the state as follows:

$$X = (v, d, q_v, q_d)$$

where q_v and q_d denote the number requests in the new request buffer and handover request buffer, respectively.

The size of the state space in this case is given by:

$$K = (N+1) * \left(\left[\frac{N}{C_{dmin}}\right] + 1\right) * (Q_v + 1) * (Q_d + 1).$$

The calculation of the rest of the quantities is the same as in the previous case. The effect of queuing is shown in the description of the rules and the rates for the possible transition as follows:

Table 3: Transition rules and rates for queuing scheme

Event	Service	Condition	Successor State	Rate
New Call Arrival	Voice	$(n_{nacc} \leq N-1) \wedge (q_v = 0)$	$v+1$	λ_v
		$(n_{nacc} > N-1) \wedge (q_v \leq Q_v - 1)$	$q_v + 1$	λ_v
	Data	$(n_{nacc} \leq N - C_{dnew}) \wedge (q_v = q_d = 0)$	$d+1$	λ_d
		$(n_{nacc} > N - C_{dnew}) \wedge (q_d \leq Q_d - 1)$	$q_d + 1$	λ_d
Incoming Handover	Voice	$n_{hacc} \leq N-1$	$v+1$	λ_{hv}
	Data	$n_{hacc} \leq N - C_{dmin}$	$d+1$	λ_{hd}
Outg. Ho. and call term.	Voice		$v-1$ &	$v \cdot (\mu_v + \mu_{hv})$
	Data		$d-1$ &	$d \cdot (\frac{C_{dact}}{C_{dmax}}\mu_d + \mu_{hd})$
Serving queues	Voice	$(n_{nacc} \leq N-1) \wedge (q_v > 0)$	$v+1, q_v - 1$	
	Data	$(n_{nacc} \leq N - C_{dnew}) \wedge (q_v = 0) \wedge (q_d > 0)$	$d+1, q_d - 1$	

- **New call requests:** A new call is accepted if there are enough free channels or if there is still a possibility to reduce the transmission rate of the active data connections ($C_{dact} > C_{dnew}$). Otherwise, the request will be queued in the corresponding buffer (voice or data) unless that buffer is full. The first part of Table 3 shows for both types of new calls the conditions, the successor states, and the transition rates. A minor difference in this table, compared to the previous two, is that in describing the successive state, we only indicate the elements that has changed e.g., $(v+1)$ instead of $(v+1, d, q_v, q_d)$ because of space limitations.

- **Incoming handover requests:** are treated in the same manner as in the previous scheme.

- **Completion of calls and outgoing handover requests:** After the release of the channels occupied by the terminating call or the outgoing handover request, if there are waiting voice calls, then they will be served immediately (as many as the released channels). If there is no more waiting voice call requests then waiting data requests will be served (as far as the free channels are enough). If there are no waiting call requests of either type then the system behaves as in the previous scheme. The last part of Table 3 indicates the above description.

4 Performance Measures

Whenever the steady state probabilities are computed, a number of interesting performance measures can be evaluated. Among these, we shall first define the following parameters for which we provide formulas for each of the three handover scheme presented in the paper:

- *New Call Blocking Probability, P_n* (P_{nv} for voice and P_{nd} for data) is defined as the fraction of new call requests that cannot be served by the base station due to the lack of free channels.

- *Handover Failure Probability, P_h* (P_{hv} and P_{hd}) is the average fraction of incoming handover requests that cannot be satisfied and thus resulting in a termination of the connection.

- *Grade of Service, GoS* is defined as a combined measure of the previous two probabilities as

  $$GoS_v = P_{nv} + 10P_{hv}$$

  $$GoS_d = P_{nd} + 10P_{hd}$$

  for each service type [18].

- *Cell Utilization, U* defined as

  $$U = \frac{\overline{n_{occ}}}{N},$$

  where $\overline{n_{occ}}$ is the average number of channels in use in the cell determined by:

  $$\overline{n_{occ}} = \sum_{x=0}^{K-1} n_{occ}(x) \cdot p(x)$$

  where $n_{occ}(x)$ is the number of occupied channels when the system in state x as defined in Section 3.

- *Average Data Transmission Rate, C_{dav}* is the average data transmission rate of a randomly chosen data connection.

The calculation of some of these measures depends on the used scheme for handover handling. We have already showed the formulas to calculate n_{occ} in Section 3. In the followings, we provide formulas for P_n, P_h and C_{dav} for each of the presented schemes:

- *Channel Reservation*

  The new call blocking probabilities for voice and data are:

  $$P_{nv} = \sum_{x:n_{occ} > N - C_h - 1} p(x)$$

  $$P_{nd} = \sum_{x:n_{occ} > N - C_h - C_{dmax}} p(x).$$

The handover failure probabilities of voice and data connections are:

$$P_{hv} = \sum_{x:n_{occ}>N-1} p(x)$$

$$P_{hd} = \sum_{x:n_{occ}>N-C_{dmax}} p(x).$$

Note that if $C_h = 0$ then blocking probability of new voice call requests and the voice handover failure probability are equal.

Obviously, the average data transmission rate in this case is C_{dmax}.

- *Scalable Data Connections*

  The new call blocking probabilities for voice and data are:

$$P_{nv} = \sum_{x:n_{nacc}>N-1} p(x)$$

$$P_{nd} = \sum_{x:n_{nacc}>N-C_{dnew}} p(x).$$

As can be seen the ratio of the new call blocking probability of voice and data, i.e., the fairness, depends on C_{dnew}. If $C_{dnew} = 1$ then $P_{nv} = P_{nd}$, and the greater is C_{dnew} the higher is the priority of voice connections over the data ones.

The handover failure probabilities of voice and data connections are:

$$P_{hv} = \sum_{x:n_{hacc}>N-1} p(x)$$

$$P_{hd} = \sum_{x:n_{hacc}>N-C_{dmin}} p(x).$$

The average data transmission rate:

$$C_{dav} = \frac{\sum_{x=0}^{K-1} C_{dact}(x).d(x) \cdot p(x)}{\sum_{x=0}^{K-1} d(x) \cdot p(x)}$$

where $d(x)$ is the number and $C_{dact}(x)$ is the actual transmission rate of the active data connections in state x.

- *Scalable Data Connections with Queuing*

  The only difference from the above is in the calculation of the new call blocking probabilities for both voice and data which becomes:

$$P_{nv} = \sum_{x\in\{x:q_v>Q_v-1\}} p(x)$$

$$P_{nd} = \sum_{x\in\{x:q_d>Q_d-1\}} p(x).$$

Table 4: Parameter values used for obtaining the numerical results

Parameter	Values
N	64
C_{dmax}	8
C_{dmin}	2
$Q_v = Q_d$	2
C_h	4, 8, 16
$1/\mu_v = 1/\mu_d$	100s
$1/\mu_{hv} = 1/\mu_{hd}$	80s
$1/\lambda_{hv} = 1/\lambda_{hd}$	80s

5 Numerical Results

In this section, we present numerical results to show the effect of using each of the proposed schemes on the above defined performance parameters by varying a number of parameters like the number of reserved channels for handover in the first scheme, and the level of priority given to handover over new calls expressed by varying the threshold C_{dnew} for the other schemes. Mainly, we shall consider the Grade of Service GoS achieved for voice and data for each scheme and we assume that the threshold for the acceptable GoS is 1% [18]. All results are presented as curves versus the offered traffic expressed in *Erlang* [17]. The fraction of voice calls is assumed to be 75% of the total offered traffic, and the remaining 25% being data calls.

The studied cell configuration is built on similar assumptions being considered by ETSI (the European Telecommunication Standards Institute) for the introduction of (moderately) high speed data services within the European wireless telephony network [18]. Table 4 provides a list of values for the parameters used in the study. The models have been built using MOSEL language [12] and solved using SPNP package [13]. For detailed description on MOSEL and SPNP, the reader is referred to the appropriate references.

We first consider a cellular network using the channel reservation scheme defined as above. Figures 1 and 2 show the GoS achieved for voice and data. The curves show clearly that this scheme cannot achieve the acceptable level of GoS even for low values of offered traffic and high values of C_h. Furthermore, Figure 3 shows clearly that increasing C_h is not a solution because this improves on the handover failure probability on the expense of new call blocking probability which results in negative overall effect on GoS. Furthermore, it decreases the utilization significantly. The main conclusion of this is that the traditional schemes that are used in cellular telephone networks will not be applicable in cellular networks offering multiple services with different data rate requirements.

Now, we consider configurations based on the proposed schemes with scalable data connections. First, Figures 4 and 5 show GoS achieved for voice and data, respectively, for different values of C_{dnew}. The value of C_{dnew} varies between $C_{dmin} = 2$ (which means that new calls and handover requests are treated similarly) and $C_{dmax} = 8$ (new calls have no advantage of the scalability). The curves show that the use of scalability decreases significantly the value of the GoS for both service types allowing high value for the offered traffic within the acceptable range of GoS (less than 1%). The figures also show that the lower the value of C_{dnew} the better is the result. An interesting observation is that for a very high offered traffic the $C_{dnew} = 2$ became the worst. This can be explained by considering the change in C_{dnew} has a controductory effect on the new call blocking probability and the handover failure probability of both voice and data. Figures 6 (new data call blocking probability) and 7 (data handover failure probability) demonstrate that as an example. Fig. 6 shows that the new data call blocking probability follows a regular behavior while the handover failure probability is extremely low for all $C_{dnew} > 2$ cases and practically has no effect on the value of GoS except for the case $C_{dnew} = 2$.

In general, the curves of GoS show very good performance and that the cellular network using this scheme is capable of handling very high value of offered traffic (40-50 Erlangs) with GoS values under the level of acceptance (1%). Obviously, the average transmission rate of the active data connections will suffer of some degradation because of the scalability introduced. The level of degradation is a function of both the offered traffic and the value of C_{dnew} (Figure 8). Finally, Figure 9 shows that the scheme allows a fairly high utilization of the capacity of the cell for all values of $C_{dnew} < 8$. As a final conclusion of the previous reasoning, we can state that good performance can be achieved allowing high values of offered traffic using this scheme. For example: For $C_{dnew} = 4$, the acceptable value of GoS can be achieved up to offered traffic value of 42 Erlang. With this traffic, the utilization is 77%, and the average transmission rate is 85% of the maximum transmission rate.

From the above results, it becomes clear that if we need higher value of offered traffic or higher utilization in the cell the lower values of C_{dnew} should be used. But this will result also in a high degradation in the average of the transmission rate of active data connections which could result in more unsatisfaction of the data users. Instead, we may use the third scheme which applies queuing besides scalability. Even with a very small buffers, $Q_v = Q_d = 2$ (which means $\sim$ 2-3 sec. average delay), for new calls, the improvement is noticeable. Figures 10 and 11 show GoS achieved for voice and data, respectively.

Here, the values of offered traffic within the acceptable range are more with 5-8 Erlangs than in the former case. Figure 12 shows that no significant change has been noticed on the average transmission rate, while the utilization is noticeably better in this case (Figure 13). To show more clearly the improvement we consider again the case of $C_{dnew} = 4$, where now the acceptable value of GoS can be achieved up to offered traffic value of 48 Erlang. With this traffic, the utilization is 88%, and the average transmission rate is 78% of the maximum transmission rate.

6 Conclusions

The paper presents new handover handling schemes for cellular mobile telecommunications networks offering voice and data service categories and build an analytical model for studying the network with the new schemes. The new schemes are based on the scalability of the transmission rate at which the mobile data terminals are able to transmit. The MTs adapt their transmission rate according to control information received from the base station which reflects the congestion situation of the cell. The results show that the traditional schemes based on giving priority to the handover request by channel reservation are not applicable to cellular networks with multiple service types with different bandwidth requirement. The analytical study of the new schemes showed that they are able of providing extremely low values for handover failure probability and new call blocking probability and therefore for GoS of both traffic types. The fact which allow very good utilization of the network at fairly high offered traffic and within the acceptable level of GoS (1%). It is also noticeable that this proposal provide good fairness among the two service types.

We think that the proposed algorithms will give excellent performance even if they are slightly simplified or the scalability was restricted due to some implementation aspects and give a good base for future generations of cellular networks.

Some additional work will be done on extending the use of this technique to other types of services, like multimedia. We shall also improve on the modeling of the handover procedure and connect it to a more realistic behavior. Work is also going on building a simulation environment to have more validation for the obtained results.

References

[1] D. Hong, S. Rapport, Traffic Model and Performance Analysis for the Cellular Mobile Radio Telephony System with Prioritized and Non-prioritized Handoff Procedures, *IEEE Transaction on Vehicular Technology,* Vol. VT-35, No. 3, August 1986.

[2] R. Steele, M. Nofal, Teletraffic Performance of Microcellular Personal Communication Networks, *IEE Proceedings-I* Vol. 139, No. 4, August 1992.

[3] S. Marano, C. Mastroianni, A Hierarchical Network Scheme for Multilayered Cellular Systems. *Proc. of the VTC 1997* Phoenix, Arizona, USA, May 1997.

[4] S. Tekinay, B. Jabbari, A, Kakaes, Modeling of Cellular Communication Networks with Heterogeneous Traffic Sources, Proc. of ICUPC, 1993

[5] S. Kumar, D.R.Vaman, An Access Protocol for Supporting Multiple Classes of Services in a Local Wireless Environment, *IEEE Transaction on Vehicular Technology,* Vol. VT-45, No. 2, May 1986.

[6] P. Narasimhan, R.D. Yates, A New Protocol for the Integration of Voice and Data over PRMA, *Proc. of PIMRC*, 1995

[7] M. Ajmone Marsan, S. Marano, C. Mastroianni, M. Meo, Performance Analysis of Cellular Mobile Communication Networks Supporting Multimedia Services, in *Proceedings of MASCOTS'98*, July 19-24,1998, Montreal, Canada

[8] F. J. Jaimes, D. Munoz Rodriguez, C. Molina, H. Tawfik, Modeling Resource Management in cellular Systems Using Petri Nets, *IEEE Transaction on Vehicular Technology,* Vol. VT-46, NO 2, 1997.

[9] S. Greiner, G. Bolch, K. Begain, A Generalized Analysis Technique for Queuing Networks with Mixed Priority Strategy and Class Switching. To appear in *Computer Communications*, Kluwer, 1998.

[10] K. Begain, G. Bolch, H. Herold, Analytical Model of Cellular Mobile Networks with Adaptive data connections, In ESM'98, Manchester, UK, 1998, SCS International.

[11] S. Tekinay, An Effective Prioritization Scheme for Handovers in Cellular Networks, in *Proceeding of ICUPC'92*, 1992, pp. 264-268.

[12] G. Bolch, H. Herold, MOSEL: Modeling, Specification, and Evaluation Language. *Technical Report* TR-i4-95-02, University of Erlangen-Nürnberg, 1996.

[13] G. Ciardo, R. Fricks, J.K. Muppala, K. Trivedi, Manual for the SPNP Package Version 4.0, Duke University, Durham, NC, USA, 1994.

[14] G. Bolch, S. Greiner, H. Jung, R. Zimmer, The Markov Analyzer MOSES, *Technical Report* TR-i4-10-94, University of Erlangen-Nürnberg, 1994.

[15] M. Krischnick, PEPSY-QNS, *Technical Report* TR-i4-18-94, University of Erlangen-Nürnberg, 1994.

[16] R.A. Sahner, K. Trivedi, A. Puliafito, Performance and Reliability Analysis of Computer Systems. An Example based approach using the SHARPE Software Package, Kluwer, 1996.

[17] R. Syski, Introduction to Congestion Theory in Telephone Systems, Second addition, North-Holland, 1986.

[18] B. Walke, ETSI Project: Broadband Radio Access Networks, *Broadband Communications'98, tutorial*, Stuttgart, Germany, April 1998.

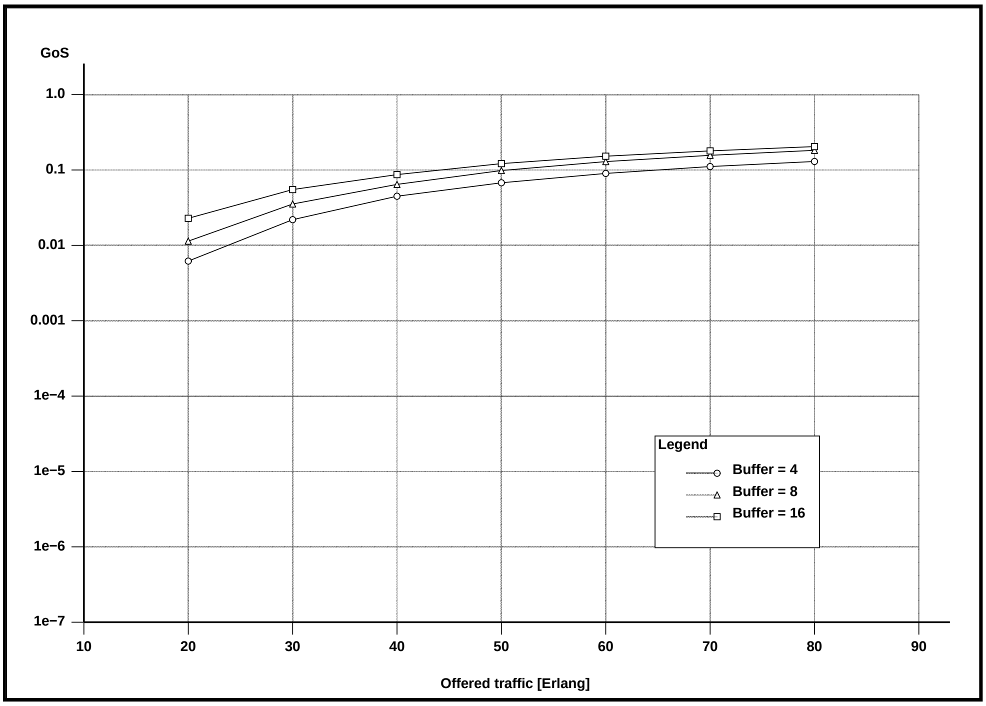

Figure 1: GoS of voice using channel reservation scheme

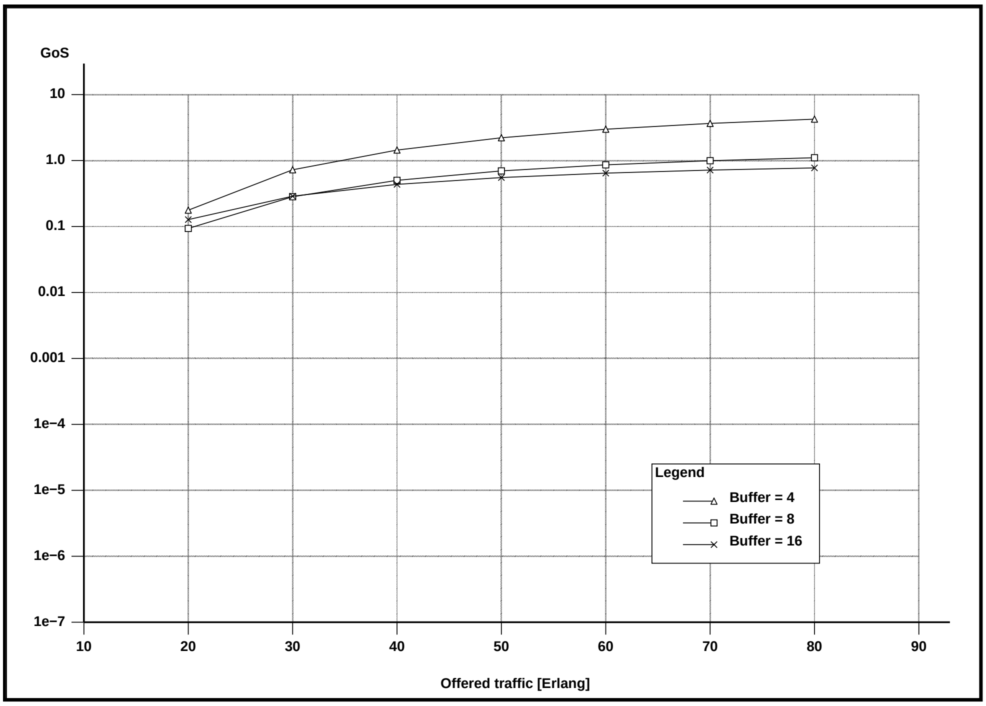

Figure 2: GoS of data using channel reservation scheme

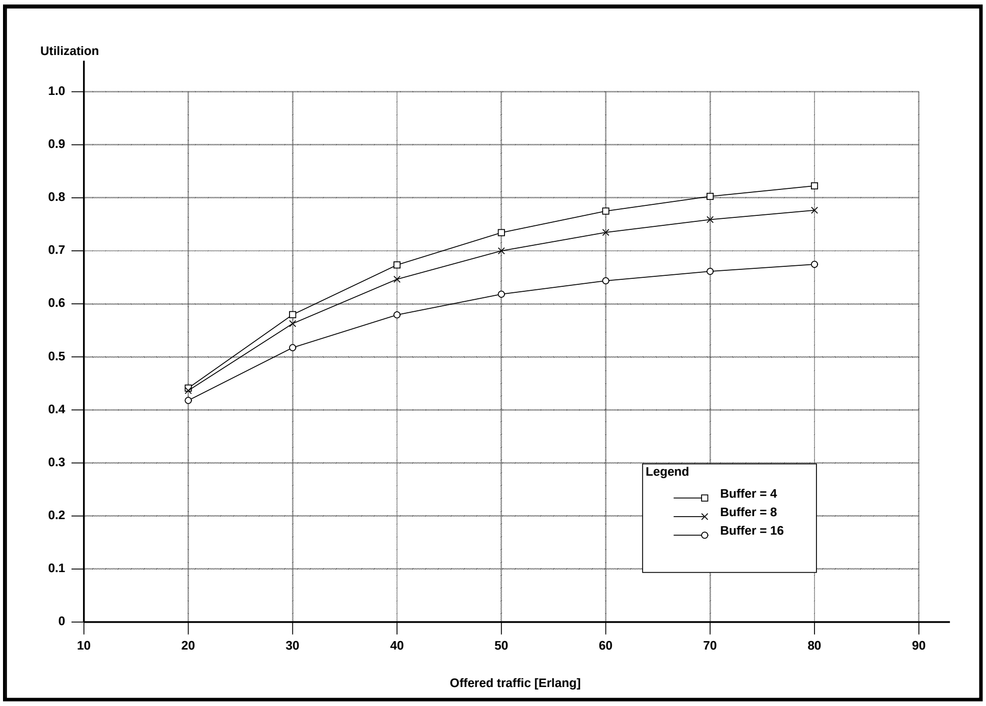

Figure 3: Cell utilization using channel reservation scheme

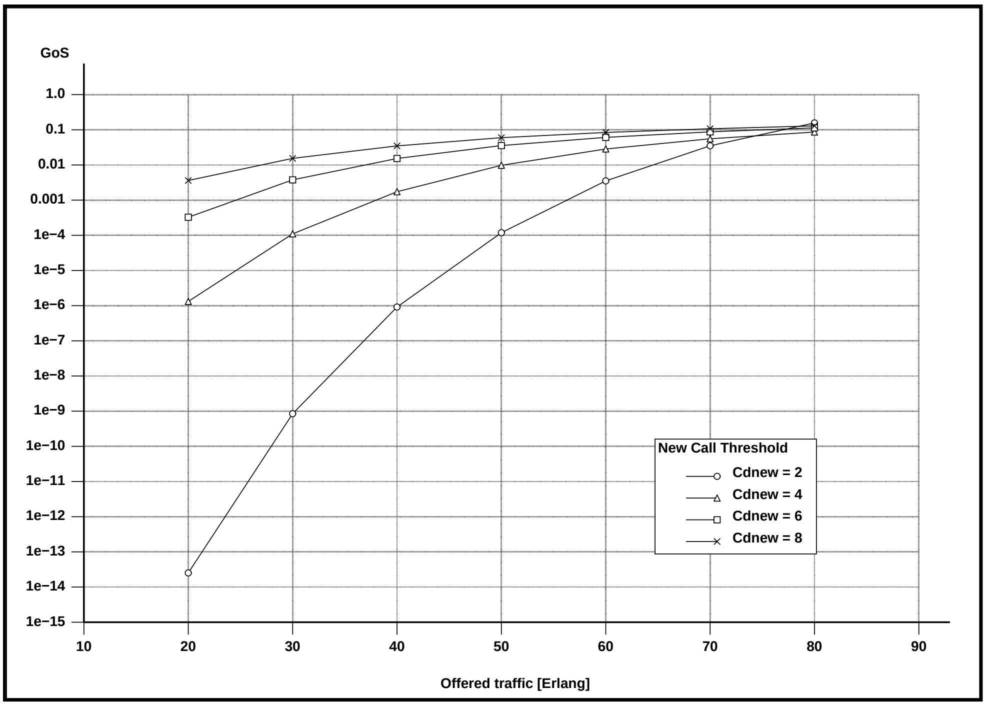

Figure 4: GoS of voice using scalable data connections

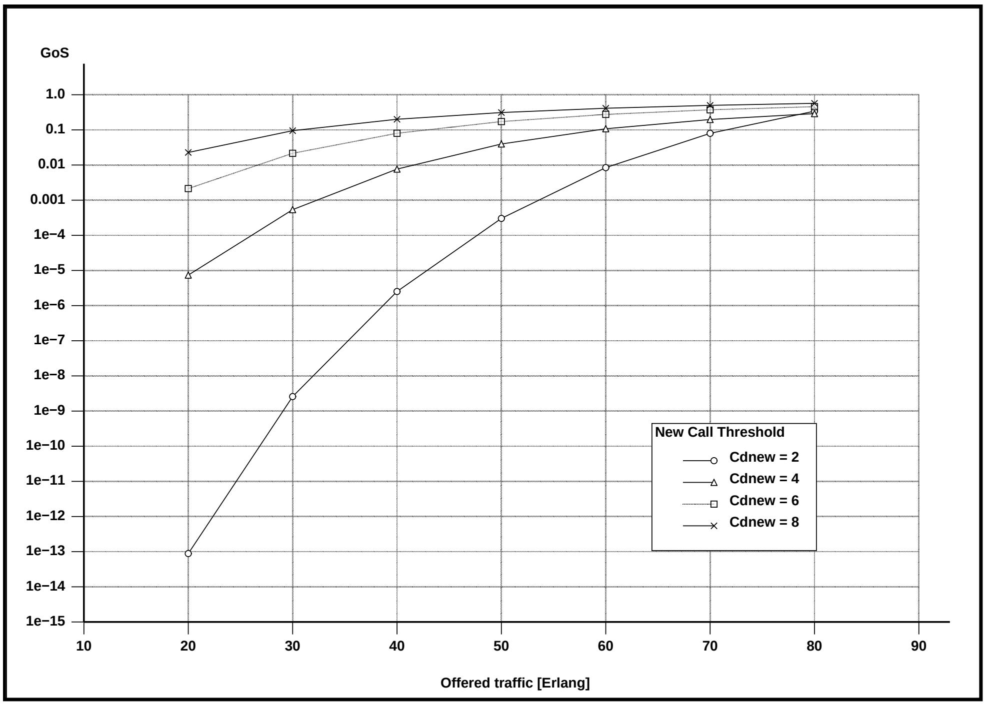

Figure 5: GoS of data using scalable data connections

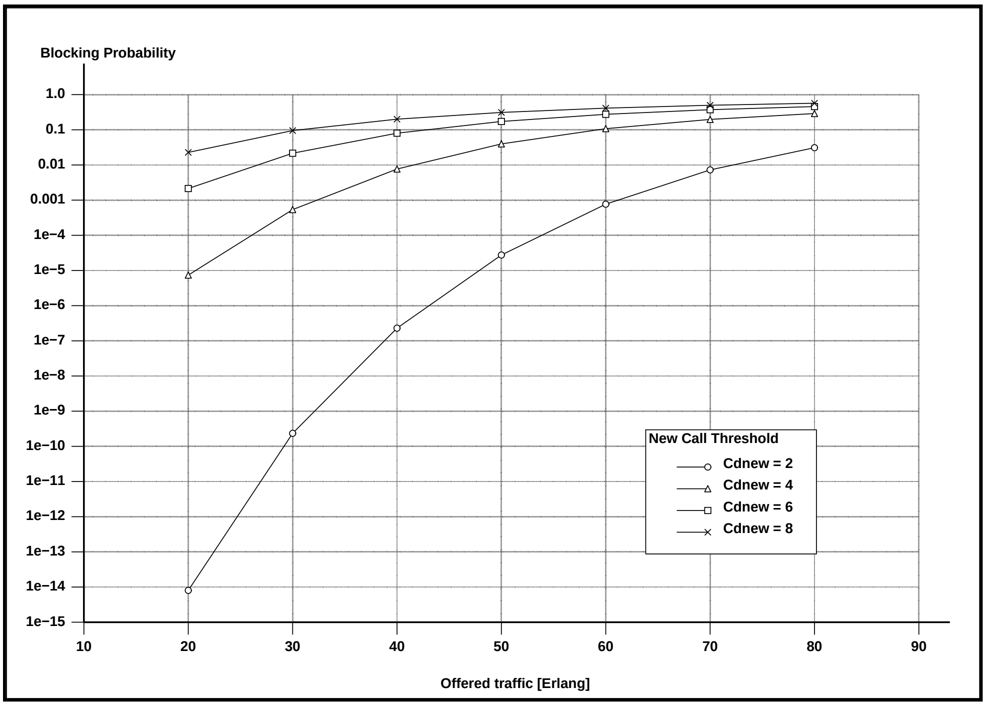

Figure 6: New call blocking probability of data using scalable data connections

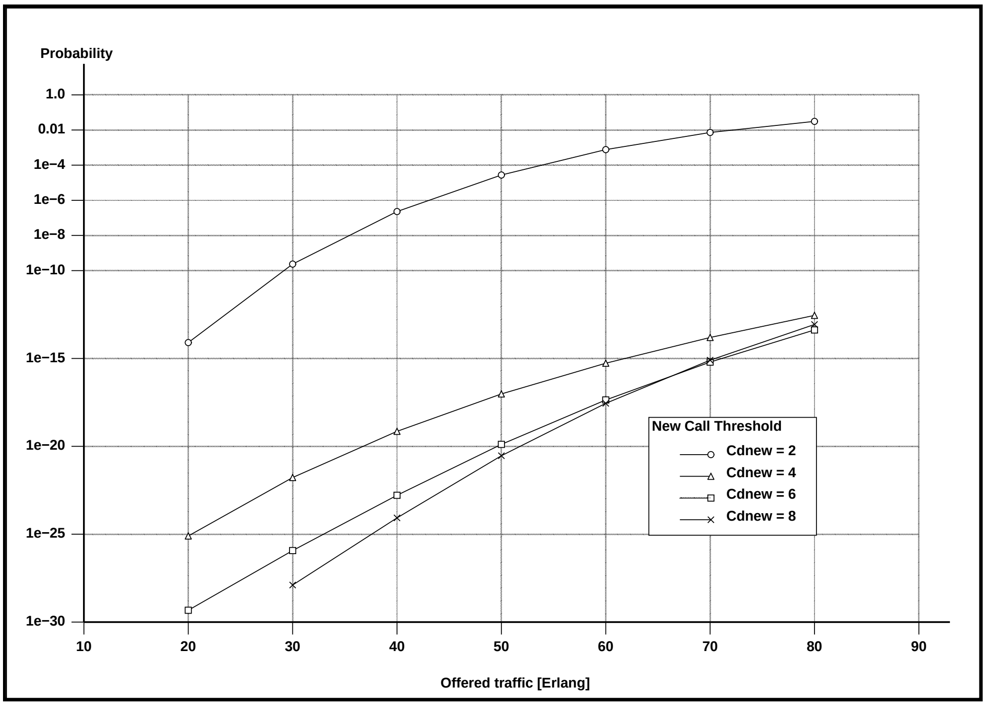

Figure 7: Handover failure probability of data using scalable data connections

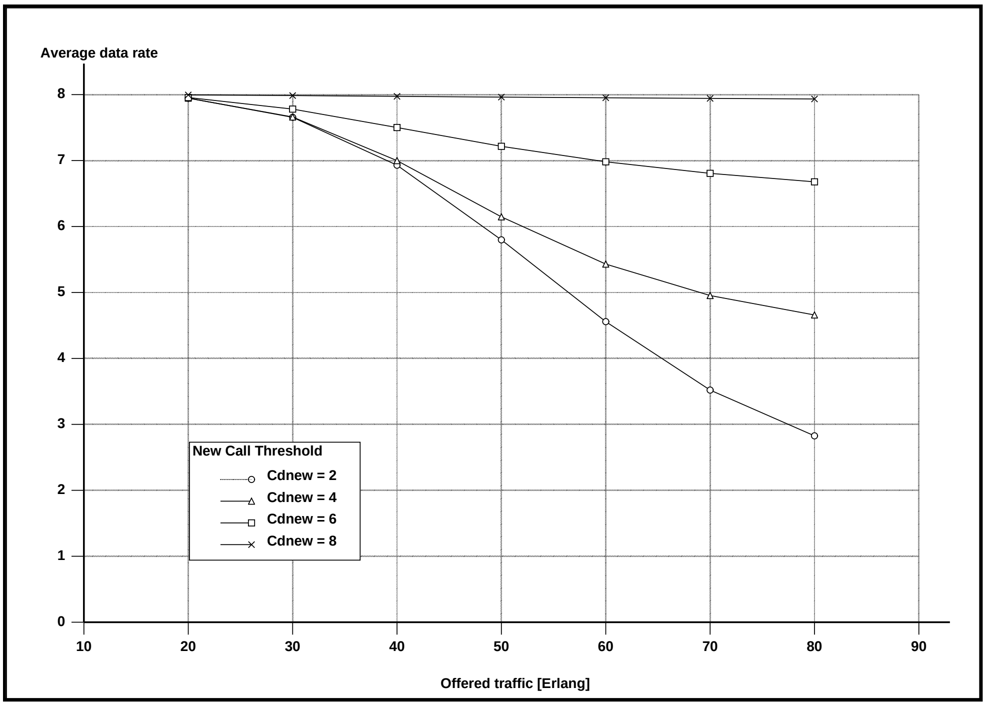

Figure 8: Average data transmission rate using scalable data connections

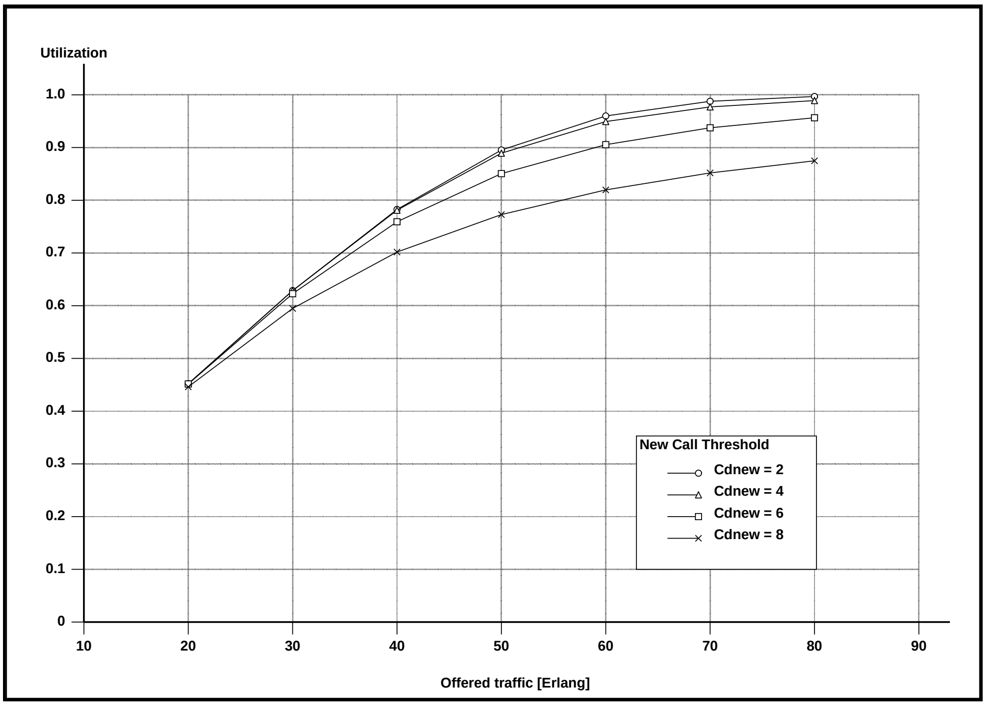

Figure 9: Cell utilization using scalable data connections

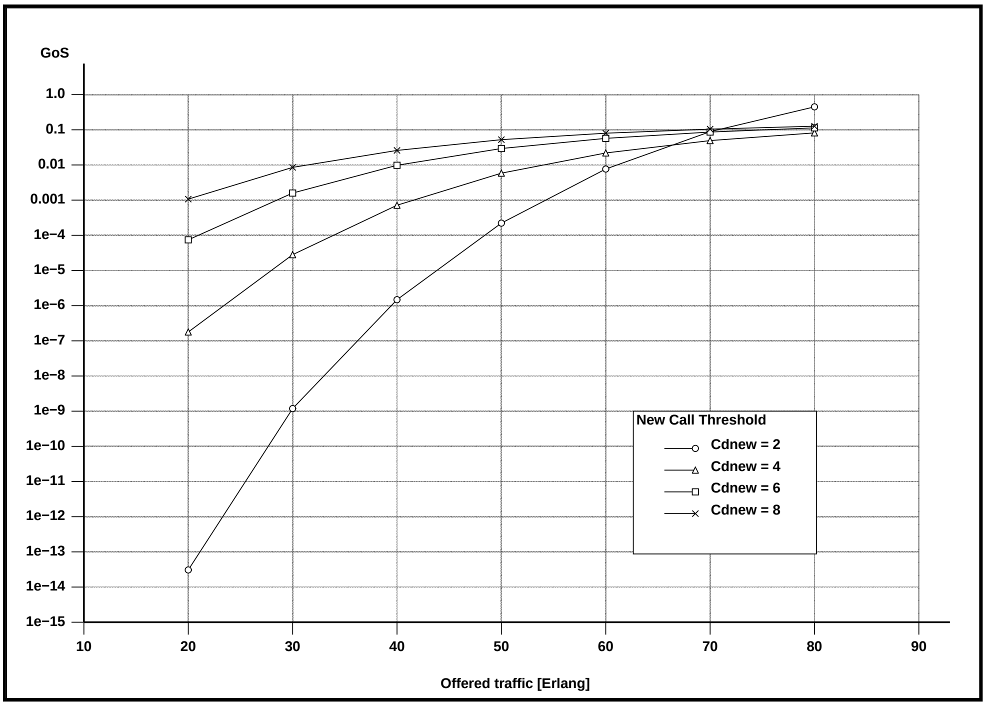

Figure 10: GoS of voice using scalable data connections with queuing

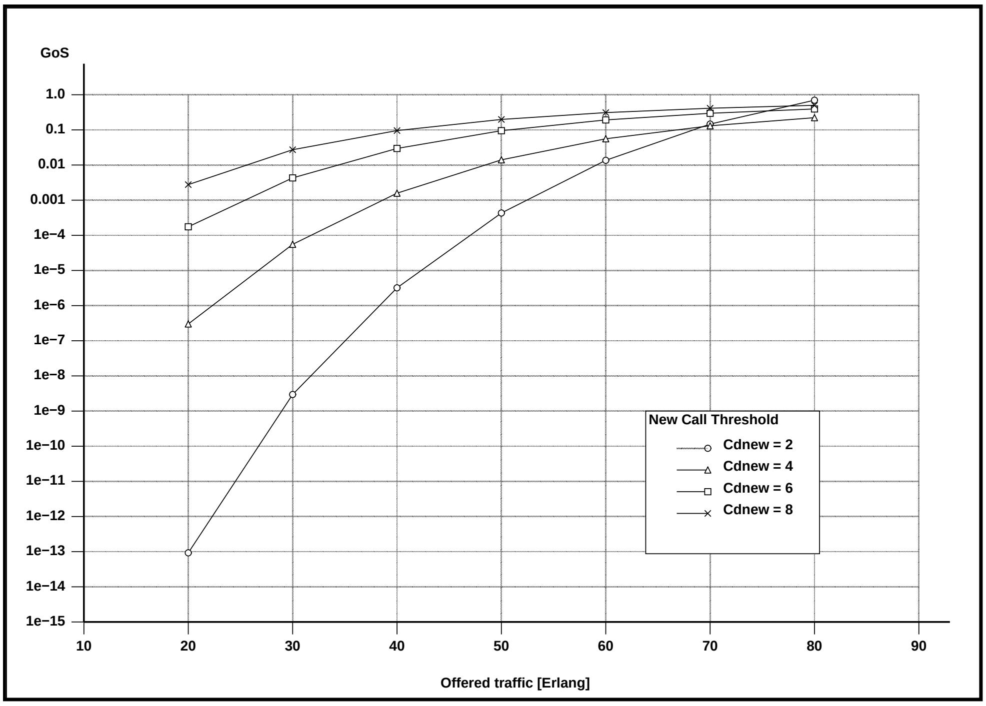

Figure 11: GoS of data using scalable data connections with queuing

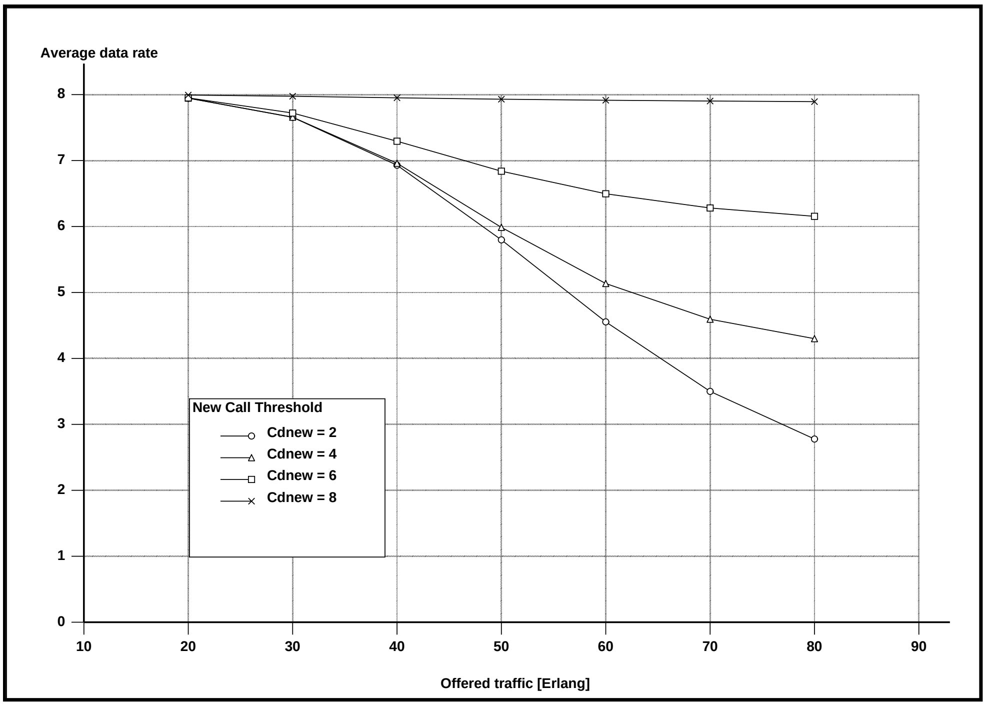

Figure 12: Average data transmission rate using scalable data connections with queuing

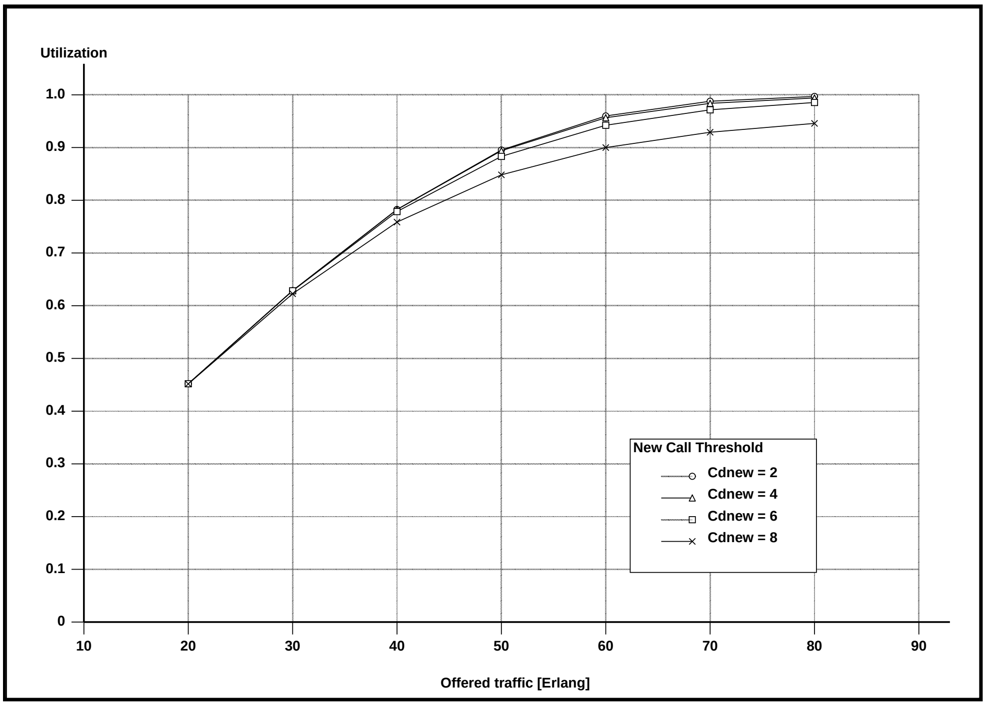

Figure 13: Cell utilization using scalable data connections with queuing