Technical QoS parameters for Cellular Systems

Samya Bhattacharya and Hari Mohan Gupta
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi, INDIA 110016
E-mail: {samya, hmgupta}@ee.iitd.ac.in

Abstract- In this study, we discuss some technical QoS parameters with definitions. Relative merits and demerits of the same are compared as well in view of their effectiveness in different cellular scenarios. Finally, the behavior of those QoS parameters are interpreted in the context of various channel allocation schemes.

I. INTRODUCTION

Quality of Service (QoS) is the ultimate deliverable for the design of a cellular system in both commercial and academia. It leads to complex performance analysis of state-of-the-art cellular systems with micro- and pico-cells [1]. A call initiated by a mobile user may suffer blocking [2] at two stages: (a) call initiation, known as fresh call blocking or fresh call congestion (\(P_{\text{FP}}\)) and (b) at the cell boundaries, which is called handoff call blocking or handoff call congestion (\(P_{\text{HP}}\)). Further, the ratio of the total number of calls lost to the total number of call requests is called traffic congestion (\(P_{\text{TCP}}\)).

The mobility of a user plays an important role in cellular system performance [3,4]. Thus it is important to model accurately the user movement in a given scenario with a suitable cellular traffic model [2]. In addition, another important aspect of cellular system design is capacity planning as the availability of channels is limited [3]. Here, a channel typically means a Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA) or Code Division Multiple Access (CDMA) slots or a combination of them. The two major types of allocation schemes are (i) Fixed Channel Allocation (FCA) and (ii) Dynamic Channel Allocation (DCA) [1]. Thus the mobility behavior and choice of proper Channel Allocation Schemes (CAS) with various strategies to implement them have profound effects on QoS. We discuss few of them in this paper.

II. MOBILITY

The probability of releasing an occupied channel in a cell due to handoff, \(\zeta\), is expressed as

\[
\zeta = 1 - \frac{1}{E[H]+1}
\]

where \(E[H]\) is the mean number of handoffs experienced by a user during a call in a non-blocking system. The parameter \(\zeta\) is called mobility. It is also referred as mobility factor or user mobility in some literatures. When the users are static, \(\zeta = 0\). It then increases as the user moves and at most reaches to one. Thus \(0 \leq \zeta \leq 1\). If \(\rho\) denotes the offered load of calls (Erlang/cell), then the effective offered load, \(\rho'\), in presence of mobility is given by

\[
\rho' = \frac{1 - \zeta}{1 - \zeta + P \zeta}
\]

It is evident from the above that the effective load with mobility is less than the offered intensity. This is because in presence of mobility, a proportion of ongoing calls drop during handoff, which, in turn, reduces the traffic load.

III. QoS PARAMETERS

In this section, few QoS parameters to quantify the performance of a cellular system are defined. These are: (i) Forced Termination Probability (\(P_{\text{FT}}\)), (ii) Call Completion Probability given that the call is initiated (\(P_{\text{CC}}\)), (iii) Mean Number of Handoffs per Call (\(E[H']\)), and (iv) different types of congestion which have already been defined in Section I.

A. Definitions

Table I shows the probability of a call being terminated or completed in \(n^{th}\) cell (\(n = 0,1,\ldots\)). The parameter \(P_{\text{FT}}\) is the probability that a call is terminated before its due completion, which is the sum of the entries of the first row in Table I. Thus \(P_{\text{FT}}\) is given by

\[
P_{\text{FT}} = P_T + \frac{(1-P_T)P_H\zeta}{1-(1-P_T)\zeta}
\]

Here \(P_T\) includes the chance of call blocking at the time of initiation (i.e. fresh call congestion). The parameter \(P_{\text{CC}}\) is a measure of how often a successfully initiated call is completed (without being terminated due to channel unavailability). Similarly, from the entries of the second row, we obtain \(P_{\text{CC}}\) as

\[
P_{\text{CC}} = \frac{1 - \zeta}{1-(1-P_H)\zeta}
\]

The Mean Number of Handoffs per Call in presence of blocking, \(E[H']\), is expressed as

\[
E[H'] = \frac{\zeta(1-P_H)-P_T}{1-(1-P_H)\zeta}
\]
B. Relative merits and demerits

Designing a cellular system needs some QoS parameters to be kept under desired bounds. For the QoS parameters discussed in this context, it is important to choose those which provide maximum behavioral information of the cellular model. Congestion or blocking probability is a good choice as this clearly indicates the failed calls or call attempts. In general, there can be three types of congestion as mentioned in Section I. In a non-prioritized design, traffic congestion \( P_{cc} \) is a more effective choice, as it clearly indicates the proportion of the lost traffic, be it fresh calls or handoff calls. In a prioritized design, call congestion of the handoff traffic \( P_{ht} \) is a more effective choice, as it indicates the proportion of the traffic that is lost because of the inability to accommodate the handoff calls. Thus, as the traffic characteristics of the system in terms of the fresh calls and the handoff calls (non-prioritized), No TCP congestion \( \text{TCP} \) is mentioned in Section I. In a non-prioritized design, traffic congestion \( P_{cc} \) is a more effective choice, as it clearly indicates the proportion of the lost traffic, be it fresh calls or handoff calls. In a prioritized design, call congestion of the handoff traffic \( P_{ht} \) is a more effective choice, as it indicates the proportion of the traffic that is lost because of the inability to accommodate the handoff calls. Therefore, the HQS scheme decreases the dropping probability of handoff calls while increasing the blocking probability of fresh calls. This scheme gives higher priority to the handoff calls. Another channel allocation scheme that prioritizes the handoff call is referred to as the Guard Channel Scheme (GCS). In GCS, each BS reserves a fraction of wireless channels exclusively for handoff calls, and the remaining channels, called normal channels, are shared between handoff calls and fresh calls. Both handoff calls and fresh calls use the normal channels first. When the normal channels are used up, a fresh call will be blocked, but a handoff call can still use the reserved channels. In this way, the dropping probability of handoff calls will be reduced. The improvement in the dropping probability of handoff calls is dependent on the number of channels reserved. However, a fresh call is blocked if there are only reserved channels left even though no handoff calls exist. Therefore, the total utilization of wireless channels is decreased. There is a tradeoff between decreasing the blocking probability of handoff calls and increasing the total channel utilization. The number of channels that should be set aside for handoff calls depends on a lot of factors such as the mobility of MSs, the call duration, the call arrival statistics, and so forth. In DCA, cells do not have fixed number of channels and the channels are not assigning specific to a certain cell. A pool of channels is made available to all cells. A specific cell acquires a channel and uses it as long as the Signal-to-Interference (SIR) is above a specified threshold. DCA could be centralized or distributed in nature. Channels are dynamically allocated to cells according to interference and service demand. The advantage of DCA is the fact that there is no need for frequency planning. Distributed DCA schemes are less stable than centralized DCA schemes because the former are affected by local changes that could spread out the whole system, hence affecting its performance and stability.

### III. Channel Allocation Schemes

CAS provides the means to efficiently access the resource in each cell. It also decides upon how to assign the available resource to achieve the highest spectrum efficiency. The underlying strategy in the fixed channel allocation is the permanent assignment of a set of channels to each cell. The radio frequencies are reused by another cell at the same reuse distance (co-channel reuse distance) for all cells. FCA has distributed network architecture. Many different methods have been proposed as variations of FCA, such as channel borrowing methods, channel borrowing with directional locking, and channel borrowing with ordering. Some of them are discussed below [5].

The simplest channel allocation scheme is the *Fully Shared Scheme* (FSS), in which all the available channels are shared by the fresh calls and the handoff calls (non-prioritized). No distinction is made between a handoff call and a fresh call. FSS is widely used in the cellular networks because of its simplicity. In addition, FSS has the advantage of maximizing the utilization of wireless channels. The disadvantage is the increased dropping rate of handoff calls. Recently, intensive research on channel allocation schemes has been conducted to decrease the dropping probability of handoff calls. One such scheme is the *Handoff Queuing Scheme* (HQS). When a Mobile Station (MS) detects that the received signal strength from the current *Base Station* (BS) is below a certain level, called the handoff threshold, a handoff operation is initiated. The handoff operation first identifies the new BS into which the call is moving. If the new BS has unused channels, the call will be transferred to the new BS. If there is no unused channel available, the handoff call will be queued until a channel is released by another call. The HQS scheme is feasible because there is a difference between the signal strength at the handoff threshold and the minimum acceptable signal strength for voice communication. This gives an MS some time to wait for a channel at the new BS to become available. A fresh call will be blocked in the new cell until all the handoff calls in the queue are served. Therefore, the HQS scheme decreases the dropping probability of handoff calls while increasing the blocking probability of fresh calls. This scheme gives higher priority to the handoff calls. Another channel allocation scheme that prioritizes the handoff call is referred to as the Guard Channel Scheme (GCS). In GCS, each BS reserves a fraction of wireless channels exclusively for handoff calls, and the remaining channels, called normal channels, are shared between handoff calls and fresh calls. Both handoff calls and fresh calls use the normal channels first. When the normal channels are used up, a fresh call will be blocked, but a handoff call can still use the reserved channels. In this way, the dropping probability of handoff calls will be reduced. The improvement in the dropping probability of handoff calls is dependent on the number of channels reserved. However, a fresh call is blocked if there are only reserved channels left even though no handoff calls exist. Therefore, the total utilization of wireless channels is decreased. There is a tradeoff between decreasing the blocking probability of handoff calls and increasing the total channel utilization. The number of channels that should be set aside for handoff calls depends on a lot of factors such as the mobility of MSs, the call duration, the call arrival statistics, and so forth. In DCA, cells do not have fixed number of channels and the channels are not assigning specific to a certain cell. A pool of channels is made available to all cells. A specific cell acquires a channel and uses it as long as the Signal-to-Interference (SIR) is above a specified threshold. DCA could be centralized or distributed in nature. Channels are dynamically allocated to cells according to interference and service demand. The advantage of DCA is the fact that there is no need for frequency planning. Distributed DCA schemes are less stable than centralized DCA schemes because the former are affected by local changes that could spread out the whole system, hence affecting its performance and stability.
It is remarked that a hybrid of FCA and DCA can also be used which is often known as flexible channel allocation. In such a scheme, each cell is assigned a set of permanent channels and the rest of the channels are allocated to the cell as in DCA.

IV. IMPACT OF CAS ON QoS PARAMETERS

Some CAS are based on adaptive technique where channels are allocated and reserved in a dynamic way using teletraffic analysis, prediction of injected traffic, and prediction of mobile terminal movement. In some prediction schemes, it is sufficient to take care of the radio resource that the mobile will need in its predicted future location. In general, the resource reservation mechanism consists of two parts: (a) some of the bandwidth reserved in the next cell the mobile is likely to visit and (b) a common pool of dynamically adjusted bandwidth used to accommodate other unpredicted flows. The next cell is predicted based on mobility pattern observed in various cells. Queuing of fresh calls and handoff requests is another approach as mentioned earlier that reduces the congestion and forced termination of calls. There is a tradeoff between the increase in service quality and corresponding decrease in total carried traffic. One of the key points of using queuing in CAC is that service differentiation can be managed with modified queuing discipline. Instead of FIFO queuing strategy, other prioritized queuing discipline can be used to maintain priority level in each service class.

When the call requests come from an infinite population of users it is considered to be memory-less Poisson distributed process. It is found in the literature [1] that in such a case, the handoff process becomes nearly Poisson or Poisson distributed (smooth traffic i.e. Variance $<$ Mean). Consequently, the handoff call congestion becomes less than or equal to fresh call congestion in FSS regardless the use of FCA or DCA [6,7]. The traffic congestion is noticed to be lying in-between fresh and handoff call congestion since it provides the proportion of lost calls containing both fresh and handoff call requests [6].

However, when the call requests come from a finite population of users i.e. a less populated area, the fresh call arrival process follows Erlang distribution [8]. Under such a scenario, a fresh call request experiences much lower congestion compared with the handoff process due to low variance of the fresh call arrival process. Therefore, to reduce the congestion experienced by the incoming handoff traffic, channel reservation policy for the handoff traffic is used. ‘Channel reservation’ of $r$ channels that the combined stream (fresh + handoff) will be competing for channels when the number of available channels is $r$ or more. When less than $r$ channels are available, only the handoff component of the stream is allowed.

It is noticed that as the number of reserved channels increases, the handoff call congestion decreases, and as a counter effect, the fresh call congestion, due to the lesser availability of channels, drastically increases. Under such cases, an acceptable level of fresh call congestion (for e.g. less than 6%) is provided to the system designer. The objective is to vary $r$ such that handoff call congestion reduces to a minimum, at the same time fresh call congestion does not exceed the acceptable level. Further it is also an issue to investigate the difference in-between rate of increment in fresh call congestion and rate of decrement in handoff call congestion for variation in $r$. Hence, a suitable number of reserved channels, depending upon the congestion estimate, must be employed to achieve optimum Quality of Service (QoS).

V. CONCLUSION

In this paper, we discussed on some important technical QoS parameters in designing of a cellular systems. We conclude that for non-prioritized design, $P_{HC}$ along with $P_{PT}$ are the most effective choices of QoS parameters. However, for prioritized cellular systems, $P_{HC}$ and $P_{PT}$ are favorable since more priority should be given to the handoff calls. Finally, we observe that channel reservation scheme is most effective for the latter type of design. We also remarked upon how a suitable number of reserved channels need to be kept, in view of the various congestion estimates, to obtain the desired Quality of Service (QoS) bounds.

REFERENCES


