

## ***Chapter Three***

### ***Call Admission Control in LTE Networks***

### **3. Call Admission Control in LTE Networks**

#### **3.1 Introduction**

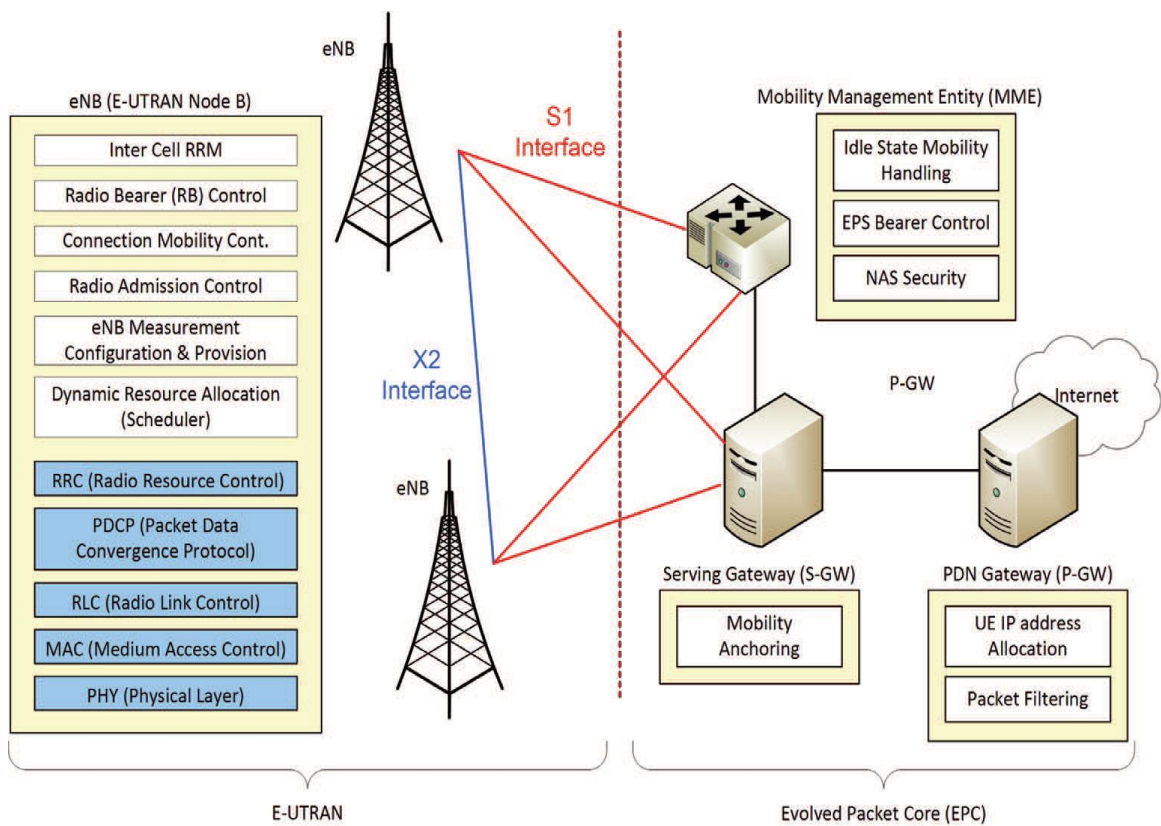
The amount of mobile users is evermore increasing and users are demanding more and more from the mobile infrastructure. To comply to this increasing demand, researchers are working on the development and improvement of a new networking technology called Long Term Evolution (LTE), which is an evolving wireless standard developed by the 3<sup>rd</sup> Generation Partnership Project (3GPP) which, along with 3GPP HSPA+, 3GPP EDGE Evolution and Mobile WiMAX (IEEE 802.16e), opens the road to 4G technologies. The standard is focused on delivering high data rates for bandwidth-demanding applications and on improving flexibility and spectral efficiency, thus constituting an attractive solution for both end users and mobile operators. An important feature of LTE that differentiates it from conventional mobile standards is the all-IP packet based network architecture, which further ensures the seamless integration of internet applications and facilitates the convergence between fixed and mobile systems [1, 5].

The radio interface of LTE is based on Orthogonal Frequency Division Multiplexing (OFDM) and supports Multiple-Input-Multiple-Output (MIMO) technology. The standard defines asymmetrical data rates and modulations for uplink and downlink, using different access schemes for each link. In particular, Orthogonal Frequency Division Multiple Access (OFDMA) is employed in the downlink, while the technically similar but less power demanding Single Carrier – Frequency Division Multiple Access (SC-FDMA) is used in the uplink. In terms of the wireless spectrum allocation, LTE supports variable channel bandwidths that vary from 1.4 to 20 MHz and can be deployed in different frequency bands [5].

## 3.2 LTE Architecture

The LTE cellular network consists of two main parts as shown in figure (3.1). These are the Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and the Evolved Packet Core (EPC) network. The typical E-UTRAN consists only of evolved Node-Bs (eNBs), which represent the Base Stations (BSs) used to provide radio access, and its main role is to manage the radio resource and mobility in the cell in order to optimize the communication among all User Equipment's (UEs) that are within its radio coverage. The eNBs can communicate with each other through the X2 interface and can access the EPC by means of the S1 interface.

The EPC provides permission to UEs to access the LTE cellular system and support for multimedia service connectivity, roaming and mobility. The EPC consists of several network entities such as the Mobility Management Entity (MME) that supports the mobility management, the Home Subscriber Server (HSS) that maintains the subscription profiles of each user, the Packet Data Network Gateway (P-GW) that represents the packet data network gateway to the Internet, the Serving Gateway (S-GW) that manages the user data tunnels between the eNode-Bs and the P-GW, under the supervision of the MME, and the Policy and Charging Rules Function (PCRF) that monitors and controls the policy rules and charging of mobile users for the services they are using [17].



**Figure 3.1:** LTE Architecture and functional split between E-UTRAN and EPC [5].

The evolution path of system architecture was towards supporting completely packet-switched services and to comply with some radio design goals such as hard handover instead of soft handover which existed in HSPA. Also all radio functionalities were concentrated in nodeBs to make the architecture more flat. Changing the system architecture followed different targets such as:

- Optimization of system architecture for packet-switched services, as in LTE there was no support for circuit-switched services.
- Support for higher throughput comparing to former technologies due to the end-user higher bit rate demands.
- Support for improvement in response time for activation and bearer set-up.

- Packet delivery delay reduction.
- Support for simplification of the whole system.
- Interworking with other 3GPP access networks and other 3GPP technologies.

As many of the targets mentioned above can be achieved by using flat architecture it was proposed for LTE. In the architecture fewer nodes are involved and therefore the latencies decrease and performance increases [19]. To meet the requirements of reduced latency and cost, LTE has a flat system architecture that contains a reduced number of network nodes along the data path. A reduction in the number of nodes makes it possible for example to reduce the call setup times, as fewer nodes will be involved in the call setup procedure. Figure 3.2 illustrates the architecture evolution of LTE over Release 6 architecture.

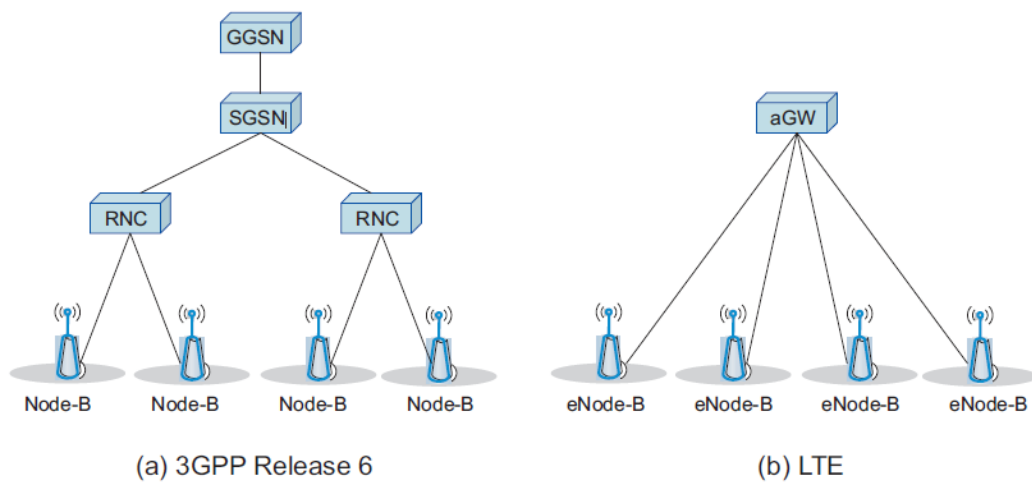
In Release 6, part of the Radio Resource Management (RRM) functionalities e.g. PS, are located in the Node-B. While, the RNC handles RRM functionalities e.g., Admission Control (AC), mobility management (locally), etc. and transport network optimization.

It further acts as a termination point for the radio protocols. The Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) act as an anchor node and visiting node in the visiting network and home network respectively. Further, SGSN handles both mobility management and session management.

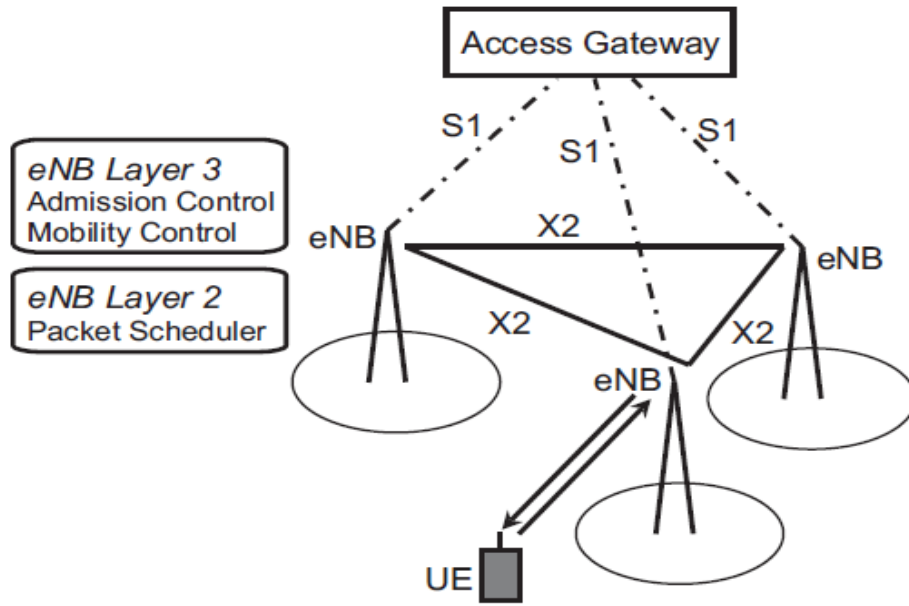
In the LTE architecture, the Access Gateway (aGW) terminates the user plane for the UE, and handles the core network functions provided by the GGSN and SGSN as in Release 6. As shown in Figure 3.3, the RRM functionalities e.g., AC, mobility control including handover, PS etc., are located in Evolved Node B (eNode-B) instead of RNC as in Release 6. The System Architecture Evolution (SAE) focuses on the enhancement of

packet switched technology. I.e. higher data rates, lower latency (both in user plane and control plane), packet optimized system, and support of multiple radio access technologies.

These goals will be achieved using fully IP based network, simplified network architecture, and distributed control [18].



**Figure 3.2:** The 3GPP Release 6 architecture and evolved system architecture for LTE reducing the number of nodes along the data path from 4 to 2 [3].



**Figure 3.3:** UTRAN LTE system architecture

LTE provides service differentiation by adopting a class-based Quality of Service (QoS) concept. In particular, each data flow between the user equipment and the P-GW (called EPS bearer) is assigned a QoS profile. A total of nine profiles are defined in the specification (3GPP, 2011) that can be mapped to different types of applications such as real time video and voice services, online gaming, etc. Each profile involves the bearer type, the flow priority, an upper bound for the packet delay and the packet error rate. The bearer type indicates whether a Guaranteed Bit Rate (GBR) will be provided to the bearer by permanently allocating network resources during the data session. The essential difference between GBR and non-GBR bearers is that, in the first case, a connection may be blocked if the network does not have the resources to guarantee the desired QoS of these connections. This concept is known as Call Admission Control (CAC) and it is an important component of radio resource management. CAC algorithms are usually implemented in eNBs and their role is to determine

whether a new connection request should be accepted or rejected, depending on the available network resources [5].

### **3.3 Call Admission Control in LTE**

Call admission control is a process to ensure and maintain certain level of Quality of Service (QoS) for real time and non real time call requests in the network. The main objective of CAC is to maintain the efficient resource allocation and to monitor the resource utilization in the high volume of traffic. CAC manages the total bandwidth with respect to the number of call request available in the base station.

The call requests are classified into new call or handoff call and real time or non real time call request. CAC allocates signal strength for eNB with a minimum threshold value, when an eNB's signal strength reaches below this threshold value the call request will be blocked [21].

The serving eNodeB carries out a cell selection process that consists of allocating the user to the cell with the lowest load level and fulfills the QoS requirements requested by the UE. In turn, an admission control may be performed by the selected target eNodeB according to the received quality of service information. In case that there is no capacity available for the handoff call in the selected cell (i.e. the admission control is not passed), another cell from the candidates cells will be selected instead. Once the decision of the handover is taken, the serving eNodeB informs the UE by the new eNodeB and orders him to ask the detachment and to achieve the handover. The target eNodeB can now start sending data to the UE and, at the same time, send a path switch message to the Serving Gateway to inform that the UE has changed cell [2].



The design of CAC for a fixed network is simple, as the call admission is based on the available resources and QoS requirements of the new calls. However, the mobile environment is more complicated than the fixed network, as the eNodeB may reserve some bandwidth to admit the handoff calls. If the eNodeB reserves some bandwidth for handoff calls, and the network happens to have few or no handoff calls, then those resources may be wasted or underutilized. On the other hand, if the eNB allocates minimum resources for handoff calls, then the handoff calls may be dropped [1].

The available resources in the base station are distributed to the available UE's in the network with maximum and minimum threshold value. When a call request is received by the base station the initial status of the available resources is checked. Quality of Service focuses on the guarantee of service provision based on the quality policy specified for the service request. The design of a call admission control depends on the following parameters;

- **Availability of Resources**

In eNB's new call and handoff call request are admitted based on the available resources. If the resources are limited, call admission decision is made with the acceptance of the available resources. While designing the call admission control mechanism, call admission criteria considers the load of the network. Prediction based decisions are employed to admit the new calls with respect to resource reservation.

- **Quality of The Network Parameters**

The connection quality plays the major role in the establishment of interference free transmission. Received signal strength (RSS) is used to evaluate the quality of the link between the network

components of the system. Quality parameters for each network element is designed and taken into account for the design of the call admission process.

- **Quality Policies**

Qos requirements are categorized with regard to the parameters like throughput, delay, fairness and bandwidth utilization. The traffic characteristics are analyzed to find the parameters for the performance degradation on the network. QoS provision is to guarantee the user request with quality policies based on the Qos demands of the user. The traffic conditions of the network are predicted to ensure the need based service with the fulfillment of required network resources.

- **Call Prioritization**

The incoming call requests are classified into real time (rt) and non real time (nrt) calls, the real time call request are provided with highest priority when compared to the non real time calls. E.g. Live video streaming calls are more prioritized than the internet browsing. Highest priorities are provided for handoff calls and emergency related calls. Reservation schemes and queuing mechanisms are introduced to deploy the priority for call request.

- **Mobility Management**

In order to reduce the call blocking and call dropping probability, the mobility factors are considered to predict the movement UEs across the base station. Mobility prediction helps the call admission

process to classify the call request either new call or handoff call, as a result it produces the efficient resource allocation.

- **Optimization Methodologies**

To enhance the performance of call admission process, wide range of optimization techniques are introduced. The main objective of the call admission framework is to provide end to end QoS with the ability to manage the transmission interference problems in the radio channel. In order to ensure the better QoS the transmission architecture involves operations like network planning, parameter configuration and optimization. Network architecture is modified based on the status of the network to generate the flow of data and control over error. Optimization process reduces the complexity of the call admission process and the parameters for each call request specified with the threshold value. The incoming calls or new calls are evaluated based on the threshold value, minimum and the maximum value for each parameter will be specified in the parameter list. An objective function is constructed by means of the objective function of the network transmission parameter [21].

### **3.4 QoS in LTE Network**

In 4G wireless networks, the QoS mainly depends on the CAC, Packet Scheduling (PS) and Bandwidth Allocation (BA) methods because they aim to distribute all the available resources, while keeping the QoS requirements of both real-time and non-real-time applications at an acceptable level.

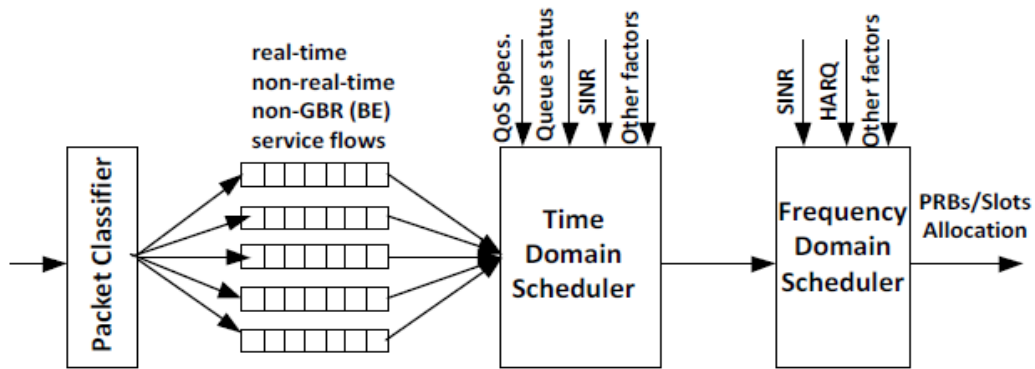
### 3.4.1 QoS-CAC Algorithms

QoS-CAC algorithm admits all connections (handoff and new calls) with QoS guaranteeing within the network. The new connection requesting for the establishment are classified based on the associated services. QoS-CAC performs the operations based on the service requests like RTPS and NRTPS queues. It provides highest priority to UGS connections requests followed by RTPS and NRTPS connection requests. The new call that arrives into a network will be classified into UGS, RTPS and NRTPS connections. The highest priority has been given for the UGS connection that deals with fixed size packets, and QoS CAC performs the RTPS Queue with the variable sized packets. The least priority is given to the NRTPS services that deal with non real time with variable sized packets balancing of traffic load among multiple cells in wireless network allows for a better utilization of the radio resources has to be performed[20].

### 3.4.2 Packet Scheduling Algorithm

In LTE a channel is shared among multiple users using orthogonal frequency division multiple access (OFDMA). It is the packet scheduler which controls, at each scheduling period, to which users the available resources are assigned. The scheduling resources in LTE downlink have a granularity of 180 kHz in the frequency domain, and 1ms (also called a transmission time interval (TTI)) in the time domain. To cope with the variable channel conditions and to optimize the overall data rate, LTE uses adaptive modulation and coding. So as a consequence of the decisions taken by the scheduler and the ability in LTE to adjust the bit rates to the varying channel conditions, the total bit rate at which data can be transmitted every scheduling time varies dynamically over time.

In general, the wireless MAC scheduler is responsible for scheduling the air interface resources among the users in both the downlink and the uplink time periods. Since OFDM technology is used in LTE networks, the scheduler effectively distributes the radio resources (slots/PRBs) in both time and frequency domains. In LTE, most of the existing MAC schedulers' first schedule radio resources in Time Domain (TD), and then Frequency Domain (FD) as shown in Figure (3.4). The TD scheduler is used to differentiate the users according to their QoS characteristics. The FD scheduler is responsible for assigning the radio resources (i.e. PRBs) based on the user's priority and channel condition [20, 24].



**Figure 3.4:** Time and frequency domain scheduling in LTE

### 3.4.3 Adaptive Bandwidth Allocation

Due to the diversity of applications and QoS requirements for the mobile users and the dynamic nature of the wireless channel quality, adaptive bandwidth allocation (ABA) would be necessary to improve the utilization of the wireless network resources. ABA can minimize the number of blocked new calls and the number of dropped handoff calls by adjusting the allocated bandwidth of ongoing calls and allowing the incoming calls to be serviced without degrading the QoS of the ongoing calls below the acceptable level.

Therefore, call admission control strategies should be designed taking this adaptive bandwidth allocation into account. With ABA, when the network conditions are favorable, the quality of a call can be upgraded by assigning more resources. However, when the network becomes congested, the amount of bandwidth allocated to some ongoing calls will be revoked to accommodate more incoming calls so that the call dropping and blocking probabilities can be maintained at the target level.

Again, adaptive bandwidth allocation is needed during vertical handoff. The acceptable bandwidth should be negotiated and the CAC strategy should be based on the result of negotiation. For example, when a call handed over to cellular networks from a WLAN, bandwidth adaptation will be required for that call [4].

QoS degradation can either be bandwidth degradation or delay degradation. In bandwidth degradation method, calls are categorized as adaptive (degradable) and non-adaptive (non-degradable) calls. Degradable calls have flexible QoS requirements (e.g., minimum and maximum data rates). For most multimedia applications, e.g., voice over IP or video conferencing, service can be degraded temporarily as long as it is still within the pre-defined range. Bandwidth degradation reduces handoff call dropping by reducing the bandwidth of the ongoing adaptive calls during network congestion. When a handoff call arrives and there is network congestion, the system is able to free some radio resource to admit the handoff calls by degrading some of the ongoing adaptive calls.

In delay degradation method, the amount of radio resources allocated to non-real-time (delay-tolerant) services is reduced during network congestion. When a handoff call arrives and there is no radio resource to accommodate the handoff call. Some non-real-time services are degraded

to free some bandwidth, which is used to accommodate the incoming handoff call [23].

#### **3.4.4 Bandwidth Reservation Schemes**

In practical scenarios, the CAC at the eNodeB calculates the required bandwidth in terms of resource elements (Physical Resource Blocks PRBs) and checks whether or not the system has enough resource elements to admit the call. However, the UE in LTE network starts random access procedure with minimum power and then increases the power level to the required signal strength. Similarly, when the user is moving at different speeds, the current channel quality may change instantly or in the next upcoming period. Therefore, the instantaneous resource elements requirement for total calls is time varying. Hence, it is not necessary to differentiate CAC based on channel quality and bandwidth measurements. Thus, reservation based CACs are more suitable for 4G wireless networks because the major requirement is to provide QoS assurance for the existing calls [1].

Existing CAC and bandwidth reservation schemes can be classified into static reservation and dynamic reservation schemes. Static schemes always endeavor to maintain a fixed amount of cell bandwidth for handoff calls; this amount, for example, can be specified as a fixed fraction of total cell bandwidth. In dynamic schemes, a variable amount of cell bandwidth is reserved for handoff calls based on the measurement information from either the local cell, the cells in the vicinity, or a combination of both. The measurement information may include, for example, the user mobility patterns and the current cell CDP (call dropping probability) and CBP (call blocking probability) information. The aim for obtaining user mobility patterns is to predict the future movement of a mobile terminal since

accurate mobility prediction can lead to highly effective bandwidth reservation. However, predicting user mobility can be complex as well as difficult. Since a handoff drop occurs mainly when a cell is overloaded, measuring the cell load may be simpler and more efficient [22].

To maintain QoS of the existing calls and to minimize CDP, the CAC in the eNB reserves some bandwidth for the mobile and high priority users and changes the bandwidth reservation adaptively based on the most recent requests from handoff and high priority users. Suppose, the reserved bandwidth is not fully utilized by handoff and high priority users, the remaining reserved bandwidth is then allocated for least priority (BE/non-GBR) users for effective bandwidth utilization. Later, when a high priority or handoff user arrives, the CAC applies bandwidth pre-emption on least priority calls to admit that call. However, while admitting new calls or handoff calls [20].