Priority Handover Schemes in Wireless Mobile Networks

Dimitar Radev, Svetla Radeva, Ivan Kurtev and Dragan Stankovski

Abstract — **Priority handover schemes in wireless mobile network are under consideration. Models of single traffic system, multiple traffic system, multiple traffic system with total priority and traffic system with State Dependent Processor Sharing are considered. Different schemes for handover handling with non-preemptive priorities are compared. The possibility of their modeling with Markov chains is discussed. Two new priority schemes for 4G Next Generation Broadband Wireless Networks are suggested.**

Keywords — **Wireless Mobile Networks, Handover, Queuing Systems, Traffic Models.**

I. Introduction

THE mobility is one of the most important features of a wireless cellular communication system. Usually, continuous service is achieved by supporting handover from one cell to another [1]. Handover is the process of changing the channel (frequency, time slot, spreading code, or their combination of them) associated with the current connection while a call is in progress [2]. It is often initiated either by crossing a cell boundary or by a deterioration in quality of the signal in the current channel [3]. Handover is defined as a change of radio channel applied by mobile terminal [4]. The new channel may be within the same cell (intra-cell handover) or in a different cell (inter-cell handover) [5].

II. Handover Principle and Concepts

By definition, handover means transfer of r connection from one radio channel to another. This definition is made before UMTS appearance and afterward it is spread with soft and softer handover [6]. The main purpose of handover is to maintain an ongoing call [7]. This is necessary as the r might be moving, sometimes with high speed and this can lead to his dropping when the r changes to another cell or area. Also it is possible that the number of r changes while the call for a r is ongoing and for the call to continue the network needs to change the frequency of an ongoing call. Finally, the r might leave the area of coverage of UMTS network and might be handed over a GSM/GPRS network.

On Fig.1 is shown the handover principle, where the mobile station MS is moving from one base station to another. The mean signal strength of the first base station decreases as the mobile station moves away from it. From the other side, the mean signal strength of the second base station increases as the mobile station approaches it.

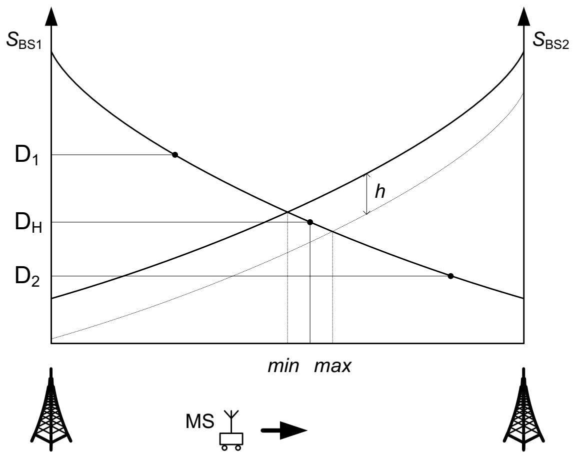

Fig.1. Potential handover between two Base Stations.

This method allows a mobile station to hand off only if the current signal is sufficiently weak less than the threshold and the signal from second base station is stronger than from the first one. The effect of the thresholds depends on its relative value as compared to the signal strengths of both base stations. If the threshold is lower than the value D_1 (Fig.1) the handover occurs at position *min*. If the threshold is lower than D_H than the mobile station would delay handover until the current signal level crosses the threshold at the position between *min* and *max*. In the case of D_2 the delay may be so long and may result in a dropped call. When the mobile station approaches the handover area a r call is served only if the new base station has stronger signal, by a hysteresis margin *h* than the current one.

III. Priority Handover in Single Traffic Systems

On Fig.2 is shown a priority queuing model of handover in single traffic system, where for it analysis is necessary more detailed to consider the handover procedure. When

This work is a part of the scientific project DSAB 02/5 of the Bulgarian National Science Fund. The authors are thankful for the financial support.

Dimitar Radev is with the University of Rousse, Professor at the Dep. of Communication Techniques and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, (phone: +359- 082-888 673; e-mail: dradev@abv.bg).

Svetla Radeva is Assoc. Prof. at the Dep. of Wireless Communications and Broadcasting at the College of Telecommunications and Post, 1 Akad. Mladenov Str. 1700 Sofia, Bulgaria, (e-mail: svetla_ktp@abv.bg).

Ivan Kurten is Assoc. Prof. at the Dep. of Wireless Communications and Broadcasting at College of Telecommunications and Post, 1 Akkad. Mladenov Str. 1700 Sofia, Bulgaria, (e-mail: i.kurtev@hctp.acad.bg).

Corresponding author is Dragon Stankovski. He is PhD student at the University of Rousse, at the Department of Communication Techniques and Technologies, 8 Studentska Str., 7017 Rousse, Bulgaria, (phone: +359-878 252 488, e-mail: draganstankovski@gmail.com).

the mobile station moves farer to the base station the power of received signal decreases and when it becomes smaller than the threshold value, the handover procedure starts.

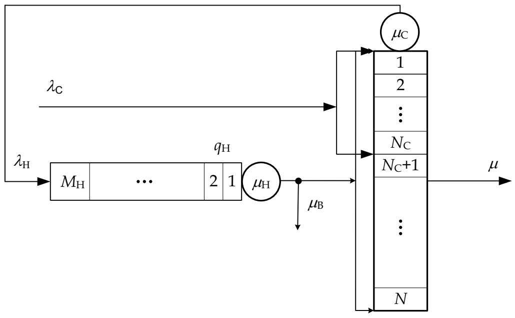

Fig.2. Handover queuing model in single traffic system

Handover area is defined as the zone on which the mean signal strength received from the mobile station is between the threshold value of the handover and the threshold value of the receiver. When the base station finds all channels of the target cell busy, then the handover request is put into the queue. If a channel become free while the handover queue is not empty than the channel is given to the first received into the queue request. If the strength of the received signal from the base station is under the receiver threshold value before the mobile station receives a channel into the target cell than the call is dropped out. The calls into the queue are served with FIFO (First In First Out) priority [8] and the queue length into the base station is fixed.

The service time for requests into channels is T_C and it is supposed that this time is exponential distributed with mean value $E[T_C] = 1/\mu_C$. If the average number of busy channels with new calls is $E[C_C]$, than from the Little's rule is received the relationship between the new calls and their service as:

$$\lambda_C = E[C_C]\mu_C , \tag{1}$$

where λ_C is arrival rate and μ_C is service rate for new coming calls.

The dwell time of the mobile station into the handover area depends on such system parameters as power and direction of movement and cell size. This dwell time of the mobile station into the handover area is T_H. For simplicity it is supposed that this dwell time is exponential distributed with mean value $E[T_H]=1/\mu_H$. If the number of busy channels for handover at the mobile station is $C_H=N-C_C$, than their average number can be determined as $E[C_H]$. Than it is received that

$$\lambda_H = \left(E[C_H] + M_{qH}\right)\mu_H , \tag{2}$$

where M_{qH} is the current number of busy cells in handover queue q_H.

The basic system parameters and blocking probabilities for calls (new and in handover) can be received from the transition diagram of one-dimensional Markov chain, defined as a sum of channels which are appliedd into the cell and number of requests into the handover queue coming.

IV. Priority Handover in Multiple Traffic Systems

A system with many cells each with N channels is under consideration. The base station has two handover queues – queue for non-real-time traffic q_N (for example, data requests) and queue for real-time traffic q_R (for example, voice calls). The new calls come into the base station with arrival rate λ_C, assigning N_C channels and service rate μ_C. When all of these channels are assigned, the new coming calls are sent to handover which distribute them into two queues. The data calls with arrival rate λ_{HN} are assigned into the handover queue q_N and are served with rate μ_{HN}. The voice calls with arrival rate λ_{HR} are assigned into the handover queue q_R with service rate μ_{HR}.

In this model there are no blocking of data calls in handover beca each request which aren't served returns back into the queue q_N. Only voice calls may be blocked in handover queue q_R with maximal length M_R when the queue is full and all N channels of the base station are occupied.

This scheme has three levels of priority at assigning channels of base station (Fig. 3). With lowest priority are new coming calls, which arrive directly at up to N_C channels with arrival rate λ_C. When channels from 1 to N_C are busy than the new calls goes to the handover. The next priority level is given to the data requests which are into the handover queue q_N. They can occupy up to N_N channels. The rest N-N_N channels are with the highest priority and can be occupied only by voice calls which are in the handover queue q_R.

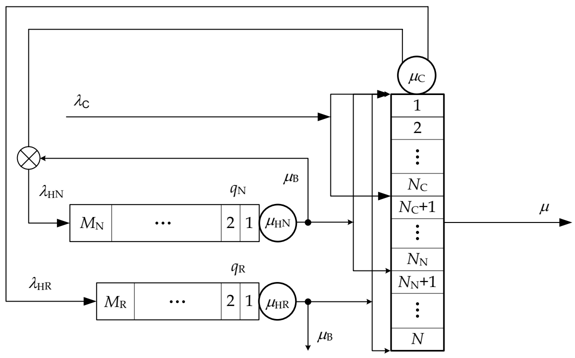

Fig.3. Queuing model in multiple traffic system

It is supposed that the holding time T_C of requests into the channels is exponential distributed and service duration of non-real-time requests in handover T_{HN} and service duration of real-time requests in handover T_{HR}, like voice calls, are as well exponential distributed e.g. their mean values are respectively $E[T_C]=1/\mu_C$, $E[T_{HN}]=1/\mu_{HN}$ and $E[T_{HR}]=1/\mu_{HR}$. The average number of occupied channels for new coming calls $E[C_C]$ can be calculated according Little's rule as is shown in (1). By analogy from (1) and (2) can be determined the average number of busy channels of voice calls in handover $E[C_{HR}]$ and average number of busy channels $E[C_{HN}]$ of data requests in handover as

$$\begin{aligned} \lambda_{HR} &= E[C_{HR}]\mu_C \\ \lambda_{HN} &= \left(E[C_{HN}] + M_{qN}\right)\mu_C , \end{aligned} \tag{3}$$

where M_{qN} is the current number of busy cells in the handover queue for non-real-time traffic q_N.

This priority scheme is considered as two-dimensional Markov chain, where each state of the system is presented with a couple nonnegative integer values (i, j). On axis i are depicted possible states of base station's N and possible states of cells in handover queue for real-time traffic q_R with length M_R. On axis j are depicted possible states of calls in handover queue for real-time traffic q_N with length M_N.

V. Handover in Multiple Traffic Systems with Total Priority

Nevertheless of existing of certain differences in 4G traffic beca of completely IP transfer of calls the handover model keeps it multimedia character. This supposes existence of two queues in handover for real-time traffic with queue length M_R and for non-real-time traffic with queue length M_N, as is shown on Fig. 4. In suggested model the channels of mobile station which are N are assigned with the help of 4 priority levels.

The threshold values for these levels are defined with the variables:

- N_{CN} is the maximal number of channels which can be applied for new coming data requests;
- N_{CR} is the maximal number of channels which can be d for new voice calls;
- N_N is the maximal number of channels which can be d for data requests in handover.

This scheme supposes that separating of real-time traffic and non-real-time traffic is realized out of cell's handover at serving the channels. This is the reason that serving of new calls is realized with different service rate - μ_{CN} for non-real-time traffic and μ_{CR} for real-time traffic.

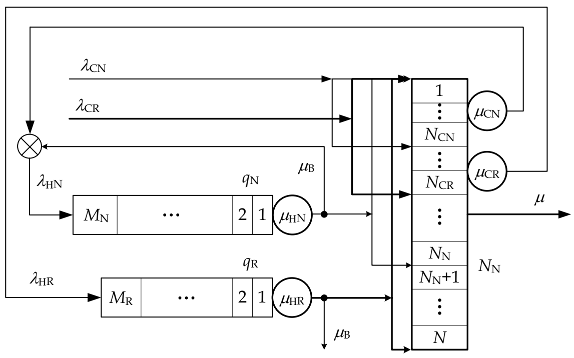

Fig.4. Queuing model with total priority

The flows of new coming calls in base station with arrival rate λ_C and of handover requests with arrival rate λ_H are determined as:

$$\lambda_C = \lambda_{CN} + \lambda_{CR}, \qquad \lambda_H = \lambda_{HN} + \lambda_{HR}, \tag{4}$$

The lowest priority has new coming data calls with arrival rate λ_{CN}. When all channels from 1 to N_{CN} are busy than these requests are going to the handover where they are separated for the queue q_N with arrival rate λ_{HN} and are served with service rate μ_{HN}.

With higher priority are new coming voice calls which occupy channels of the base station from 1 to N_{CR} ($N_{CR} > N_{CN}$) with arrival rate λ_{CR}. When all of these channels are busy they are going to the handover queue q_R with arrival rate λ_{HR} and are served with service rate μ_{HR}.

The next priority level is given to the requests from the handover queue q_N which occupy channels of the base station from 1 to N_N. The voice calls which are into the handover queue q_R has the highest priority and can occupy each of the all N channels of base station.

It is supposed exponential distribution for holding times of data requests T_{CN} and voice calls T_{CR} into the cell's channels and exponential distribution for holding service times of data requests T_{HN} into handover and holding service times of voice calls into handover T_{HR}. The mean values of these holding times can be presented as $E[T_{CN}]=1/\mu_{CN}$ and respectively $E[T_{CR}]=1/\mu_C$, $E[T_{HN}]=1/\mu_{HN}$ and $E[T_{HR}]=1/\mu_{HR}$.

Analogical as at previous schemes can be determined the average number of busy channels with: new coming data requests $E[C_{CN}]$; new coming voice calls $E[C_{CR}]$; handover data requests $E[C_{HN}]$; handover voice calls $E[C_{HR}]$, where:

$$\begin{cases} \lambda_{CN} = E[C_{CN}]\mu_{CN} \\ \lambda_{CR} = E[C_{CR}]\mu_{CR} \\ \lambda_{HR} = E[C_{HR}](\mu_{CN} + \mu_{CR}) \\ \lambda_{HN} = \{E[C_{HN}] + M_{qN}\}(\mu_{CN} + \mu_{CR}) \end{cases} \tag{5}$$

The parameters of this priority scheme can be determined from the balance equations of steady-state solution of the system, described with three-dimensional Markov Chain. Each steady-state solution is presented with three nonnegative integer values (i, j, k).

The states on axis i present the possible coming of real-time requests into channels and handover queues, that's why they are from 0 to $N+M_R$. On axis j are presented all channels at the base station (from 0 to N), and on axis k are depicted possible states of the handover queue for non-real-time traffic which are from 0 to M_N.

VI. SDPS Handover with Total Priority

Let suppose that the handover service times are exponentially distributed and the system can be presented as three dimensional Markov chain. This leads to many shortcomings and in certain cases it is impossible to determine effective performance characteristics and working parameters of the handover.

The considered up to now schemes show us that modeling and analysis of the handover corresponds with computer and communication systems, where lately is widely d processor sharing systems [9]. This is the reason that in this investigation is suggesting a new model of handover handling where is realized processor sharing of traffic flows. The model is considered as M/GI/ system with processor sharing of service in dependence of the states - SDPS (State Dependent Processor Sharing).

With accordance to the previous sections here is suggesting a new scheme for handover handling with state dependant processor scaring. In this scheme are d r parallel queues and servers for handover serving, where $r>2$. The local exponential handover service time into the model is calculated via the sojourn time of a request in handover T_H of a M/GI/2-PS traffic system, e.g. system with two queues with common processor sharing of service times.

At the broadband mobile wireless networks the sojourn time of requests in handover T_H increases significant with increasing of offered load ρ. One of the ways for removing this is the adequate increasing of the number of flows into the system. The necessity of this is proved more detailed in [10].

As id shown on Fig. 5 all queues has the same maximal length M, which is achieved with equal for all mean deterministic service time $m_H=d_H$, e.g. the service in handover is presented as a M/G/r-SPDS system.

The number of queues r in this system is determined from the condition of steady state solution of the system and depends of her offered load ρ, where

$$\rho = \lambda_H . d_H = (\lambda_R + \lambda_{NR}).d_H < r. \tag{6}$$

In this scheme is removed the necessity of different queue length of the queues for teal-time traffic M_R and queues for traffic with delays (data traffic) M_N. The new coming call's traffic is not separated and this decreases the service time into the channels μ_C. Moreover, with implementation of FIFO rules for queues serving we can assign different priorities into the scheme. When there are no free channels at the base station the calls wait for service at the handovers queues. When handover queue q_1 becomes full, the calls arrive into the handover queue q_2, afterward arrive into the handover queue q_3, etc. up to the handover queue q_r. The waiting service at the handover queues requests are deleted from the handover queue when they passed through the handover area before their arrival into a channel or if theirs communication is completed before call requests passed through the handover area.

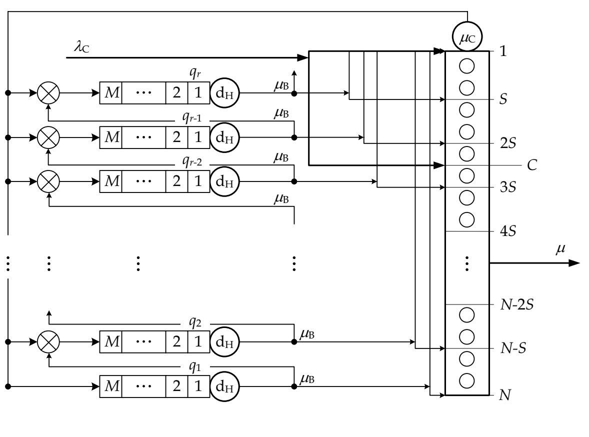

Fig.5. Total priority model with SDPS handover

When there are no free channels at the mobile cell, the new calls are served according to FIFO rules at the handover q_1. If there are free channels, according to FIFO rules are served the requests from handover queue q_1. If there are free more than (N-S) channels at the mobile cell etc. If there are more than $3S$ free channels, than are served the requests from handover queue $q_{(r-2)}$. When the number of free channels is more than $2S$, than the requests from handover queue $q_{(r-1)}$ are served according to FIFO rules. If there are more than S free channels, than are served the requests from handover queue r according to FIFO rules.

In this way can be realized priorities for real-time traffic flows and non-real-time traffic flows (data requests). From the relationship between offered load, new data requests and new voice calls can be determined the exact number of handover queues for real-time traffic and for non-real-time traffic.

VII. Conclusion

In this investigation is analyzed call's handling in handovers of GSM 2G, UMTS 3G and 4G Next Generation Broadband Wireless Networks.

Different schemes with non-preemptive priorities are discussed and compared. The possibilities for modeling of these schemes with one-dimensional, two-dimensional and three-dimensional Markov chains are presented.

Two new priority schemes for call's handling in handovers of 4G NGWN are suggested – scheme with total priorities and State Dependent Processor Sharing scheme. The advantages at implementation of these schemes are presented.

References

[1] R. Naja, "Mobility Handling and Resource Allocation in Wireless Multiservice Networks", ENST Bretagne, France, Ph. D. Thesis, 2003.

[2] 3GPP TS 23.009: *Handover Procedures*.

[3] Y. Zhang, "Handoff Performance in Wireless Mobile Networks with Unreliable Fading Channel," *IEEE Transactions on Mobile Computing*, vol. 9, No. 2, 2010, pp. 188-200.

[4] USA Patent US007197307B2, "Hard Handover Method and Controller", 2007.

[5] Q-A. Zeng, D.P. Agrawal „Handoff in wireless mobile networks", Handbook of wireless networks and mobile computing, *John Wiley & Sons*, New York, NY, 2002.

[6] D. Wong, T.J. Lim, „Soft Handoffs in CDMA Mobile Systems", *IEEE Personal Communications*, pp.6-17, 1997.

[7] G. Tamea , T. Inzerilli , P. Rea , R. Cusani, „Vertical handover among broadcast networks, Proceedings of the 6th International Conference on Symposium on Wireless Communication Systems", Siena, Italy, pp. 418-422, 2009.

[8] P.J. Gutierrez, "Packet Scheduling and Quality of Service in HSDPA", Department of Communication Technology, Institute of Electronic Systems, Aalborg University, Ph. D. Thesis, 2003.

[9] A. Brandt, M. Brandt, "Approximations for the second moments of sojourn times in M/GI systems under state-dependent processor sharing" ZIB-Report 09-25, 2009.

[10] D. Radev, I. Kurtev, D. Stankovski, S. Syarova, „A Handover Scheme for Broadband Wireless Mobile Networks", *International Symposium on Electronics and Telecommunications*, ETC 2010, Timisoara, Romania, 2010, (in press).