CHAPTER 6

MAC LAYER PROTOCOLS

6.1 INTRODUCTION

Medium access control layer which deals with the allocation of radio channel resources, plays a very important role for QoS guarantees in WATM networks [128]. Three MAC layer protocols, namely, Combined Hybrid ARQ and Adaptive scheduling, RAPP power control algorithm and Distributed call admission control are presented in this chapter. These protocols enhance the overall quality of WATM networks.

It is desired that a MAC protocol design should be coupled with an error control method. The present work describes an improved combined hybrid ARQ error control and adaptive scheduling scheme for time division multiple access/time division duplex (TDMA/TDD) MAC protocols in WATM networks. The improved algorithm has more CRC bytes than the earlier one.

The main challenge in wireless communication is to overcome the disadvantage of the hostile characteristics of the radio channel. Specifically, power control is one of the major concerns during retransmission in wireless networks. The present work explains a version of Retransmission Algorithm based on Power Priorities (RAPP) for power control in WATM networks. The algorithm is implemented in a Rayleigh fading environment.

In general, admission decisions in WATM networks are made to ensure guaranteed QoS for heterogeneous traffic [129]. In the present work, the improved DCAC scheme is described for WATM networks in which resource reservation is dynamically allocated according to the user mobility information to meet the QoS
requirements for heterogeneous traffic. Along with blocking and dropping probabilities, admission delay is also analysed for the improved scheme in a Rayleigh fading environment.

6.2 COMBINED HYBRID ARQ AND ADAPTIVE SCHEDULING

A combined hybrid automatic repeat request error control and adaptive scheduling scheme is proposed here for time-division multiple access/time-division duplex medium access control protocols in WATM networks. This scheme is appropriate for supporting real-time traffic with strict QoS requirements. As the channel BER in wireless environment is very high and varies from $10^{-2}$ to $10^{-5}$, existing schemes have been using FEC to overcome the noisy channel condition at the cost of valuable bandwidth. The proposed scheme guarantees the throughput requested by a real-time traffic user while keeping the bandwidth consumption at a minimum for a TDMA system.

6.2.1 System Model

The proposed scheme is based on a finite-state Markov channel model. In this model, only the transitions of two adjacent states are allowed. The switching process between states is described by the transition probability matrix.

![Figure 6.1 Two-state Markov model](image)

Figure 6.1 shows the two-state Markov channel model. In this model, the channel state varies between two states namely, '0' (bad) and '1' (good), which could be further referred to as fading and non fading states respectively. The transition matrix of the model can be given as,
The punctured convolutional codes are used with Viterbi algorithm to enable adaptive encoding and decoding without modifying the basic structure of the encoder and the decoder. The error correction is performed by shortened cyclic codes with variable degrees of shortening. When the channel bit error rate increases, the system generates additional parity bits for error control [130].

The adaptive error control algorithm is illustrated by the block diagram in Figure 6.2. Information is generated by the data source and encoded using one of the M Punctured Convolutional coders. An interleaver is used to randomize the burst error. The forward channel is composed of the interleaver and the modulator. In order to implement the adaptive coding scheme, it is necessary to use a return channel. It is assumed that the return channel is error-free.

![Figure 6.2 Adaptive Error Control](image)

The Channel State Estimator (CSE) determines the current channel state based on counting the number of erroneous blocks. Once the channel state has been estimated, a decision is to be made by CSE to change the code. Such changes and the corresponding messages are to be sent to the encoder and the decoder. In FEC schemes, only error correction is performed, while in hybrid ARQ schemes, retransmission of erroneous blocks is requested whenever the decoded data is labeled as unreliable.
6.2.2 Medium Access Control Protocol in WATM

A MAC protocol is a set of rules to control the access of a shared wireless communication medium among various users. These protocols need to provide access to the users moving within the cell and handoff calls. At the same time, they must have features to include ATM service standards. Multiple access schemes and protocols can be static or dynamic. In general, multiple access schemes can be classified into three main categories namely, Fixed assignment, Random access and Demand assignment.

Unlike fixed assignment, demand assignment schemes avoid wastage of bandwidth by assigning it on demand. Demand assignment protocols also avoid bandwidth wasted due to collision by providing the connections with free contention bandwidth during active periods unlike random assignment. Because of these features, MAC protocols based on demand assignment are most appropriate for the requirements of integrated wireless networks.

Demand assignment users are required to provide information regarding their needs for bandwidth which will be assigned on demand. It is widely accepted that the dominant services in the broadband environment are VBR services. Therefore for VBR users, bandwidth will be assigned according to requirement. Whenever the VBR user enters an idle period, the assigned bandwidth will be allocated to another user [131].

In demand assignment protocols, the channel bandwidth is time-slotted and represented by a single or several block(s), depending on the channel rate and the type of applications. Figure 6.3 shows the radio channel bandwidth allocation. Each block is divided into uplink and downlink periods (channels), each of which is further divided into sub periods. This type of classification might be performed on a slot-by-slot or period basis. Downlink traffic can be transmitted in a separate channel (using a different frequency band), or it can time share a single block with the uplink traffic by using time-division duplex (TDD). The latter provides better flexibility in controlling the available bandwidth by dynamically allocating the length of each period.
The uplink period is a mobile unit-to-BS transmission, subdivided into request access (RA) and data transmission access (TA) sub periods. In the RA sub period, users transmit their RA packets to the BS for requesting bandwidth. The BS will identify the successful users (those not involved in collision) and assign them a bandwidth, if available.

The time taken for the transmission from BS to mobile unit is the downlink period. The downlink period is subdivided into acknowledgment (ACK) and data downstream (DD) sub periods. Basically, in the ACK sub period, the BS notifies the RA terminals about their request status. In the DD sub period, the BS transmits downstream data to the destined terminals. Downlink transmissions are governed by the BS and generally supported by a contention-free time-division multiplexing (TDM) broadcast mode. Since a TDM channel can be operated with high efficiency and low delay, the downlink channel is not a critical performance driver of the system.
As explained, the uplink period is subdivided into RA and TA sub periods. It is to be noted that the number of terminals under BS coverage is considered to be much larger than the available channel bandwidth. However, not all the terminals are active simultaneously. In this case, the MAC scheme used in the RA sub period is random.

![Diagram of Scheduler](image)

**Figure 6.4 Scheduler**

In general, demand assignment protocols provide better channel utilization than fixed assignment and random assignment schemes. Because of the frequent bandwidth assignment, there will be an overhead due to channel access delay and control signal exchanges, which might lead to inefficiency. The channel access delay is classified into connection and packet access delay. The channel access delay is defined as the time required for the RA and TA packets to access the channel. The connection access delay is experienced in the RA sub channel and will have a significant impact on the retransmission requests. The packet access delay can be defined as the time required for active VBR, ABR, and UBR applications to resume transmission after getting back from an idle period. The packet access delay is experienced in the TA sub channel. This problem has a great impact on the queue length at the terminal premises. For real-time active VBR services, packets will be stored at the output queue at the user terminals until they get another reservation for transmission after getting back from an idle period. Therefore, the output queue might go off before the bandwidth reservation is granted.
For non-real-time active ABR services, data might be stored at the higher layers; therefore, packet access delay would not affect the terminal's output queues.

The proposed MAC protocol operates based on time slots. When a user is in the noisy channel, the system assigns an extra slot for that user, which is twice as many slots as needed for the user in the quiet channel. The extra slot is assigned to provide QoS guarantees for the user in the noisy environment. Considering the small size of a cell covered by a base station, the propagation delay between the base station and a mobile user can be assumed as being negligible.

**UPSTREAM STRUCTURE**

<table>
<thead>
<tr>
<th>Signaling slots</th>
<th>Queue status slots used for feedback</th>
<th>M message slots</th>
</tr>
</thead>
</table>

**DOWNSTREAM STRUCTURE**

<table>
<thead>
<tr>
<th>M message slots</th>
<th>Allocation slots</th>
<th>Signaling slots</th>
</tr>
</thead>
</table>

Figure 6.5 Upstream and Downstream Structures

The scheduler shown in Figure 6.4 has complete information about channel state information (CSI). Figure 6.5 shows the general structure for upstream and downstream. The upstream structure consists of signaling mini slots for the channel access that uses the slotted ALOHA protocol. Active mobile users use queue status mini slots to feedback their queue information to the base station. Based on the information of the queue status and the channel state, the scheduling algorithm allocates the appropriate number of slots to guarantee the QoS requested by active users. The number of queue
status mini slots is equal to the maximum number of mobile users that can be served within a block. The message slots are used for the information transmission. The scheduler needs to reserve certain number of slots for active users possibly in the noisy state. The reserved slots are used for users in the noisy channel to meet their QoS requirements. If some slots are still left unassigned, they will be allocated to new calls.

6.2.3 Performance Analysis of Combined Hybrid ARQ and Adaptive Scheduling

The main function of a hybrid ARQ protocol is to properly combine ARQ and FEC techniques. There are two types of hybrid ARQ schemes namely, type I and type II hybrid ARQ. In type I hybrid ARQ, each packet is encoded for both error-detection and error-correction capabilities. Retransmission is initiated once the error correction fails as indicated by the error detection to recover all the errors incurred. In type-II hybrid-ARQ protocol, a code word is encoded using two different codes, \(C_1\) and \(C_2\). These codes have corresponding decoding operations \(D_1\) and \(D_2\) which can recover the original code word from noise corrupted versions of \(C_1\) and \(C_2\) respectively. \(C_1\) is initially transmitted while \(C_2\) is set aside. Upon receiving \(C_1\), the receiver attempts decoding operation \(D_1\) to recover the transmitted data. If the attempt is successful, the receiver sends an acknowledgment (ACK) to the transmitter; otherwise, a negative acknowledgment (NACK) is sent. The transmitter responds to the NACK by sending \(C_2\). The receiver then attempts to decode \(C_2\), using decoding operation \(D_2\). If successful, the receiver finds the corrected version of \(C_2\) to recover the desired information and sends an ACK to the receiver. If unsuccessful, the receiver combines \(C_1\) and \(C_2\) to create \(C_3\), a new code word with lower rate. If the third decoding operation \(D_3\) is successful, the data is recovered and an ACK is sent to the receiver; otherwise, a NACK is sent and the entire process is repeated.

In this work, a punctured RS code with a finite Galois Field \(GF(2^8)\) is considered. Specifically, a type II hybrid ARQ has been adopted. The throughput of type I and type II hybrid ARQs with the developed two-state Markov channel model are estimated. Infinite transmitter and receiver buffers and noiseless feedback channel with selective-repeat protocol have been assumed in the ARQ retransmission.
The overall throughput for the two-state Markov model shown in Figure 6.1 can be expressed as,

\[ \eta = p_0 \eta_0 + p_1 \eta_1 \]  

(6.2)

where, \( p_i \) for \( i=0,1 \) and \( \eta_i \) for \( i=0,1 \) are the steady state probability and throughput of state \( i \) respectively.

The throughputs \( \eta_i \) for \( i=0,1 \) of type II hybrid ARQ can be analyzed first. Let \( T_{si} \) denote the total number of transmissions (including the original transmission) required for a packet to be decoded successfully in state \( i(i=0,1) \). Then the throughput \( \eta_i \) for \( i=0,1 \) can be expressed as,

\[ \eta_i = \frac{1}{E[T_{si}]} \frac{k}{n} \]  

(6.3)

where, \( E[T_{si}] \) is the average number of transmissions and \( k/n \) is the coding rate of codes \( C_1 \) and \( C_2 \). Figure 6.6 shows a state diagram for the behavior of a type II hybrid ARQ decoder. In this figure, \( D_1, D_2 \) and \( D_3 \) denote the decoding operations for codes \( C_1, C_2 \) and \( C_3 \) respectively and \( P_R^{(1)}, P_R^{(2)}, P_R^{(3)} \) are the probabilities of the generation of a retransmission request by \( D_1, D_2 \) and \( D_3 \) respectively. Now \( E[T_{si}] \) may be computed as follows:

\[ E[T_{si}] = \frac{1 + P_R^{(1)} + P_R^{(2)} + P_R^{(3)} - P_R^{(1)} P_R^{(2)} P_R^{(3)}}{1 - P_R^{(1)} P_R^{(2)} P_R^{(3)}} \]  

(6.4)

\[ \eta_i = \frac{k}{n} \frac{1 - P_R^{(1)} P_R^{(2)} P_R^{(3)}}{1 + P_R^{(1)} + P_R^{(2)} + P_R^{(3)} - P_R^{(1)} P_R^{(2)} P_R^{(3)}} \]  

(6.5)

where,

\[ P_R^{(1)} = P_R^{(2)} = \sum_{j=\lceil n/2 \rceil + 1}^{n} \binom{n}{j} P_{si}^j (1 - P_{si})^{n-j} \]  

(6.6)

\[ P_R^{(3)} = \sum_{j=\lceil n/2 \rceil + 1}^{2n} \binom{n}{j} P_{si}^j (1 - P_{si})^{2n-j} \]  

(6.7)
Then, \( \eta_i = \frac{1}{E[T_{sl}]} \cdot \frac{k}{n} = (1 - P_{rt}^{(t)}) \cdot \frac{k}{n} \quad i = 0, 1 \) (6.8)

As the characteristics of the wireless channel vary with time, it is desirable that the error control method and the scheduling algorithm be adaptive according to the varying channel conditions.

The scheduling scheme in the WATM MAC protocol must allow isolation and integration of the different service categories. Fair-queuing or General Processor Sharing (GPS) strategies, especially Self Clocked Fair-Queuing (SCFQ), have to provide such properties. The basic idea of SCFQ scheduling is that packets are served based on their service tags which are assigned to packets in the head-of-line position of a queue.

In particular, the service tags can be calculated by the following equation:

\[
F_i = \frac{L_i}{q_i} + \max (F_i', v(a_i))
\]

where,

- \( F_i \) is the service tag of the \( i^{th} \) packet
- \( L_i \) is the length of the \( i^{th} \) packet
- \( a_i \) is the arrival time of the \( i^{th} \) packet
- \( v(t) \) is the service tag of the packet in service at time \( t \);
- \( F_i' \) is the weight of connection \( i \).

In wireless environments, the bandwidth requirement of a real-time connection varies according to the channel condition. In WATM networks, adaptive SCFQ is particularly desirable for the case of bandwidth assignment of retransmission. In this work, SCFQ scheduling with adaptive weights for a wireless ATM MAC protocol is presented. Specifically, the service weight of each connection dynamically changes according to the channel condition as well as its QoS requirements.
It has been shown that GPS is able to make worst case performance guarantees on throughput and delay by the use of leaky bucket rate control. In particular, suppose there are $Z$ connections sharing one link with the capacity of $C_h$ and that connection ‘$i$’ is assigned with a service weight. Then, the given connection ‘$i$’ guaranteed with a service rate is given by,

$$g_i = \frac{\gamma_i C_h}{\sum_{j=1}^{i} \gamma_j} \geq \gamma_i$$  \hspace{1cm} (6.10)

Accordingly, the service weight for a constant bit rate connection ($\gamma$) can easily be determined by its peak cell rate. For an available bit rate connection, its service weight could be equal to minimum cell rate. Likewise, in order to provide at least a minimum QoS to unspecified bit rate (UBR) connections, some bandwidth may be reserved but with a small service rate. On the other hand, the weight determination of a variable bit rate (VBR) connection is not straightforward. Suppose a VBR connection is leaky bucket policed then its service weight should be associated with its traffic.
parameters, such as peak cell rate, sustained cell rate, burst tolerance, maximum transfer delay, etc. For example, the VBR source is characterized by \((m, B_s, B_p, D_p)\), where \(m\) is the maximum burstiness of the source, \(B_s\) is the average cell rate of the source, \(B_p\) is the peak cell rate of the source and \(D_p\) the maximum tolerable delay of the source. The regulation of the leaky bucket is characterized by \((\Phi, \rho)\), where \(\Phi\) is the token buffer size and \(\rho\) is the token generation rate. WATM cells that have passed the leaky bucket are stored in the ready to transmit (RTT) buffer and will be served at the prescribed rate. Then the delay of a packet is found to be the difference between the time it reaches the other side over the wireless channel and the time it arrives at the RTT buffer.

The maximum burst size of a leaky bucket policed VBR source is given by,

\[
m_i = \frac{\Phi}{B_p - \rho} \cdot B_p
\]

Then, under the assumption that \(g_i \geq \rho \geq B_s\), the maximum queue length of the RTT buffer is,

\[
L^* = \frac{m_i (B_p - g_i)}{B_p}
\]

Moreover, if the request and reservation delay are ignored, then the maximum delay is given by,

\[
D_p = \frac{L^*}{g_i} = \frac{m_i}{B_p g_i} (B_p - g_i)
\]

Once this delay requirement is met, the guaranteed service rate can be determined as,

\[
g_i = \frac{B_p m_i}{B_p D_p + m_i}
\]

Finally, the service weight of the VBR source can be obtained as,

\[
\chi = \max \left( \frac{B_p m_i}{B_p D_p + m_i}, B_p \right)
\]
In a non-fading duration, each packet most likely needs only one transmission. But more transmissions might be necessary in the case of fading duration. As a result, the service weight of real-time connection should be increased by a factor of two during fading. On the other hand, the service weight of a non-real-time connection during fading will be reduced to one-half or one-fourth (or even less) of the one without fading. Specifically, the combination of adaptive fair-queuing and hybrid type II ARQ scheme for a connection can be illustrated by the flowchart shown in Figure 6.7, where the channel state can be determined by whether it can decode the information bytes correctly from the received packet by \( C_1 \) or \( C_2 \) code. Simulations presented in the next section will show that this scheme is quite reasonable in terms of improved throughput and delay performance of the overall system.

In simulation, the channel capacity is assumed as 1 Mb/s, and a variable block length is employed with maximum number of 20 data slots, including the block header and mini slots. The transceiver turn around time is ignored but the guard time and physical layer preamble are taken into account. The Rayleigh fading channel is considered for simulation. Infinite transmitter and receiver buffers and noiseless feedback channel are assumed. For service traffic, it is assumed that six real-time VBR (rtVBR) and six UBR connections exist in the system. The rtVBR and UBR connections are modeled as on–off and Poisson sources, respectively. Accordingly, the traffic load of the system is about 70% (18% rtVBR traffic and 52% UBR traffic).

Figure 6.8 shows the system throughput of different retransmission and scheduling schemes in terms of different coding rate (\( R=k/n \)) of \( C_1 \) code. Here the WATM packet size is 55 bytes.

It is observed that the system reaches the maximum throughput when \( R=1 \) (\( k=55, n=55 \)) for both type I and type II hybrid ARQ. More importantly, the combined Hybrid type II ARQ and adaptive scheduling yields the best performance especially with low coding rate.
BS performs traffic scheduling based on current weight of each connection

BS sends allocation information and ACKs of uplink traffic to MTs

Set channel state flag = good
Weight of VBR connection is divided by 2
Weight of UBR connection is multiplied by 4

BS transmits downlink traffic based on hybrid type-II ARQ

Each mobile transmits packets based on type-II hybrid ARQ and ACK for downlink traffic

BS receives the packets sent by MTs and the ACKs for the downlink traffic

Y

Channel state flag = 'bad'?

Y

Packets are decoded correctly by CI or C2?

N

N

N

Set channel state flag = bad
Weight of VBR connection is multiplied by 2
Weight of UBR connection is divided by 4

Figure 6.7 Flowchart for the Hybrid Type II ARQ and Adaptive Scheduling scheme
Figure 6.8 Throughput of different retransmission schemes

Figure 6.9 shows the average delay of UBR traffic with respect to different coding rates. Simulation results indicate that the average delay of UBR traffic for the case of type I hybrid ARQ is minimized when the coding rate $R=0.85$, while the delays for the case of type II hybrid ARQ and type II hybrid ARQ with adaptive fair-queuing are both minimized when the coding rate $R=1$. This clearly shows that type I hybrid ARQ always has a larger delay.

Figure 6.10 shows the mean delay of UBR traffic as a function of traffic load which is increased by the increase of the mean rate of UBR traffic. It can be observed that the UBR traffic has the largest mean delay in the case of type I ARQ. The mean delay of UBR traffic can be significantly reduced by the use of type II hybrid ARQ and further reduced by using the combination of type II hybrid ARQ and adaptive fair-queueing, which has the best performance, especially under a heavy load condition.
Figure 6.9 Average delay of UBR Traffic Vs coding rate

Figure 6.10 Average delay of UBR services Vs Traffic load
Figure 6.11 shows the cell loss rate (CLR) of VBR traffic and average delay of UBR traffic in terms of traffic load which is increased by increasing the source rate of the VBR traffic. By the use of a combination of type II hybrid ARQ and adaptive fair-queuing, the CLR of VBR traffic can be kept under a low value. It is further noticed that in some traffic load regions, the type II ARQ without adaptive fair-queuing has better UBR delay performance than type II hybrid ARQ with adaptive fair-queuing, as the VBR traffic suffers a larger CLR.

6.3 RETRANSMISSION ALGORITHM BASED ON POWER PRIORITIES

The main challenge in wireless communication is to overcome the disadvantage of the hostile characteristics presented by the radio channel. In this regard, power control is one of the important issues in the environment of retransmission. The proposed Retransmission Algorithms based on Power Priorities (RAPP) may be used for power control at the retransmission of request packets sent by remote terminals to the base station (BS) in WATM networks.
6.3.1 Priorities using different power levels

The proposed power control scheme considers a TDMA / TDD access scheme with fixed block size of 64 slots. As part of this framework, a scheduling mechanism that assigns slots to the terminals in a block-by-block basis has been developed for the base station. It allows and promotes the satisfaction of QoS guarantees for voice, data and video services. The scheduler handles the five standardized types of services (CBR, rt-VBR, nrt-VBR, ABR, and UBR), assigning priorities according to Table 6.1, where a bigger number implies a higher priority.

<table>
<thead>
<tr>
<th>Type of service</th>
<th>CBR</th>
<th>rt-VBR</th>
<th>nrt-VBR</th>
<th>ABR</th>
<th>UBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

The idea behind RAPP is to help the WATM cells that have failed to reach the receiver on the request procedure. After each failure, there are a set of power levels from which a contending terminal may randomly choose increased power until the predetermined limit. Providing a terminal the opportunity to use a higher transmission power increases its capture probability at the receiver. At the same time, by potentially increasing the power level after each retransmission, the fairness in serving the earlier requesting terminals first is assured to some extent.

An important motivation to use the RAPP algorithm is to take advantage of the capture effect which has been shown to increase the throughput. Using different power levels has the advantage as compared to back off mechanisms in that no waiting periods are necessary and still one of the terminals has a very good chance of being successful. Further capture effect is considered in the receiver’s side. QoS restrictions are imposed by the WATM network such as the need to guarantee a maximum delay, cell loss ratio and to limit the maximum number of retransmissions which is very important for delay sensitive services.
The following rule is proposed to specify the different power levels \( \{P_1, P_2, \ldots, P_m\} \) to be used in RAPP algorithm so that contenders who have waited for longer time still have an advantage over new ones, regardless of the total number of contenders and the power levels they are using.

\[
P_i = z P_1 \quad \text{for} \quad 1 \leq i \leq m
\]

where ‘m’ is the number of power levels given by,

\[
m_p = \log_z (P_u / P_l) + 1
\]

This results in power vector as follows:

\[
[P_1, P_2, \ldots, P_m] = [P_1, z P_1, z^2 P_1, \ldots, P_m]
\]

As an example, it is assumed that \( P_1 = 0.625 \) mW, \( P_m = 160 \) Mw and \( z = 2 \) to capture the power vector generated. The \( m_p \) power levels are determined according to the minimum \( (P_L) \) and maximum \( (P_u) \) allowed transmission power levels for the terminals. The minimum power \( (P_L) \) must satisfy SNR and bit error rate (BER) constraints. Assuming a given \( P_L \) and \( P_u \), the minimum power level to be used by a terminal \( P_1 \) should be above \( P_L \), while the maximum power level \( P_m \) should be below \( P_u \).

### 6.3.2 RAPP Algorithm

Two priorities are defined in this protocol; the higher priority is for real time applications (CBR, and rt-VBR) and the lower priority is for non real time applications (nrt-VBR, ABR, and UBR). The set of power levels is divided into two subsets, each corresponding to one of the priorities. Let the subsets be defined as: \( \{P_1, P_2, \ldots, P_k\} \) for low priority and \( \{P_{k+1}, P_{k+2}, \ldots, P_m\} \) for high priority terminals.

In its first attempt, each terminal will send its request through one of the request slot using the lowest possible power level corresponding to its priority. After each
collision, the terminal will select a new power level from among those allowed by the RAPP protocol according to its priority level and the number of failed attempts. The operation of the retransmission algorithm is as follows:

- If the mobile terminal (MT) corresponds to a high priority service, it will select between $P_{k+1}$ and $P_{k+2}$ in its first retransmission. After each new collision, it will add to the allowed set of power level that is right above the highest one the set currently has. When there are no more power levels to add, the MT will limit itself to using one of the highest power levels ($P_m-1$ and $P_m$). If the MT fails to reach the base station (BS) again, the process is stopped for a random time selected from the set $\{T, 2T\}$, where $T$ is block duration. After this silence, if the packet has not expired, the station will start trying again, choosing between the highest power levels. If transmission is again unsuccessful after a predetermined number of attempts ($m-k$), it will stop again for a period selected from the set $\{T, 2T, 4T\}$ and so on.

- On the other hand, if the MT corresponds to low priority service, it will select between $P_1$ and $P_2$ in its first transmission. After each new collision, it will shift the allowed set upwards but still keeping two options in the set. When it is not possible to move upwards, i.e., when $P_k$ is already a member of the allowable set, the MT will transmit one for the last time using $P_k$ and then, if the transmission is again unsuccessful, it will stop sending requests for a random time. When this happens, a full cycle will have been completed. The process is restarted later, repeating the procedure from the first transmission.

The above procedure is shown in Figure 6.12. In all cases, regardless of the priority the MT belongs to, it will randomly select its power level considering that all the options in the allowable set are equally possible.

It is to be noted that there is no differentiation among the MTs that correspond to the same priority group. Therefore CBR and rt-VBR users share the same privileges.
However, RAPP has to do with the delivery of request to the base station only. It is the responsibility of the scheduler at the base station to make sure that the five priorities shown in Table 6.1 are respected when it comes to actually allocating the bandwidth.

![Diagram of power levels for RAPP algorithm]

**Figure 6.12 Power levels for RAPP algorithm**

### 6.3.3 Performance Analysis of RAPP

The achievable throughput $S$ in a slotted ALOHA environment under traffic generated according to a Poisson process with average arrival rate $G$ is given by,

$$S = \lambda e^{-\lambda}$$  \hspace{1cm} (6.19)
When capture effect is taken into account,

$$S = \lambda e^{\lambda} \{ 1 + \sum_{m=1}^{n} \frac{\lambda^m}{m!} P[capture/C_m] \}$$

Equation (6.20)

In the Eq.(6.20), $C_n$ is used to denote collision with $n$ contenders. The capture probability is calculated for multipath fading distribution. Finally, when several mini-slots ($m$) are provided and the users are allowed to choose randomly among them, the total load $\lambda$ is evenly divided among the mini-slots, which results in the following expression for the throughput per mini-slot:

$$\eta = \frac{\lambda}{M} e^{\frac{-\lambda}{M}} \left[ 1 + \sum_{m=1}^{n} \frac{(\lambda/M)^m}{m!} P[capture/C_n] \right]$$

Equation (6.21)

Figure 6.13 Throughput Vs Traffic load under RAPP

The relationship between throughput and traffic load is shown in Figure 6.13. The generated traffic is the combined traffic, created by the five types of traffic generators plus retransmissions. Here the throughput ($\eta$) is the measure of the effective
utilization of the request channels. The maximum throughput corresponds to the reception at the base station of 3 or 7 successful request packets per frame. As already mentioned, the TDMA access scheme is considered. From Figure 6.13, the throughput achieved with RAPP is a maximum value of 0.65 which means that it corresponds to around 4.5 out of 7 possible packets being correctly received per frame. The simulations are carried out for different number of available mini slots (m=3, 6, 7). The performance of a Retransmission algorithm based on power priorities has been analysed. Simulation results show that the throughput achieved is greater in this scheme.

6.4 DECENTRALISED CALL ADMISSION CONTROL SCHEME

It is well known that ATM is a connection-oriented service. Before a user starts transmitting over an ATM network, a connection has to be established. This is done at the call set-up phase. The main objective of this set-up procedure is to establish a path between the sender and the receiver which may involve one or more ATM switches. At each of these ATM switches, resources have to be allocated to the new connection. An important part of the connection establishment procedure is the Call Admission Control (CAC) function. CAC is defined as the set of actions taken by the network during call set-up phase to determine whether a connection request can be accepted or rejected[132].

In this work, a Decentralised Call Admission Control (DCAC) scheme is presented for WATM networks which statistically multiplex user mobility information to estimate future demand and available resources. Resource reservation for handoff calls is dynamically adjusted according to the user mobility information to meet the QoS requirements for heterogeneous traffic. The proposed decentralized CAC is an online scheme which is executed partially on serving Access Point (AP) and partially at Mobility Supporting (MS) ATM switch.

6.4.1 Estimation for new connections

To estimate the bandwidth, a mechanism is required to gather cell counts at each connection and update the traffic distribution at the MS-ATM switch. Therefore each switch is provided with a cell-count request timer that is associated with it. The cell-
count request timer starts when the switch is engaged. When a cell-count of switch \( j \) expires, a cell-count request is issued to the switch for each connection that has node \( j \) as its destination. A queue is used to keep track of the number of cell-count requests issued \((X_1, X_2 \ldots X_n)\) and the number of responses received. When enough cell-count samples are collected, the effective bandwidth of the connection can be calculated using the following equations:

\[
\text{Effective Bandwidth of the connection} = \frac{\lambda_T(\delta)}{\delta}
\]  

(6.22)

where \( \delta \) is jitter bound given by,

\[
\delta = -\frac{\log(\Gamma_o)}{b}
\]

The arrival rate is given by,

\[
\lambda_T(\delta) = \frac{1}{T_i} \log\left(\frac{1}{n}\right) \sum_{i=1}^{n} e^{\delta x_i}
\]  

(6.23)

where,

- \( T_i \) is the interval between cell counts
- \( \Gamma_o \) is the cell-loss ratio
- \( b \) is the buffer space
- \( X_i \) is the cell-count requests issued
- \( n \) is the number of cell-count samples
- \( \lambda_T \) is the arrival rate

When a new connection is received at the AP, the AP will attempt to reserve the bandwidth for the connection from the Radio Resource Manager (RRM). The amount of bandwidth required by a new connection which is referred to as the equivalent capacity can be obtained in the following way. If a new connection with the parameters \( (R, b_1) \) inputs a link with buffer capacity ‘\( X \)’, where ‘\( R \)’ is the peak bit rate of the connection
and ‘t₁’ is the average duration of the connection active period (mean burst of the period), the equivalent capacity C of that connection would be,

\[
C = \left( R \alpha (1 - \rho) - X_C \right) + \sqrt{\left( \alpha b_1 (1 - \rho) - X_C \right)^2 + 4X_C \rho \alpha t_1 (1 - \rho) + (2 \alpha t_1 (1 - \rho))}
\] (6.24)

where,

- \( R \) is the peak bit rate of the connection
- \( \alpha \) is the smoothing co-efficient
- \( t_1 \) is the average duration of the connection
- \( \rho \) is the source utilization
- \( X_C \) is the buffer capacity

The equivalent capacity in Eq.(6.24) is used as an estimate of the bandwidth requested by the new connection which is then compared with the effective bandwidth of the destination port calculated in Eq.(6.22). The result available is then used for the next ‘update period’ until the next update of available bandwidth is carried out. There is a possibility that the connection might be accepted by the proposed method but later rejected by the actual CAC as this method is the estimation.

6.4.2 CAC with user mobility information

The proposed CAC is focused on CBR and VBR traffic. For ABR traffic, all new and handoff requests are treated as the same. There is no resource reservation for ABR handoff call. All existing ABR calls will share the remaining resources after serving CBR and VBR calls fairly and efficiently. Three aspects are included in the proposed CAC scheme:

i) Resources reservation for potential handoff calls updated periodically at each BS.

ii) New CAC performed whenever a BS receives a new call request.

iii) Handoff CAC performed whenever a handoff event occurs.
To admit a call (new or handoff), two conditions must be satisfied. Guaranteed QoS for a particular call should be provided and QoS performance of all existing calls should still be guaranteed after the admission. The same conditions are applied to reserve resources for any CBR or VBR class.

The available time is divided into equal intervals of length \( \tau \) and the time sequence is given by \( t = 0, \tau, 2\tau, \ldots \), so that there is at most one handoff event for any active mobile terminal \( 'x' \) in each of the intervals. It is assumed that at the beginning of each \( \tau \) interval, the handoff probabilities, \( P_{ab}(t) \), of each mobile \( 'x' \) from cell \( 'a_1' \) to cell \( 'a_2' \) are calculated. The handoff probabilities are known to the mobile's current cell \( 'a_1' \) and the probabilistic information can be exchanged to all neighboring cells in order to reserve necessary resources for handoff calls.

Let cell \( 'a_1' \) be the reference cell and \( n_{ic} \) and \( n_{jv} \) are defined as the number of equivalent handoff calls in the \( i^{th} \) CBR class and \( j^{th} \) VBR calls respectively, which need cell \( 'a_1' \) to reserve resources in the next \( \tau \) interval. The number of equivalent handoff calls for a traffic class is the average number of calls in the class. This will handoff to cell \( 'a_1' \) in the next \( \tau \) interval and can be calculated as the accumulated handoff probability of all the calls in the class that will handoff from the neighboring cells to cell \( 'a_1' \) in the next \( \tau \) interval.

At the beginning of every \( \tau \) interval, resource reservation is updated at each BS for all classes of the traffic from the highest (the first CBR class) to the lowest (the last VBR class) priorities. The algorithm starts with CBR for bandwidth reservation. If the current calculated class is CBR class, it accordingly calculates the jitter bound. If there is no CBR class then it will reserve for VBR and determines the delay bound and QoS which must be satisfied for any class.

In the proposed model, an issue of paramount importance is the estimation of call blocking probability (\( P_B \)) and handoff dropping probability (\( P_{hd} \)). As long as real time
measurements are not available, the parameters can be selected based on the average speed of mobiles, their movement patterns and the coverage area of wireless cells. Once the network is deployed and traffic measurements are available, these probabilities can be refined based on the statistical analysis of such measurements. For example, the data regarding the movement and holding time can be collected based on the statistical analysis of the data. Then $P_B$ and $P_{hd}$ can be defined accordingly. Here the location, time of the day and other factors can be taken into consideration. An example of this would be to collect mobility details and call holding time statistics in highway or street micro cells, and define $P_B$ and $P_{hd}$ based on the time of the day to capture the peak hour traffic characteristics and incorporate that in the call admission algorithm.

In fact these algorithms can be readily incorporated in existing wireless and cellular networks. Each base station communicates the number of wireless users in its domain to all its neighboring base stations from the Mobile Switching Centre (MSC) in a periodic manner. As new information becomes available, the call admission controller of a base station recalculates the admission threshold to reflect the change in the system. Finally the rate of the change in the number of users in neighboring cells can be further used to refine the movement parameters ($P_B$ and $P_{hd}$) in order to increase the performance of the call admission algorithm defined in terms of wireless utilization.

6.4.3 Performance Analysis of DCAC

The performance analysis of the proposed Decentralised Call Admission Control scheme includes delay analysis and call rate analysis.

The WATM processor control system operates on a delayed basis when requests are queued or scanned at regular intervals. There are 2 levels of control namely, level 1 and level 2. At level 1, a processor control is involved in scanning and interfacing with customers. At level 2, central processors are involved when all the data for a call request are received. The delay in the processor can be calculated as,
Average call delay = \[ \frac{1}{2} \left( \frac{1}{\mu_1} \right) \] (6.25)

where,

\[ \frac{1}{\mu} \] is the holding time
\[ \mu \] is the average call duration
\[ \alpha \] is the traffic load

It is extremely difficult to accurately estimate how a system will perform under real-time traffic conditions. For that, a traffic simulation can be used in which the total capacity is determined by the probabilities of false traffic and peak traffic. The total capacity is evaluated for the entire call processing. The channel holding time is defined as the time during which a new or handoff call occupies a radio channel in a given cell and it is dependent on the mobility of the user. While this is similar to the call holding time in the fixed telephone network, it is often a fraction of the total call duration in a wireless cellular network. Most research work on CAC and bandwidth reservation assumes that the channel holding times in all cells are independent and identically distributed according to an exponential distribution [133]. With the help of this total capacity and holding time, admission delay is calculated and used in the simulation.

Call blocking rate is defined as the rate at which the calls are blocked due to the unavailability of resources. The actual blocking probability data must be used to check the outcome from the Erlang B model. Although the difference can be up to 15 percent, the Erlang B model is still considered as a good model for obtaining useful estimates.

The counting of a dropped call is after the call is established but before it is properly terminated as described below. The definition of the phrase “call is established” means that the call is setup completely by the setup channel. If there is a possibility of a call drop due to no available voice channels, this is counted as a blocked call, not a dropped call. Call drops are caused by factors such as:
- Unsuccessful completion from set-up channel to voice channel
- Blocking of handoffs (switching capacity)
- Unsuccessful handoffs (processor delay)
- Interference (foreign source)
- Improper setting of system parameters

The percentage of call drops (P_{cd}) is expressed as the ratio of number of call drops before completion to total number of accepted calls handled by set-up channels. Sometimes call drops due to weak signals are called lost signals. This case happens when the mobile or portable units are at a standstill and the radio carrier is changed from a strong setup channel to a weak voice channel due to the selective frequency fading phenomenon. The simulation of call blocking rate and call dropping rate is performed for DCAC and compared with centralised CAC.

Consider a wireless mobile ATM network where several BSs are connected to an ATM backbone network through a Mobile Switching Center (MSC). The multiple access mode used is Time-Division Multiple Access (TDMA). Each BS can support different types of traffic. This work is confined to the service area of a particular MSC. The wireless system is characterized by a broadcast channel in the forward direction (or downlink), where the BS is capable of broadcasting packets to all users and a multiple access channel in the reverse direction (or uplink), where all users are capable of sending packets to the BS. The measures of performance are the call blocking probability (P_B) and handoff dropping probability (P_{hd}). Call blocking and handoff dropping probabilities are simulated and comparison is made with the centralised scheme. This is extended to handle heterogeneous services and its performance is also analysed. To achieve a complete view of the results, the following two scenarios are considered:

(a) A Centralised CAC method that uses the MS-ATM switch to select the route and also to decide whether to accept the connection or not.

(b) A Decentralised online CAC that uses the estimated available bandwidth from the last update to decide whether to accept the connection or reject it.
There is a possibility that a connection might be accepted by the proposed method but later rejected by the actual CAC because this method is an estimation. The main objective of the generic CAC introduced by the ATM Forum is to make a quick decision on whether to execute the real CAC.

One dimensional (1D) cellular arrays are considered for simulation purpose. The simulated system has five radio cells arranged in a circle to avoid the boundary effect. Mobile unit can be in any one of the cells with equal probability. Let the initial location where mobile unit originates its call request be uniformly distributed in the 1D region of its current cell, where D is equal to 1500 m. The mobile can move in either direction of the 1D region with equal probability, and its velocity is uniformly distributed between 5 and 20 m/s. The interval used to update the estimation is $\tau = 60s$.

The traffic classes in the system are CBR, VBR and ABR. It is assumed that the call arrival process follows a Poisson distribution with parameters $\lambda$, where $1/\lambda$ is varied within the range of (0-30) sec and call duration of $1/\mu = 50$ sec. For any particular call, it can be any of the three traffic classes. In the simulation, the call blocking rate and handoff dropping rate for three traffic classes are determined by using the performance bounds.

Figure 6.14 shows the difference between the estimated bandwidth and the available bandwidth. It illustrates that the proposed method works accurately for a network with heavy load conditions. If the update of the available bandwidth is done at shorter intervals, then this estimation will be very close to the accurate values. Furthermore, if the network being examined does experience abrupt changes in traffic loads, then the update period can be reduced even more to meet the network needs.

The decentralised and centralised methods are compared for the number of connections admitted by each of them during fixed periods. Figure 6.15 shows clearly how the proposed CAC scheme admits more connections when compared to other
methods. The connections are admitted on the basis of QoS. A call is admitted only when it satisfies the predefined quality of service.

Figure 6.16 shows the number of connections admitted for the centralized CAC and the proposed decentralized CAC. Figure 6.17 shows the considerable delays resulting from the centralised to the proposed online decentralised method. Assume that three APs have forwarded connection requests to the MS-ATM switch at the same time. If the switch is to use the centralised approach, then the time required to take the admission decision will increase rapidly in the centralised approach in comparison to a slight increase for the decentralised approach as shown in Figure 6.17.

Decentralised CAC admits more connections than centralized CAC. This is shown in terms of the throughput. The result in Figure 6.18 also shows the delay in the admission decision as the number of channels in each AP increases for both approaches. It is clear that there is a reduction in the admission response times in the decentralised approach compared to the centralised approach.

The call blocking and call dropping probabilities for centralised and distributed CAC schemes are simulated and compared. At lower load conditions, the centralized CAC system has a higher blocking probability than the decentralised call admission control algorithm. By increasing the system capacity, the performance of both systems becomes comparable at low load conditions. However, at high load conditions, the decentralised CAC provides better performance. The results in Figure 6.19 and Figure 6.20 show that by increasing the system capacity, the performance of the DCAC is efficient.

The calls are admitted only when it satisfies their corresponding bounds. Based on that analysis, the blocking and dropping are simulated as shown in Figures 6.21 and 6.22. The CAC scheme that uses jitter bounds can reduce the forced handoff dropping rate of both CBR and VBR calls. The CAC scheme using user mobility information can achieve a much lower call blocking rate and higher resource utilization, while keeping
the handoff dropping rate at a very low level. Results show that the proposed scheme achieves a much lower handoff call dropping rate than the centralised CAC.

Figure 6.14 Estimation of Bandwidth

Figure 6.15 Number of connections admitted for the two methods
Figure 6.16 Comparison of the Admission delay

Figure 6.17 Holding time Vs Admission delay
Figure 6.18 Comparison of the Throughput

Figure 6.19 Comparison of Blocking probability
Figure 6.20 Comparison of Dropping probability

Figure 6.21 Call blocking probability for heterogeneous services in DCAC
6.5 SUMMARY

Three MAC layer protocols have been discussed in this chapter namely, Combined Hybrid ARQ and Adaptive scheduling, RAPP power control algorithm and Distributed call admission control. Results show that these protocols enhance the overall quality of WATM networks. The results of these schemes are summarized as follows:

- A combined hybrid ARQ error-control and adaptive scheduling scheme is presented for WATM networks. Instead of treating the error recovery and scheduling schemes in separate DLC and MAC layers, the scheduling scheme is implemented in conjunction with a retransmission scheme. In particular, with the aid of proper channel modeling, type II hybrid ARQ is chosen as the error recovery scheme to combat fading effects while adaptive fair-queuing is adopted as the scheduling scheme to achieve a fair and efficient resource allocation. Specifically, the weight of a connection used in the fair-queuing algorithm
dynamically varies according to the current channel condition, as well as the type of service. Various simulation results show that the proposed scheme can achieve a high throughput with minimum delay and cell loss rate compared to conventional algorithms.

- The proposed Retransmission algorithms based on power priorities (RAPP) may be used for the power control of the retransmission of request packets sent by remote terminals to the base station (BS) in WATM networks. Simulation results show that the throughput achieved is greater in this scheme.

- In this work, a new decentralised CAC is presented. The available bandwidth is calculated at each node and updated regularly to be used by the CAC to route the incoming call. The method proposed takes a quick decision on the connection acceptance before the actual CAC takes the final decision. Mobility information is likely to be very useful in CAC for wireless ATM networks, as it is used to estimate future resource requirements. The accuracy of resource estimation which is essential to the CAC scheme was determined by the available mobility information and the update time. Simulation results illustrate a considerable improvement in the CAC responsiveness and in the overall network performance. Under heavy traffic conditions, the proposed method performs very well. A comparative analysis of the decentralised approach with the centralised approach for different measures such as admission delays and throughput has been carried out. The benefits of the proposed DCAC have been proved by the results. Moreover, the proposed CAC is simple enough that the admission decision can be made in real time without much computational difficulties. Furthermore, the handoff dropping and blocking probabilities are limited by the proposed CAC scheme to a minimum level almost independent of load conditions.