Chapter Two

Literature Review

2. Literature Review

2.1 Introduction

Supporting multimedia applications with different quality of service (QoS) requirements in the presence of diversified wireless access technologies (e.g., 3G cellular, IEEE 802.11 WLAN, Bluetooth) is one of the most challenging issues for the forth-generation (4G) wireless networks. In such a network, depending on the bandwidth, mobility, and application requirements, users will be able to switch among the different access technologies in a seamless manner. Efficient radio resource management and call admission control (CAC) strategies will be key components in such a heterogeneous wireless system supporting multiple types of applications with different QoS requirements [4].

A call admission control scheme aims at maintaining the delivered QoS to the different calls (or users) at the target level by limiting the number of ongoing calls in the system. One major challenge in designing a CAC arises due to the fact that the network has to service two major types of calls: new calls and handoff calls. The QoS performances related to these two types of calls are generally measured by new call blocking probability and handoff call dropping probability. In general, users are more sensitive to dropping of an ongoing and handed over call than blocking a new call. Therefore, a CAC scheme needs to prioritize handoff calls over new calls to keep the handoff dropping probability below some threshold. Also, the new call blocking probability should be maintained below or at the target level. After all, the resource utilization should be maximized while achieving the QoS requirements [1].

In contrast to the traditional voice-oriented circuit-switched cellular wireless networks, the 4G networks will be based on packet- switching at the wireless interface and will be internetworked with IP-based Internet. Therefore, while designing a CAC scheme for such a network, packet-level performance measures (e.g., packet dropping probability and packet transmission delay) at both the wireless interface and the wired interface (e.g., at the IP-aware wireless router/base station) will need to be considered in addition to the call-level performance measures.

For the evolving 4G networks, the traditional and existing CAC algorithms (e.g., those for 3G systems) needs to be modified with a view to

- Handling the handoff calls between the networks (i.e., *vertical handoff*). Resource reservation and call admission control become more complicated due to the presence of heterogeneous wireless access environment in which the mobile terminals have the ability to connect to different types of networks.
- Prioritizing the different types of calls. Some calls might use real-time applications which have more strict QoS requirements than those using non-real-time applications. CAC must be performed based on different QoS requirements.
- Taking into account packet-level performances. 4G wireless networks will operate purely on packet-based data transfer. The CAC algorithm must consider not only call-level QoS but also the packet-level QoS. In other words, the CAC algorithm needs to evaluate the availability of the network resources by taking into account the packet-level performance statistics.
- Internetworking with the IP-based Internet. The CAC algorithm should be aware of the availability of the network resources at the

wireless-Internet gateway and in the wired network so that wireless resources are not wasted due to dropping of packets at the wired-part of the network [4].

2.2 Related Work

Although the admission control concept has been extensively studied in the literature, only a limited number of contributions are developed within the context of 4G networks. The objective of this section is to highlight the recent work on this field in order to provide the reader with an up to date State of the Art.

One of the first attempts towards introducing admission control in the fourth-generation cellular mobile networks has been made by **Jeong et al in 2005[6]**. In their work, the authors present a CAC scheme that supports the QoS requirements of the accepted connections in IEEE 802.16e wireless systems. The objective of the proposed CAC is to maximize the utilization of the resources, considering as basic parameter the capacity estimation of the cell. Furthermore, during the admission and scheduling process, the base station distinguishes the delay-sensitive real-time (RT) from the delay-tolerant non-real-time (NRT) connections. The proposed scheme achieves to fulfill the QoS demands of the connections, but in temporary overloaded situations, only NRT class connections can be admitted, thus excluding entirely the RT traffic.

In 2009 Qian et al [7] proposed a novel radio admission control scheme for multiclass services in LTE systems. The authors introduce an objective function to maximize the number of admitted users and propose a CAC algorithm that implements a service degradation scheme whenever a limitation of resources occurs. In their paper, there is a service differentiation approach, with different portions of bandwidth devoted to each traffic class. However, in the presented numerical results there is no plot that distinguishes the blocking probabilities for the different types of traffic, thus not providing any information about the actual handling of the multiclass services.

Anas et al in 2008[8] proposed an admission control algorithm for LTE utilizing the fractional power control (FPC) formula agreed in 3GPP , In their work, GBR (Guaranteed Bit Rate) is the only considered QoS of the bearer, while each user is assumed to have a single-bearer. The main idea of their proposed algorithm is that the current resource allocation can be modified in order for the new user to be admitted without violating the power restriction for the physical uplink shared channel (PUSCH).

Lei et al in 2008[9] introduced a resource allocation algorithm along with a connection access control scheme for LTE systems with heterogeneous services. Their proposed CAC assigns different portions of bandwidth for real-time and non-real-time connections, thus balancing the ongoing connections of different traffic classes and facilitating the support to potential handoff users. However, the results show that the cell throughput remains the same whether the proposed admission control scheme is applied or not.

In **Kwan et al 2010 [10]** Presented a novel predictive admission control scheme. The authors propose a new cell load measurement method and mechanisms for predicting the load increase due to the acceptance of new connections.

In the same content, a resource-estimated CAC algorithm is proposed in **Bae et al 2009[11]**. Specifically, whenever a service request occurs, the resource-estimated CAC algorithm calculates the required amount of resources in order for the request to be served. This amount is determined based on the service type, the modulation and the coding scheme level of the particular user. However, the results show that the proposed CAC is beneficial only in terms of packet delay, since the average data rate and the cell utilization are decreased.

Regarding the bandwidth reservation, a downlink CAC algorithm with look-ahead calls for 3GPP LTE mobile networks is presented in **Sallabi and Shuaib 2009[12]**. The proposed algorithm handles the advance resource reservations, providing a high probability that the advance calls will be immediately served once their session is ready to start. Nevertheless, it is hard to derive useful conclusions since there is no reference or comparison to other admission control methods.

Sílvio Martins Reis and Paulo Roberto [13] presented a Connection Admission Control (CAC) algorithm for the IEEE 802.16estandard, based on dynamic bandwidth reservation. These reserves are obtained by segmenting the amount of the channel's bandwidth by thresholds, which are dynamically adjusted according to the admissions of handoff and new connections demonstrating that the proposed algorithm can avoid the waste of network resources, increase its efficiency and provide QoS, in terms of bandwidth, for applications [5].

2.3 General Model of CAC

In general, a CAC scheme has three main components; information management, resource reservation, and admission control as shown in Figure 2.1. These three components collaborate with each other by exchanging information with a view to achieving specific CAC objectives such as minimizing QoS degradation or maximizing the revenue.

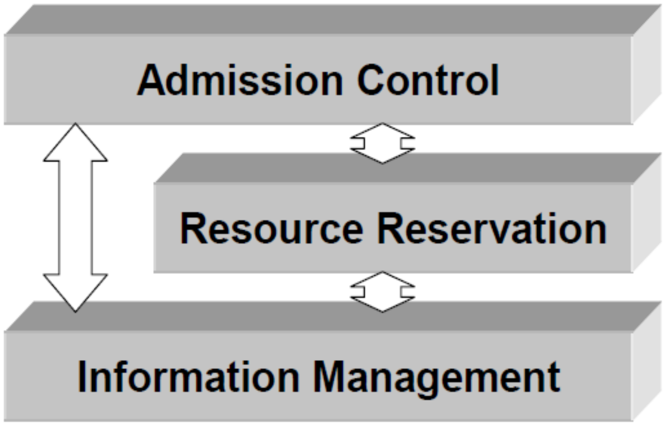

Figure 2.1: Call Admission Control Mechanism Components

2.3.1 Information Management

This is required for storing, exchanging, and maintaining the state of the network. The type and amount of information (e.g., number of channels used by ongoing calls, position and velocity of a mobile) depends on the types of the CAC algorithm. This information can be either exchanged among the cells or used locally.

2.3.2 Resource Reservation

This is required to reserve the resources according to the users' needs. This component can consult the information management component and use optimization and prediction techniques to calculate the amount of resources to be reserved for handoff calls and maximize the resource utilization.

2.3.3 Admission Control

The admission control component is responsible for making decision on whether an incoming call (either a new call or a handoff call) can be accepted or not. This component consults the information management and resource reservation components.

Most of the admission control algorithms are rule-based, i.e., based on examining the predefined conditions. The outcome of the algorithm could be either to accept the call, reject the call, or queue the call until resources become available. As an example, for a guard channel-based CAC scheme, the information management component maintains the current number of busy channels, and the resource reservation component reserves a pool of channels for handoff calls. The admission control module uses information from both the components to decide whether a call can be accepted or not.

2.4 Classification of CAC

The admission control schemes can be classified by a number of different approaches. Some of these approaches are shown in figure (2.2).

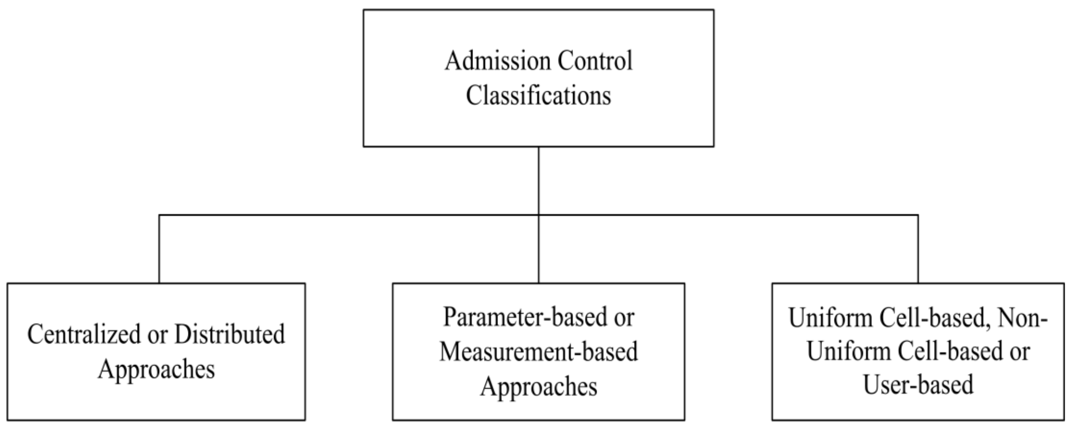

Figure 2.2: Classification of CAC

2.4.1 Centralized and Distributed Approaches

A call admission control algorithm can operate in a centralized or in a distributed fashion. In the former case, the CAC algorithm is executed in a central site (e.g., mobile switching center (MSC)). In this case, information from every cell needs to be transferred to the central site and the CAC needs to be performed remotely from the local cell. In case of distributed CAC, the CAC algorithm is executed locally at the base station of each cell. A distributed CAC algorithm can follow either a collaborative or a local approach. In the former case, information is exchanged among the neighboring cells for resource reservation and admission control while the decision is made locally. In the latter case, information collection and decision making are done locally. Although the collaborative approach can provide more accurate information for CAC decision, it incurs more communication overhead.

2.4.2 Traffic Descriptor-Based and Measurement-Based Approaches

A call admission control policy can use either a traffic descriptor based or a measurement based approach. In the traffic descriptor-based approach, it is assumed that the knowledge about the traffic pattern of the incoming calls is available. Therefore, a CAC algorithm can simply determine the expected amount of resource usage by summing all the resources used by all ongoing calls and the incoming calls together. If this is less than some predefined threshold, the incoming call is accepted, otherwise the call is rejected. Although traffic descriptor-based CAC is simple, it is relatively conservative, since the ongoing calls will not use maximum amount of resource as specified in the descriptor all the time.

Instead of using explicit traffic descriptors, the information on the traffic pattern can be obtained by measuring the characteristics of the call. In that case, call admission control decision can be made dynamically based on the actual state of the network. Measurement-based CAC schemes are based on this principle.

Most of the CAC algorithms in the hard-capacity cellular networks are traffic descriptor-based. Most of CAC algorithms in CDMA systems are measurement-based in which SIR information is measured and used to ensure the QoS of the ongoing calls.

The CAC algorithms used in WLANs mostly adopt a measurement-based approach.

2.4.3 Classification Based on the Granularity of Resource Control

A taxonomy of CAC algorithms by considering the granularity of the resource control. Three different criteria were used to categorize a CAC algorithm.

The first criterion is the type of information used by decision making process for call admission control. Generally, CAC algorithms consider resource usage of the mobiles, but some of the algorithms consider the mobility patterns of the users.

The second design criterion is the spatial distribution (position and movement) of the mobiles which can be either uniform or non-uniform.

The third criterion is based on how the information is organized and manipulated by the CAC algorithms.

The information can be on the aggregate of flows or on the per-user basis and the CAC algorithms use the information on the group of mobiles or on individual mobile, respectively. Based on these criteria, the granularity of the CAC algorithms is different. For example, the algorithm which considers the resource usage of all ongoing calls assuming uniform spatial distribution (e.g., the guard channel scheme) has the largest, and the algorithm which considers the mobility of individual user assuming non-uniform spatial distribution has the smallest granularity of resource control.

2.5 Traditional CAC Approaches in Cellular Networks

There are many approaches for call admission control in cellular wireless networks such as Guard Channel Approach, Partitioning and Sharing Approach, Collaborative Approach Based on Estimation, Non-Collaborative Approach Based on Prediction, Mobility-Based Approach, and Pricing-Based Approach.

2.5.1 Guard Channel Approach

To prioritize handoff calls over new calls, some channels (referred to as guard channels) are reserved for handoff calls. Specifically, if the total number of available channels is C and the number of guard channels is $(C - K)$, a new call is accepted if the total number of channels used by ongoing calls (i.e., busy channels) is less than the threshold K, while a handoff call is always accepted if there is an available channel. According to this channel reservation, the threshold must be chosen such that the handoff call dropping probability is minimized while the system can admit as many incoming calls as possible.

Although the guard channel scheme with a static threshold is easy to implement, it may not be efficient. Higher channel utilization could be achieved through adaptation of the threshold according to the state of the network.

A more general scheme used is the fractional guard channel scheme. In this case, an incoming call is accepted with certain probability which depends on the number of busy channels. In other words, when the number of busy channels becomes larger, the probability for accepting a new call becomes smaller and vice versa. This helps to keep the handoff call dropping probability lower than the desired value and also avoid congestion in the cell.

2.5.2 Partitioning and Sharing Approach

This approach for reserving channels and admitting new calls is based on the concepts of *complete sharing* and *complete partitioning*. In case of complete sharing, the handoff calls and new calls can use all the available channels. In contrast, in case of complete partitioning the channels reserved for handoff and new calls are not shared between these types of calls.

Instead of using pure complete sharing, which is unable to prioritize the calls, and complete partitioning, which is relatively conservative, a hybrid model for resource reservation and CAC is used. In this hybrid scheme, the channels are divided into three categories: channels dedicated for new calls, channels shared among the new calls and handoff calls, and channels reserved for handoff calls. By combining complete partitioning and complete sharing, resource reservation becomes flexible to control the performance of the system. This type of hybrid resource reservation can handle calls with different priority levels as well.

2.5.3 Collaborative Approach Based on Estimation

This is a distributed approach for call admission control. In this case, information is exchanged among the neighboring cells for resource reservation and admission control, while the admission control decision is made locally. This approach use estimates of call dropping and call blocking probabilities. The maximum number of ongoing calls N is estimated from the following:

$$Phd = \frac{1}{2}\mathrm{erfc}\left(\frac{N-m}{\sigma}\right) \qquad (2.1)$$

Where *Phd* is the target call dropping probability, and m and σ denotes the mean and variance of the number of calls in the home cell, respectively. The mean and variances are approximated from the number of users in the home cell and the neighboring cells.

The call blocking probability *Pnb* (t) at time t *-1* to t is estimated locally as follows:

$$Pnb\,(t) = (1 - w)\,Pnb\,(t - 1) + w\frac{s(t)}{r(t)} \qquad (2.2)$$

Where $s(t)$ and $r(t)$ are the number of blocked calls and the number of calls that arrived during time interval t -1 to t, respectively, and w is the weight used for calculating the exponential weighted moving average. The decision on whether an incoming call is accepted or rejected is made based on (2.1) and (2.2).

2.5.4 Non-Collaborative Approach Based on Prediction

In a Pico-cellular wireless network with high user mobility, exchanging information among the cells to make resource reservation and admission control might incur significant control overhead. Therefore, CAC algorithms designed based on local information (e.g., history of bandwidth usage) would be desirable. In such a case, resource reservation is based only on local information in the home cell which is used to predict the resource needed in the future. In [13] Two prediction techniques were used *Wiener filtering* and *time series analysis* (e.g., ARMA (autoregressive moving average) model). In the former case, the prediction can be done directly from the historic data, whereas in the latter case the time series model needs to be constructed and the corresponding parameters need to be estimated so that the prediction can be performed based on this model afterwards. Such a local predictive approach to call admission control was shown to perform as good as a collaborative approach when the traffic fluctuation is moderate.

2.5.5 Mobility-Based Approach

Mobility-based approaches exploit the user mobility information for efficient call admission control. Figure 2.3 shows the shadow cluster for a mobile in cell *C* which is moving towards north. The cells in a shadow cluster usually have different levels of predicted traffic intensity. For example, in figure (2.3), the predicted traffic intensity in cells denoted by *B* in the vicinity of cell *C* will be more than that in the cells denoted by *A*. To calculate the shadow cluster and the corresponding levels of intensity, the information on call holding time, current direction, velocity, and position of the active mobile terminal need to be considered.

The base station in the home cell must inform its neighboring base stations of the location and mobility of the active mobile terminal (e.g., this information can be obtained from global positioning system). In a cell, the amount of resources reserved for handoff calls is based on the number of calls moving to that cell and the corresponding probabilities. The estimated resource usage $Cuj(t)$ in cell j at time t is expressed as follows:

$$Cuj(t) = Cuj(t-1) - C_{uj}^{out}(t) + C_{uj}^{in}(t) \qquad (2.3)$$

Where $C_{uj}^{out}(t)$ is the estimated amount of resources that will be freed by active users whose calls will be either terminated or handed over to other cells and $C_{uj}^{in}(t)$ is the estimated amount of resources that will be occupied by active mobile terminals moving from neighbors cells within the shadow cluster.

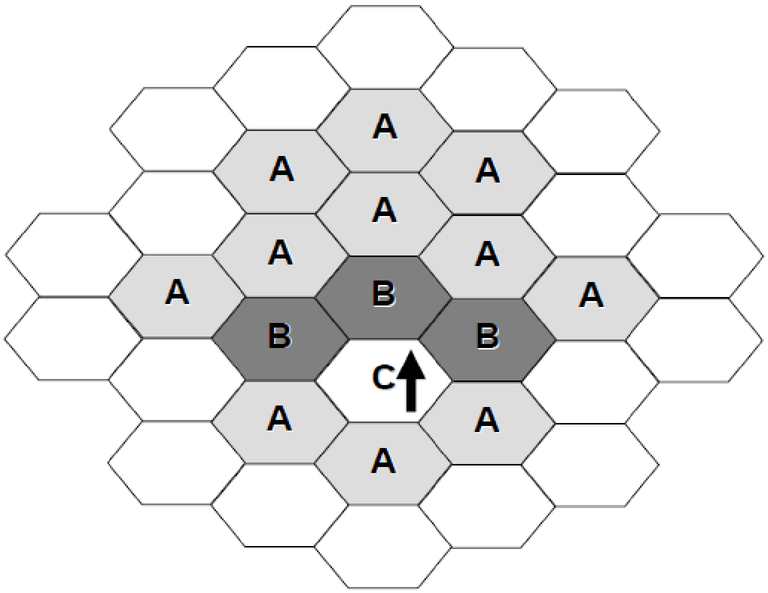

Figure 2.3: Shadow Clustering: C is the home cell of active mobile terminal, B is the bordering neighbor, and A denote non bordering neighbor.

2.5.6 Pricing-Based Approach

A pricing-based approach to call admission control maximizes the utility of the wireless resources. The utility is generally defined as the users' level of satisfaction with perceived QoS. For example, the utility is a decreasing function of new call blocking and handoff call dropping probabilities.

However, maximizing the utility of the network might not maximize the revenue of the service provider. The tradeoff between the user satisfaction and revenue is illustrated in Figure (2.4). Specifically, for higher user satisfaction, more resources should be allocated to each user. In contrast, to maximize revenue under flat rate pricing, the allocations need to be degraded to accommodate more number of users. Therefore, a CAC scheme can be designed such that the optimal point can be obtained.

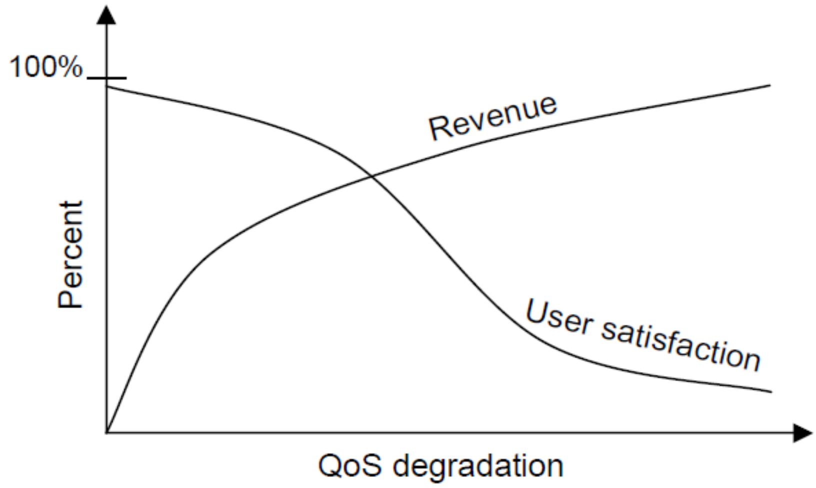

Figure 2.4: Tradeoff between QoS degradation and network revenue.

In [15], the optimal point between utility and revenue was determined in terms of the new call arrival rate, and a pricing scheme was developed to achieve this optimal effective arrival rate in the network. In this case, the QoS metric *Pb* referred to as the grade of service (GoS) is defined as follows:

$$Pb = \alpha Pnb + \beta Phd \qquad (2.4)$$

Where α and β are the weights corresponding to the new call blocking and handoff call dropping probabilities, respectively, and $\alpha + \beta = 1$.

Based on this optimal new call arrival rate, the pricing scheme is developed by changing the cost of a call. The pricing scheme adjusts the fee dynamically by taking the state of the network into account. If the network is congested, it will charge *peak hour price* $p(t)$ which is higher than the normal hour price $p0$. According to this pricing scheme, the demand function which describes the reaction of users to the change of price is expressed by

$$D\,[p\,(t)] = \exp\left(-\left(\frac{p(t)}{p0} - 1\right)^2\right) \qquad (2.5)$$

The peak hour price is calculated by considering the state of the network. With this pricing scheme, a user has an incentive not to initiate a call during peak hours so that the congestion of the network can be avoided.

The main ideas of all the above CAC approaches are summarized in Table (2.1).

Table 2.1: Different Approaches for call admission control in cellular wireless networks.

Approach	**Main Idea**
Guard channel	Some portion of the wireless resources is reserved for handoff calls so that handoff call dropping probability can be maintained below the target level.
Fractional guard Channel	New calls are gradually blocked according to the current status (i.e., the number of ongoing calls) of the network.
Collaborative	The neighboring cells exchange information about the network status so that the resource reservation can be made in advance accurately.
Non-collaborative	By using prediction techniques (e.g., ARMA model, Wiener filtering) to project the amount of the resources required locally so that the resources can be reserved in advance without the need for information exchange among neighboring cells.
Mobility-based	Mobility information (i.e., position and direction of movement) of the mobiles can be used to enhance the accuracy of the resource reservation.
Pricing-based	Dynamic pricing is used to limit the call arrival rate so that the maximum utility and revenue of the system is achieved.

2.6 Advantages of Using Admission Control Schemes

One of the main reasons of using CAC schemes is to provide an acceptable quality of service for the consumer and to ensure convenient revenue for the service provider. CAC is also used to prioritize some services or user classes while maintaining a fair resource sharing between them.

2.6.1 Guaranteeing Quality of Service (QoS)

CAC is essential to guarantee the signal quality in interference-limited wireless networks. For instance CDMA wireless networks have a soft capacity limit so that the more loaded the network is the more deteriorated is the signal quality for users in terms of the interference level or the signal to interference ratio (SIR). Hence CAC schemes admit users only if it can maintain a minimum signal quality to admitted users (including the new call and existing calls). In this case the admission criterion can be the number of users (per cell and/or per group of neighbor cells) interference level or SIR total transmitted power by Base Station (BS) or received power by either BS or the mobile station .

Since dropping an active call is usually more annoying than blocking a new call CAC is employed in bandwidth-limited wireless networks to control the probability of handover failure (Phf). This can be implemented by reserving some resources for handover calls exclusively. The admission criterion can be either the number of users (per class in a multiple-class system) or an estimate of handover failure probability .Resources availability can also be used as a criterion for admission. Whatever the used admission criterion handover calls receive less strict admission conditions compared with a new call which might lead to an increase in the new call blocking rate.

When packet-oriented services are provided by wireless networks, network overloading can cause unacceptable excessive packet delay and/or delay jitter. The throughput level at the network or user level can also be dropped to unbearable levels. Therefore CAC should be used to limit the network level to guarantee packet-level QoS parameters (packet delay, delay jitter and throughput). In this case the number of user's resource availability and/or an estimate of the packet-level QoS parameters can be utilized as an admission criterion.

CAC schemes are used in wireless networks offering data services to guarantee a minimum transmission rate. The use of CAC to ensure a minimum transmission rate has been studied extensively in wire line network .The problem however is more complicated in wireless networks because of user mobility (implying handover and link quality variations) limited bandwidth and mutual co-channel interference.

2.6.2 Revenue-Based Admission Control Schemes

From the network perspective a new call admission has both rewards and penalties. The reward comes from the utilization of network resources for a certain amount of revenue. However there is a potential penalty particularly at high network loading values which can cause deterioration in the QoS offered to the already admitted users and even potentially some call dropping .Hence CAC can be used to increase the network revenue function based on the potential reward and penalty of admitting new calls. In this case the admission criterion can be the number of users or an estimate of the probability of QoS deterioration (e.g. lower transmission rate than the acceptable level).

2.6.3 Prioritize Some Services/Classes

Giving higher priority to some services or classes can be one of the objectives of CAC schemes. For instance there is a common belief among network operators and researchers as well that voice services should be given higher priority over data services since the former are more rewarding (at least for the time being) to the network operators. Also giving higher priority to some classes inside the same service might be needed to differentiate between different user classes based on some criteria such as the subscription fees or the urgency of the call .

2.6.4 Fair Resource Sharing

Fairness among different users in the same class (with different channel conditions and mobility characteristics) and among users of different classes is one of the objectives of CAC. CAC can be employed to admit/reject users based on the allocated resources such that no user class dominates the system resources [16].