

# Mean Waiting Time Analysis in Finite Storage Queues for Wireless Cellular Networks

J. PYLARINOS<sup>1</sup>, S. LOUVROS<sup>1</sup>, K. IOANNOU<sup>1</sup>, A. IOANNOU<sup>3</sup> A.GARMPIS<sup>2</sup>  
and S.KOTSOPOULOS<sup>1</sup>

<sup>1</sup> Wireless Telecommunication Laboratory, Department of Electrical and Computer Engineering,  
University of Patras,  
Rion 26500, Patras  
GREECE

<sup>2</sup> Department of Applied Informatics in Management & Economy,  
Technological Educational Institution of Messolonghi,  
Messolonghi, 30100  
GREECE

<sup>3</sup> Hellenic Telecommunications Organization S.A.  
Marousi, 99 Kifissias Ave. Postal Code 151 24 Athens  
GREECE

*Abstract:* - Quality of service is obtained by introducing a threshold in the maximum waiting time of a handover call in the queue. In case the handover call mean service time at each queue position is found to be greater than this threshold, this call can not be serviced by this cell. This innovative idea introduces a more flexible traffic management since it enables a delay prediction in case of congestion.

*Key-Words:* -Wireless Networks, QoS, Mean Waiting Time

## 1 Introduction

In case of microcellular networks where frequent handovers is a fact, Quality of Service (QoS) may degenerate below an acceptable level due to brief service interruptions. As the frequency of these interruptions increases the perceived QoS is reduced. This issue places additional challenges on the design and dimensioning of microcellular wireless networks. Increasing the handover rate, the probability of an ongoing call to be dropped due to a lack of free channel is high. In order to preserve this blocking a queue is introduced to accommodate the handover calls. However the queue size should be kept small and the time spent in the queue should be below a certain threshold since handover attempt and waiting should not last for ever. This threshold should be decided after a mathematical justification over certain conditions from theory. In this paper a mathematical analysis is conducted in order to model the waiting time in the queue as function of mean service time, queue length and offered traffic load.

## 2 Mathematical Analysis of the Proposed Handover Procedure

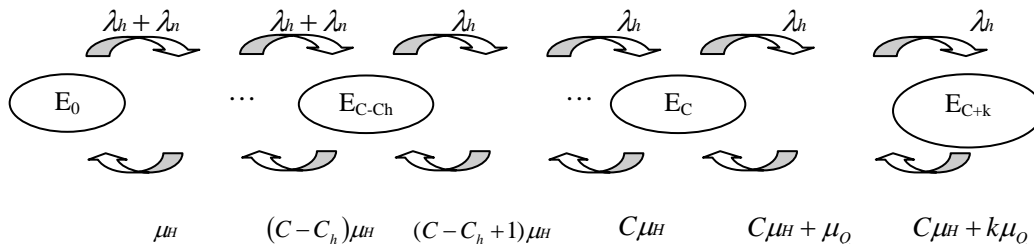
Priority can be given to handover attempts by assigning  $C_h$  guard channels exclusively for handover calls among the  $C$  channels in a cell. Both the new and the handover calls can share the remaining  $C - C_h$  channels. A handover attempt may be queued in a queue of  $k$  size if the state number in the cell is equal to  $C$  (All channels in the cell are busy). The channel holding time  $T_H$  in a cell is defined as the time duration between the instant that a channel is occupied by a call and the instant it is released by either completion of the call or a cell boundary crossing by a portable, whichever is less. It is proved [1] that the pdf of  $T_H$  can be approximated to a negative exponential distribution with mean  $\overline{T_H} = 1/\mu_H$

The service time  $T_Q$ , the time that an attempt remains queued at a position  $q$ , depends normally

on whether or not a channel becomes available as long as the mobile is still in the handover area. A handover attempt that joins the queue will be successful, if both of the following events occur before the mobile moves out of the handover area:

1. All of the attempts which joined the queue earlier than the given attempt have been disposed
2. A channel becomes available when the given attempt is at the first position in the queue.

Using the steady-state equations from Figure 1, we conclude:



**Figure 1:** State transition diagram of the queuing traffic model

$$P_j = \begin{cases} \frac{\rho^j}{j!} P_0, 0 \leq j \leq C - C_h \\ \frac{\rho^j a^{j-(C-C_h)}}{j!} P_0, C - C_h < j \leq C \\ \frac{\rho^j a^{j-(C-C_h)}}{C! \prod_{i=1}^{j-C} (C\mu_H + i\mu_Q)} P_0, C < j \leq C + k \end{cases} \quad (1)$$

$$P_0 = \frac{1}{\sum_{j=0}^{C-C_h} \frac{\rho^j}{j!} + \sum_{j=C-C_h+1}^C \frac{\rho^j a^{j-(C-C_h)}}{j!} + \sum_{j=C+1}^{C+k} \frac{\rho^j a^{j-(C-C_h)}}{C! \prod_{i=1}^{j-C} (C\mu_H + i\mu_Q)}}$$

where  $\lambda_n$  and  $\lambda_h$  are the new calls and the handover calls arrival rate respectively,  $\lambda = \lambda_n + \lambda_h$  is the total call arrival rate ( $\lambda_h = a\lambda$ ) and the offered load  $\rho$  in a communication system is defined as  $\rho = \frac{(\lambda_n + \lambda_h)}{\mu_H}$ .

On the basis of our consideration,  $T_Q$  should have an upper bound. In order to have an effective system, a handover call must not be allowed to

remain at a buffer position more than a maximum time threshold. Moreover, the queue size has to be limited because it is more realistic and practical than the infinite buffering. The maximum value of the mean service time  $\overline{T_Q} = 1/\mu_Q$  is obtained by the mean waiting time  $\overline{W}_h$  in the queue. Waiting time of a queued handover call is defined as the time of an arbitrarily selected handover call between the moment it is accepted and begin waiting in the queue to the moment it successfully accesses a free

channel. Given that the state of the system is when the call arrives and waits in the queue, we denote the waiting time by  $W_h(j)$  [3]:

$$W_h(j) = -\frac{1}{\mu_Q} \ln(1 - R_h(j)) \quad (2)$$

where  $R_h(j)$  is the dropping probability of an arbitrary selected handover call, given that the system state is  $j = C + q$  ( $0 \leq q \leq k - 1$ ) just at the instant the call is accepted by the system and waits in the queue. Consequently, the average waiting time of a handover call, denoted by  $\overline{W}_h$ , can be obtained by:

$$\overline{W}_h = \frac{\sum_{j=C}^{C+k-1} P_j W_h(j)}{\sum_{j=C}^{C+k-1} P_j} \quad (3)$$

As we can easily conclude,  $\overline{W}_h$  is, among other, a function of the mean service time  $\overline{T}_Q = 1/\mu_Q$ . Thus, setting an upper bound  $(\overline{W}_h)_{MAX}$  for the waiting time in the queue, we can solve for  $\overline{T}_Q$  and find the corresponding maximum allowable mean service time at every position  $q$ . Of course, this solution of  $\overline{T}_Q$  should be inside the interval  $[0, +\infty)$ .

As we already mentioned, the blocked handover calls join a queue. A handover attempt that enters the queue at the position  $q$  will be successful, if it manages to reach the first position of the queue and get a channel before its mean service time becomes greater than the calculated from equation (3) value. In order to derive the probability of a handover failure in the queue  $R_h(C + q)$ , we assume that:

$$1 - R_h(C + q) = \left[ \prod_{i=0}^q P(i/i+1) \right] \times \Pr(\text{call\_remains\_in\_queue}) \quad (4)$$

The probability of transition from position  $i + 1$  to  $i$  is denoted by  $P(i/i+1)$  in equation (4) and is contributed by two probabilities [2]:

- I. The remaining channel holding time of any of the  $C$  calls in progress is smaller than each of the following:
  - The remaining channel holding time of any of the other  $(C - 1)$  calls in progress.
  - The service time of any of the  $i$  waiting handover calls.

- The service time of the waiting handover call of interest.
- II. The remaining service time of any of the  $i$  handover calls waiting in the queue is smaller than each of the following:
    - The channel holding time of any of the  $C$  calls in progress,
    - The service time of any of the other  $(i - 1)$  waiting handover calls.
    - The service time of the waiting handover call of interest.

Thus, the transition probability can be obtained by:

$$P(i/i+1) = \frac{C\mu_H + i\mu_Q}{C\mu_H + (i+1)\mu_Q} \quad (5)$$

The second term in equation (4) is a logical condition that can have only two values. If the mean service time at this position is smaller than or equal to the maximum mean service time threshold, this term is set to "1". Otherwise, it is set to "0". Thus:

$$\Pr(\text{call\_remains\_in\_queue}) = \begin{cases} 1 & , \overline{T}_Q \leq (\overline{T}_Q)_{MAX} \\ 0 & , \text{otherwise} \end{cases} \quad (6)$$

Substituting equations (5) and (6) into (4),  $R_h(C + q)$  can be calculated and then the mean waiting time of a handover call in queue  $\overline{W}_h$  can be estimated as a function of  $\overline{T}_Q$ . Setting a maximum value for  $\overline{W}_h$  (after certain logical constraints),  $\overline{T}_Q$  can be calculated as a reverse function of  $\overline{W}_h$ .

## 5. Simulation Results

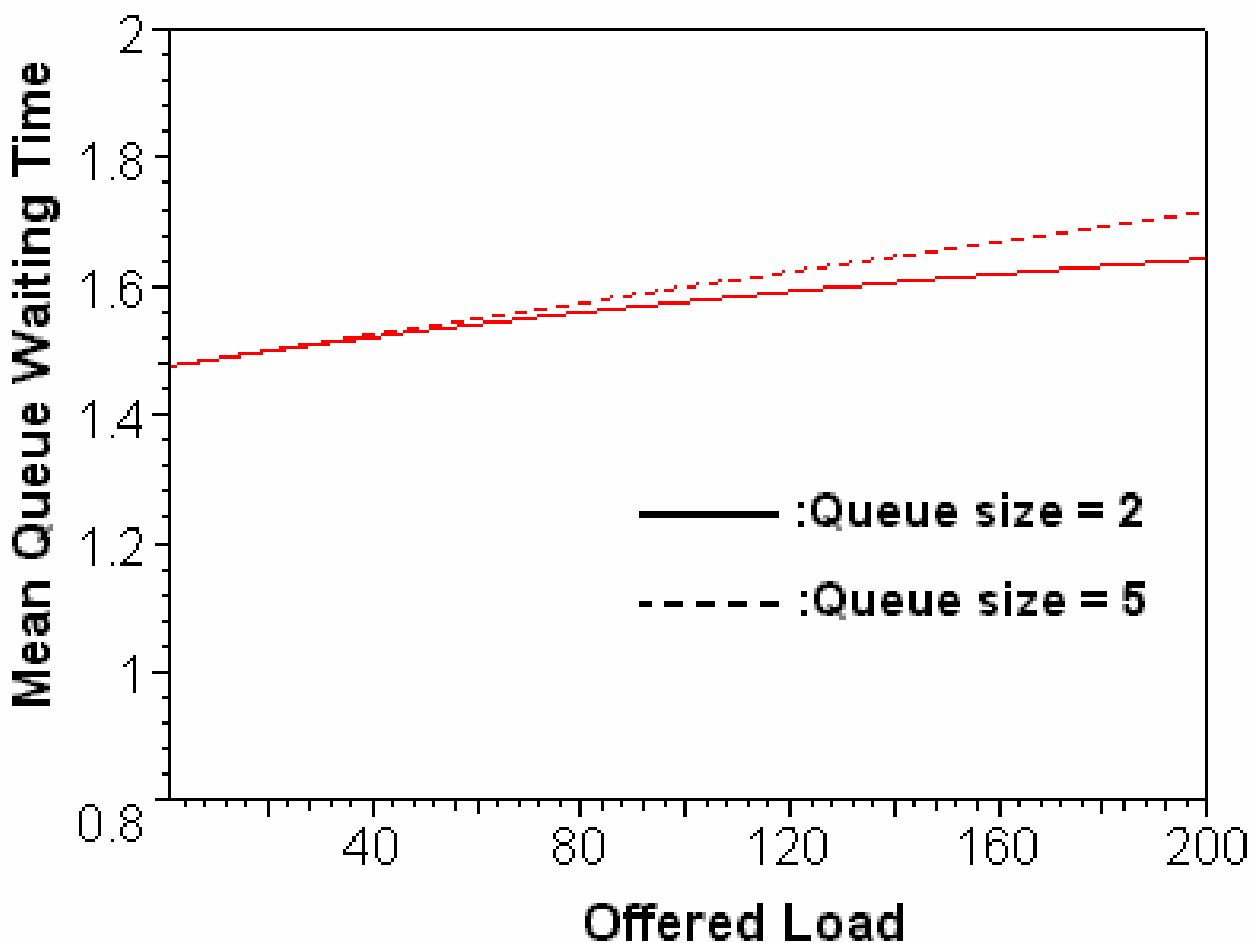
The following assumptions have been made during simulation:

- The total number of available channels in the cell is  $C = 20$ .
- $C_h = 2$  channels are reserved only for handover calls.
- The handover call to total call is  $a \cong \frac{1}{3}$ . This value is based on statistical measurements in real cellular systems.
- The mean channel holding time is  $\mu_H \approx \frac{1}{85} \text{ sec}^{-1}$ , according to mean value analysis of real cellular network data.

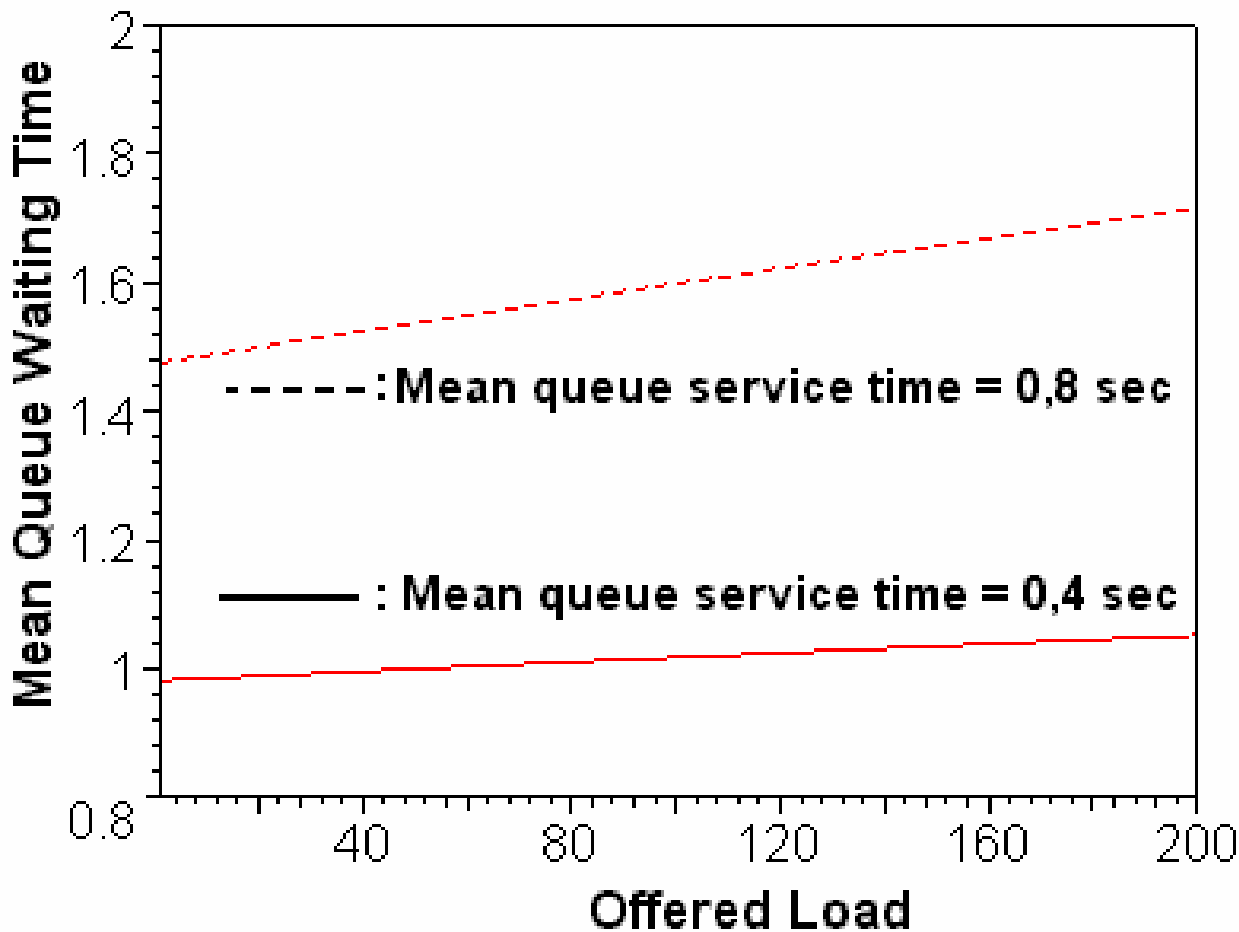
The relation between service and waiting time is presented in figures 2 and 3. In figure 2 we notice that, as the offered load increases, the increment in  $\overline{W}_h$  is more significant for a queue size of  $k = 5$  than for a size of  $k = 2$ . This is because we accommodate more subscribers in a larger queue and the service is delayed. However the more the handover calls to accommodate the more likely is the probability of forced terminating a handover call due to exceeding the maximum service time. In figure 3, the mean waiting time is plotted as a function of offered load with parametric values of specific service time at each queue position ( $k=2$  or  $k=5$ ). It is obvious that the mean service time of 0.4 sec is too low (resulting from a very low threshold  $(\overline{W}_h)_{MAX}$ ) and as a result the mean waiting time is increasing very slowly (around the neighborhood value of 1 sec) with offered load. However a small increment of the mean service time up to 0.8 sec

results in a large increment of the mean waiting time, showing how sensitive is the mean waiting time in small perturbations of service time. This has to be taken into consideration in case of dimensioning the network.

The calculation of the maximum service time in queue is a very useful parameter to calculate the handover blocking probability. Handover blocking probability is a major criterion of Qos of cellular networks. Hence modeling the handover blocking probability as a function of maximum service time gives a powerful tool to the designer to calculate the capacity of the system and evaluate the system response. Moreover if we consider also the queue length  $k$  then the handover blocking probability will be a function of three variables (mean service time, queue length and offered load), providing the designer with a more powerful tool to design the network resources, taking care also of special compromises among the parameters.



**Figure 2:** Mean Queue Waiting Time vs. Offered Load for different queue sizes



**Figure 3:** Mean Queue Waiting Time vs. Offered Load for different mean queue service times.

reneging/dropping”, IEEE/ACM Trans. on networking, vol. 2, No 2, April 1994.

## References

- [1] Hong D. and Rappaport S.S.: “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non prioritized handover procedures” IEEE Trans. on Vehicular Technology, vol. VT-35, pp.77-91, 1985.
- [2] Panoutsopoulos I., Kotsopoulos S., Ioannou C. and Louvros S.: “A new proposed priority technique to optimize the handover procedure in personal communication systems” IEE Electronics Letters, vol. 36, No 7, pp.669-670, March 2000.
- [3] Chang Chung-Ju, Su Tlan-Tsair and Chiang Yueh-Yiing: “Analysis of a cutoff priority cellular radio system with finite queuing and