# Quality of Service in 3G Wireless Networks

Soshant Bali and John Korah
Department of Electrical and Computer Engineering,
Virginia Polytechnic Institute and State University.

## Abstract

*The main difficulty in providing multimedia services in 3G wireless systems is due to the mobility of hosts and the unpredictable nature of the radio link. Providing an uninterrupted service while moving from one cell to another is also a challenge. A number of resource allocation and scheduling algorithms have been suggested. This paper is an overview of the most important ones. The first part is a short introduction to the general network architecture, protocol stack of radio link and QoS service classes. Algorithms for fast, efficient handover are compared. The next section deals with the working of higher layer protocols like Mobile IP and Mobile RSVP. The report ends with a description of a resource-scheduling algorithm for WCDMA.*

## 1. Introduction

Looking into the recent years, two main drivers for the telecommunications market can be identified: Internet and Portable cellular telephony. 3G Network enables the integration of these services. It provides high-speed wireless Internet access to mobile users. The same network can be used to make voice calls using conventional mobile phones. Moreover 3G Networks are seen as the enabler of multimedia services like videoconferencing, streaming audio and video etc. These Networks must provide services to many different types of applications without letting data from one application affect the services provided to other types of applications. In other words, 3G Networks must provide support for Quality of Service.

## 2. Network Architecture (Author: Soshant Bali)

3G Partnership Project (3GPP) refers to 3G mobile systems as Universal Mobile Telecommunication System (UMTS) [1], [2]. UMTS Network can be partitioned into two basic components: UMTS Core and UMTS Terrestrial Radio Access Network (UTRAN). UMTS Core is an IP based wired network similar to the Internet. It provides Quality of Service (QoS) based on the Differentiated Services Architecture. A service provider can choose to use "ATM-SVC's as an internal dedicated transport bearer. In that case inter operation with IP networks is based on differentiated services" [3]. UTRAN is the air interface between Mobile Stations (MS) and the Core. It defines protocols and physical layer technologies used here.

## 3. UTRAN Protocol Stack (Author: Soshant Bali)

UTRAN uses a layered structure to support different data services [4]. The lowest layer in the protocol stack is the physical layer. Media Access Control (MAC) layer is built on top of the physical layer. MAC can perform priority handling between different data flows. On the top of MAC is the Radio Link Control (RLC) layer. It uses Automatic Repeat Request (ARQ) to ensure optimum utilization of radio resources. This feature is needed to overcome the problems in using TCP directly over wireless links [5]. On the top of RLC is the Radio Resource Control (RRC) layer, which performs resource reservations for applications (establishment/release of radio access bearer).

## 4. UMTS QoS Classes (Author: Soshant Bali)

Different QoS classes supported in UMTS are Conversational, Streaming, Interactive and Background. Conversational class is the most delay sensitive class. Applications like telephony and videoconferencing use this class. Such applications have stringent delay requirements. End to end delays suffered by packets belonging to this class should not exceed 150ms. Applications using the Streaming class have less stringent delay bounds. They can tolerate larger delays because of their non-interactive nature. For this reason, the network is allowed to delay streaming class packets.

Two non real time classes supported by UMTS are Interactive and Background. Interactive applications like Web browsing, Telnet etc. use the Interactive class. Downloads and emails use the background class. The idea behind having two non real time classes is to give priority to the data for which a user is waiting and cannot proceed till it arrives.

## 5. Problems in Providing Quality of Service (Author: Soshant Bali)

A well-known problem in providing Quality of Service in 3G Networks occurs during handoffs. When a user wants to make a new call, he sends a resource allocation request to the base station. The base station checks if it has enough resources to support the new call. If resources are available, the call is admitted, otherwise it is "blocked". Admitted Mobile Stations keep the resources until the call is completed or the user moves out of the cell. When a call gets completed in the cell, resources are released and become available to serve another caller. If a user crosses cell boundaries while the call is still in progress, the MS sends a new resource reservation request to the target cell's base station. The base station checks if resources are available to support the handoff call. If resources are available, the call is accepted otherwise it is "dropped". "Dropped" calls are more objectionable than "Blocked" calls. There are ways by which one can decrease the dropping probability at the expense of an increased blocking probability. One way of doing this is to reserve a fraction of the total resources for handoff calls. Resources not reserved for handoff calls are available for both new calls and handoff calls. New calls cannot use the resources reserved for handoff calls. This decreases the handoff call dropping probability as a portion of the resources in every cell are reserved for such calls. More the amount of resources reserved, lesser is the dropping probability. But this comes at a price. New call blocking probability increases with an increase in amount of reservation. Therefore it is important to find the right amount of resources that should be reserved. If too many resources are reserved, a large fraction of new calls will get blocked. If on the other hand too little resources are reserved then a large fraction of handoff calls will get dropped. The optimum amount is usually a function of the traffic load. Adaptive admission control algorithm determines this optimum value based on the number of active mobile stations in the neighboring cells. If this number is large then it is expected that many users will cross cell boundaries and request resources. Large amount of resources are reserved in anticipation of handoff calls. If on the other hand, there are a very few active mobile stations in the neighboring cells, a small number of users are expected to cross over. For this reason, very small amount of resources are reserved for handoff calls.

This is also not the best way to reserve resources. There might be a large number of stationary users in the neighboring cells and as stationary users never cross-cell boundaries, it will be wasteful to reserve resources for them. Early resource reservation scheme collects location information like position of Mobile Station, direction of movement, velocity etc, and reserves resources only for users heading towards the target cell. This avoids over booking of resources and improves bandwidth utilization.

A semi reservation scheme seems to produce even better results than the Early Resource Reservation scheme [6]. It determines the probability that a user will cross over to the neighboring cell and then performs semi reservation based on it. To determine this probability, it divides the users into three basic classes: stationary, walking and vehicular. Probability of handover is different for each of the three classes. It is computed based on the past history of handovers, for e.g. if a large fraction of the most recent vehicular handovers were to cell 5, then there is a high probability that the next handover will also be to the same cell. This scheme produces better results than any of the schemes discussed earlier and is a good contender for being included as a 3G standard.

## 6. Mobile IP [8] (Author: John Korah)

The third generation wireless envisions the transfer of data and voice traffic in a packet switched fashion. In the Internet, each host has an IP address using which, data is routed towards it. Since the hosts are stationary, the physical location of the host does not change. Hence the routing tables (refreshed in relatively large intervals of time) in routers do a fine job, guiding data to the destination address. But in mobile systems, nodes keep changing their location. It is inefficient to change the routing tables every time the mobile node moves. A protocol called Mobile IP has been proposed by IETF as a solution.

A home network of a Mobile Terminal (MT) is its 'parent network'. The IP address of the MT is registered with the home agent (HA), which is a node or a router in the home network. When the MT moves to a new network, the network becomes the foreign network with the corresponding home agent becoming the foreign agent for the MT. The MT now registers its IP with the foreign node. Let us examine a simple fixed node to MT data transfer. The mobility agent in the network announces its presence by sending out advertisements messages. The advertisement consists of its IP address. The MT receives this message and determines if it is in the home or foreign network. If it is in the foreign network, the IP address of the mobility node is taken as the care-of-address. Alternatively, the care-of –address can be obtained by DHCP. The MT then registers its care-of address with the home agent by sending a registration request message via the foreign agent. If the registration is accepted, a registration reply is send back. The home agent makes a mobility binding for the MT, which consists of the original IP address, care-of-address and a time-to-live (TTL) value. This binding lasts only for TTL and need to be refreshed periodically. When the data is send by the fixed node, it is routed to the home network, intercepted by the home agent and tunneled to the foreign network using the care-of-address, either directly to the MT or through the foreign node. The return data transfer is done by normal shortest path routing and does not retrace the path taken by the message from the fixed node.

## 7. Mobile RSVP [7] (Author: John Korah)

Mobile IP has solved the problem of routing data to the Mobile Node through the Internet. The protocol is suitable for best effort traffic. But 3 G wireless systems are expected to provide multimedia and other real time, delay constraint services. Hence reservation of resources has to be carried on. RSVP is the popular protocol for resource reservation on the Internet. But RSVP lacks the ability to deal with the mobility factor. Hence an enhancement of RSVP which works on Mobile IP has been proposed by Mahmoodian and Haring. The main aspect of this protocol is that the base station (BS) supporting the Mobile Terminal is responsible for making sure that the QOS guarantees are maintained when the MT moves to a nearby cell. This it does by passively reserving resources in all the neighboring cells.

The various steps in M-RSVP operations are:

**1. Connection Setup**: A simple duplex transfer between a fixed node (FN) and MT is considered. The care-of –address of the MT is got by the FN from the Home Agent (HA). The FN then send a PATH message to the MT at its care-of –address. The PATH message is similar that of RSVP and contain the FLOW SPECS. It initiates the QoS states in all the routers through which it passes. The receiver then sends back the RESV message through the same path, which sets up the reservation in each of the router.

**2. Passive Reservation**: To ensure that the QoS guarantees are maintained when it moves to a neighboring cell, the BS sends advanced passive reservation PATH messages to all the neighboring cells. Some authors like Talukdar and Badrinath have suggested a method in which the mobility of the cells are predicted and passive reservation done only in select cells. But here non-deterministic mobility is assumed and reservations done in all the cells. When the resources are passively reserved, it means that the reserving agent would receive the first priority when it actually needs the resources. In the meantime it may be used by other traffic, mostly best effort traffic. The reservation has a TTL and need to be periodically refreshed.

**3.Hand-over procedure**: When the MT is 'handoffed' to a neighboring cell, the QoS information is relayed to the new BS. Also the passively reserved path to the cell is activated by the old BS using the ACTIVATE message. The other reservations are cancelled by using the RELEASE message or allowed to time out. The new BS at once passively reserves resources in the neighboring cells. The old BS stores the new care-of –address of the MT so that packets transmitted during the handover are not dropped. When the fixed node is informed of the new mobility binding by the HA, it tries to reserve resources along a new path to the present location of MT.

**Dynamic Resource Sharing**

A resource-sharing algorithm has to be developed to regulate the passive reservations. There are three classes of traffic-best effort, guaranteed and active hand offs in the increasing order of priority. Total capacity is shared by all the classes but when a higher priority call comes up, the lower priority calls are dropped. Resources are passively reserved only when substantial best effort traffic is present (which can be dropped when a passive reservation is activated).

**Advantages**

1. There is very less chance of an active call being dropped.
2. Handoff can be done in a smooth manner.

**Disadvantages**

1. There is a lot of signaling and protocol overhead involved.
2. The passively reserved resources (before it is activated) cannot be used by high priority data services. Hence there are increased chances for the blocking of new calls in the cell.

## **8. Dynamic Resource Scheduling (DRS)** (Author: John Korah)

One of the differences between wired and wireless networks is the presence of radio links in wireless networks. And this introduces one of the complexities in QoS implementation. Managing radio resources is difficult because it continuously changes due to fading, interference and path loss. The DRS algorithm discussed below has been proposed for WCDMA by Gurbuz and Owen.

The QoS parameters for any service are defined in terms of delay and bandwidth constraints. In WCDMA, each MT transmits with a particular spread code. The smaller the spread code length, the larger the bandwidth. Delay constraints can be translated in terms of Bit Error rate (BER). Smaller the BER, smaller is the delay. The BER is in turn related to the power of transmission (expressed as Power Index in the algorithm). The greater the power, lesser is the BER. To be noted is the fact that total transmission power in a cell is limited and a constant. Hence by allocating power efficiently, more calls (which includes data services can be accommodated) can be allocated. Hence spreading codes have to be allocated carefully.

There are two types of DRS:

**1. Fixed Spreading Code DRS**

**2. Variable Spreading Code DRS**

The resource scheduling is done in two steps:

**1. Time scheduling**: Resource scheduling for each radio frame.[9] When an application on the MT needs QoS, it translates the traffic descriptors such as delay and bandwidth to WCDMA layer parameters such as Signal to Interference ratio and symbol rate respectively. It is then relayed to the base station using the control channel RACH (Random Access Channel). The requests are put in the queue according to their delay constraint $\gamma_i$.[11] Hence the request with the most stringent delay constraint is at the head of the queue because it would require the most resources (in terms of power). There is a separate queue for UBR (best effort traffic) as shown in fig1.

**2. Power Scheduling**

1.Each service request is allotted a stream code according to its bandwidth requirements.

2.Then the power index $g_i$ is given by the formula.

$$g_i = \frac{\gamma_i}{\gamma_i + G_i}$$

Gi – spreading factor

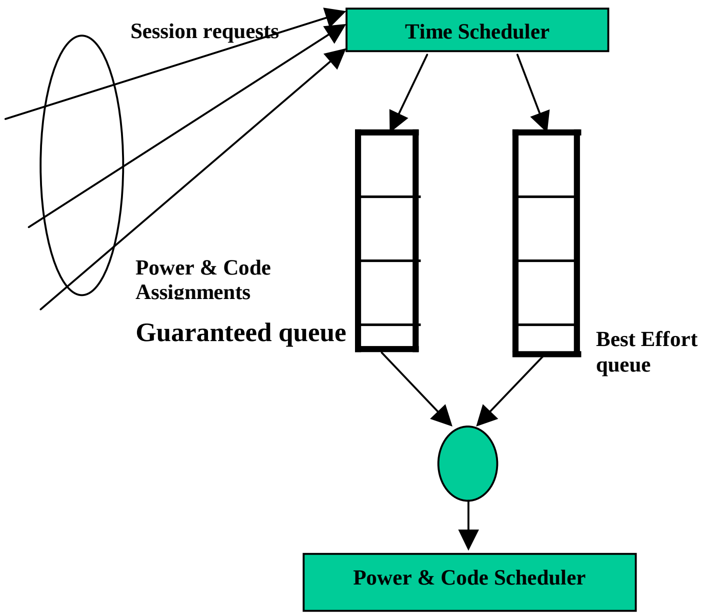


**Fig 1:** Base Station Resource Scheduler

3.Then the Connection Admission Connection is applied

$$\Sigma g_i < 1 - \frac{\eta_0 . W}{Min_i(P_i . h_i / g_i)}$$

$\eta_o$ – power spectral density of Gaussian white noise
$h_i$ – path loss

4.If the CAC is applied, then the service request is accepted.
5. The transmission power is calculated as

$$Pi = \frac{gi.\eta_o.W}{hi.(1 - \sum_i g_i)}$$

$h_i$ – path loss
W- total system bandwidth

In Fixed DRS, the spread code that is fixed during connection setup remains constant. This is inefficient in the case of VBR service where the bandwidth and BER changes dynamically. Hence Variable DRS is applied. This is similar to Fixed DRS except that the spread code allotment is done every time the traffic descriptors change. Hence Bandwidth is used efficiently. The excess bandwidth can be used for background traffic like web browsing.

## 9. Conclusion

The lossy and unpredictable nature of wireless links makes providing Quality of Service a very challenging task. In this paper we presented some of these problems and discussed the solutions that are proposed in 3G forums. The biggest hurdle in maintaining QoS occurs during handoffs. We prsented some algorithms and discussed in detail the merits of each when compared with the others. We also discussed 3G's answer to the TCP over wireless link problem: reliabile link layer. We also discussed Mobile RSVP and some radio resource scheduling algorithms.

## 10. References


1. 3G TR 23.821 v1.0.1 "Architecture Principles for Release 2000", 2000-07.
2. Lieve Bos and Suresh Leroy, "Towards an all IP based UMTS Architecture", IEEE Network, Jan/Feb 2000.
3. Vijay K. Garg and Oliver T. W. Yu, "Integrated QoS Support in 3G UMTS Networks", Wireless Communications and Networking Conference, IEEE, Volume: 3, 2000.
4. Vijay K. Garg and Mirdu Garg, "Role of MAC/RLC in achieving Higher-data Rates and QoS in Third Generation (3G) UMTS System", IEEE International Conference on Personal Wireless Communications, 2000.
5. Hari Balakrishnan, Venkata N. Padmanabhan, Srinivasan Seshan and Randy H. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links", IEEE/ACM Transactions on Networking, December 1997
6. Geng-Sheng Kuo, Po-Chang Ko and Min-Lian Kuo, "A Probabilistic Resource Estimation and Semi-Reservation Scheme for Flow-Oriented Multimedia Wireless Networks", IEEE Communications Magazine, Feb. 2001.
7. Ali Mahmoodian and Gunter Haring, "Mobile RSVP with Dynamic Resource Sharing", Wireless Communications and Networking Conference 2000 IEEE, Volume: 2, 2000 Page(s): 896 –901.
8. Charles E. Perkins, "Mobile IP", IEEE Communications Magazine, Volume: 35 Issue: 5, May 1997.
9. Ozgur Gurbuz, Henry Owen, "Dynamic Resource Scheduling for Variable QoS Traffic in W-CDMA", IEEE International Conference on Communications, 1999.
10. Ozgur Gurbuz, Henry Owen, "Dynamic Resource Scheduling Strategies for QoS in WCDMA", Global Telecommunications Conference, 1999.
11. Ozgur Gurbuz, Henry Owen, "A Resource Management Framework For QoS Provisioning in W-CDMA Systems", IEEE 49th Vehicular Technology Conference, 1999.