CHAPTER 5

HYBRID TECHNIQUES USING CLD AND BWA

5.1 INTRODUCTION

In hybrid technique, bandwidth adaptation along with prioritized weight assignment and Cross-Layer Design (CLD) are done to minimize the delay experienced by mobile multimedia applications.

A joint cross-layer approach between application layer and MAC layer for enhancing Quality of Service (QoS) for multimedia applications was implemented. Bandwidth adaptation (BWA) scheme based on application utility function is used in the application layer as it reflects the satisfaction degree of end-users with the allocated bandwidth. Prioritized Weight Assignment (PWA) in MAC layer assigns higher weight to prioritized multimedia data based on the allocated bandwidth.

The application utility function is the cross-layer parameter which jointly adapts the application layer function in coordination with the MAC layer to minimize the delay experienced by the multimedia data. Simulation results have clearly demonstrated the superior performance of the proposed joint cross-layer schemes such as bandwidth adaptation with Prioritized Weight Assignment (PWA) by comparing them with some existing ones.

The rapid development of wireless communication technologies and its services has greatly enriched the diversity of wireless applications evolving
from the traditional voice service to a wide range of multimedia services including data, voice, and video.

Different multimedia services over networks have different bandwidth requirements. For example, applications like audio phone and video conference require strict end-to-end performance guarantees; hence it is crucial for the networks to provide reliable and timely packet transmission. On the other hand, applications such as E-mail and file transfer can adapt their bandwidth to various network loads since they can tolerate certain delays. As a result, providing QoS to multimedia applications according to their bandwidth requirements is becoming an important resource management issue for wireless networks.

Bandwidth adaptation is one of the most promising resource management methods to provide QoS guarantees to multimedia traffic in wireless networks. The main feature of bandwidth adaptation is that it can explore the adaptive nature of multimedia applications and dynamically adjust their allocated bandwidth to deal with network resource fluctuations. Examples of such multimedia services include the International Organization for Standardization’s (ISO’s) Motion Picture Experts Group 4 (MPEG-4) [ISO00-1] and the International Telecommunication Union’s (ITU’s) H.263.

5.2 UTILITY-BASED MULTIMEDIA TRAFFIC MODEL

A wireless network featuring limited and varying bandwidth resource, different applications have different bandwidth requirements. To provide QoS support to multimedia applications according to their bandwidth needs, an explicit traffic model is needed to reflect the QoS sensitivity of the applications for bandwidth allocation. So, a utility-based multi-class traffic model is proposed to differentiate multimedia traffic according to their adaptive characteristics.
5.2.1 Utility Function

Utility function is defined as a curve mapping the amount of bandwidth received by the application to the performance as perceived by the end-user. The key advantage of utility function is that it can inherently reflect the QoS requirements of the end-to-end user and quantify the adaptability of the application. The shape of the utility function varies according to the adaptive characteristics of the application. According to the bandwidth requirements, multimedia traffic can be classified into two broad classes:

- Class I – real-time traffic
- Class II – non-real-time traffic.

Class I traffic can be further classified into two subclasses

- Adaptive Real-Time traffic
- Hard Real-Time traffic.

5.2.2 Adaptive Real-Time Traffic

Adaptive real-time traffic refers to the applications that have flexible bandwidth requirements. The utility curve is depicted in Figure 5.1. The term intrinsic bandwidth $b_{intr}$ is used for adaptive real-time applications. The intrinsic bandwidth gracefully adjusts the bandwidth in the event of network congestion and is the acceptable limit for adaptive real-time applications. The minimum bandwidth $b_{min}$ is the minimum limit below which the quality degrades.
Figure 5.1 Utility curve for Adaptive real time traffic

Typical examples are interactive multimedia services and video on demand. The utility function of adaptive real-time traffic is modeled as follows:

\[ u(b) = 1 - e^{\frac{k_1 b^2}{k_2 + b}} \]  

(5.1)

where \( k_1 \) and \( k_2 \) are two positive parameters which determine the shape of the utility function and ensure that when the maximum bandwidth requirement \( b_{\text{max}} \) is received, the achieved utility \( u_{\text{max}} \) is approximately equal to 1.

5.2.3 Hard Real-Time Traffic

Hard real-time traffic refers to the applications with stringent bandwidth requirements. Examples include audio/video phone, video conference and telemedicine. The following utility curve and function is used to model hard real-time traffic.
Non-real-time traffic refers to the applications which are rather tolerant of delays. Most traditional data applications such as E-mail, file transfer and remote login belong to non-real-time traffic and they can work without guarantees of timely packet delivery. The following utility function is used to model non-real-time traffic.

\[
u(b) = \begin{cases} 
1, & \text{when } b \geq b_{\min} \\
0, & \text{when } b < b_{\min}
\end{cases}
\]  

(5.2)

**5.2.4 Non-Real-Time Traffic**

Figure 5.2 Utility curve for Hard real time traffic
Figure 5.3 Utility curve for Non-real time traffic

where $k$ is a positive parameter which determines the shape of the utility function and ensures that when the maximum bandwidth requirement $b_{\text{max}}$ is allocated, the achieved utility $u_{\text{max}}$ is approximately equal to 1.

5.3 PRIORITIZED WEIGHT ASSIGNMENT (PWA)

The data packets to be transmitted are classified into two major categories as real time and non real time data. Real time and non real time data coming from the source are separated based on type of service. Transmission of real time data alone will reduce the delay of the transmitted real time data in the network.

5.3.1 Priority Assignment

By assigning higher priority to the real time data and queuing those data separately and transmitting the real time data before the non real time...
data will reduce the loss of the real time data during the transmission and also
decrease the delay in the transmission of the data. For separating the real time
data the timing sequence of the packet coming from the network layer is
considered. Higher weights are assigned to high priority real time data and the
other data are assigned lower priority.

Thus, the data in the higher priority are sent before the lower
priority and thus the loss of data and delay in transmitting the data are reduced
to a certain level.

5.3.2 Bandwidth Assignment

In this approach, bandwidth to the certain level of the data is
assigned as

\[ B_i = (P_i / \sum_j P_j).B \]  \hspace{1cm} (5.4)

where \( B_i \) is the bandwidth of the \( i \)th data, \( P_i \) is the priority of \( i \)th data, \( P_j \) is the
priority assigned to overall data, \( B \) is the overall bandwidth assigned.

5.3.3 Weight Assignment

Like priority, weights are also assigned to the packets such that the
weighted packets are transmitted first. Based on the priority, more weights are
given to higher priority packets and are transmitted first. Weights are assigned
in such a way that, the weights of the higher priority must be higher than the
lower priority packets as

\[ W_i \geq W_{i+1}, i=1,2,... \]  \hspace{1cm} (5.5)
The assigned weights must be updated such that the data can be transmitted without delay and effective output can be obtained.

The updating of the weights can be done as

\[ W_1^{(\text{new})} = W_1.(RT_1)^{-\alpha} \]

\[ W_i^{(\text{new})} = \min \left[ \left( \frac{W_i}{W_{i-1}} \right) W_{i-1}^{(\text{new})}, W_i.(RT_i)^{-\alpha} \right] \]  \hspace{1cm} (5.6)

Using the above approach, bandwidth can be assigned to specified data and will allow the data to be transmitted with a collision free medium and thus the delay can be reduced up.

5.4 SYSTEM MODEL

In wireless networks with multimedia traffic, each call is assigned a utility function with shape depending on its traffic class. When a call requests a connection to the network, it is assumed to provide the following information

- Traffic class
- Bandwidth requirements
- Utility function

With adaptive bandwidth allocation paradigm, if there is enough bandwidth available in the network, the call is allocated its maximum bandwidth \( b_{\text{max}} \); otherwise, depending on how much the network is overloaded, the call is allocated a bandwidth ranging from its minimum bandwidth \( b_{\text{min}} \) to its maximum bandwidth \( b_{\text{max}} \).
Although there is still no standard overall architecture for the end-to-end implementation, some partial solutions have been proposed. The adaptation of multimedia applications can be achieved at different Open System Interconnection in wireless networks.

**Figure 5.4 ISO-OSI Layers with bandwidth adaptation**

The function of each OSI layer is as follows:

**Physical layer:**

At the physical layer, adaptability can be achieved by choosing appropriate modulation techniques e.g. PSK.

**Data link layer:**

At the data link layer, error control mechanisms e.g. retransmission can be used to protect against the varying error rates of wireless links.
Network/Transport layer:

At the network/transport layer, routing methods can be used to adapt the applications when there is user mobility.

Application layer:

At the application layer, most multimedia applications can adapt to the changing networking conditions using various multimedia coding techniques.

To support the QoS provisioning in wireless networks, bandwidth adaptation often needs to make tradeoffs between multiple QoS objectives. Each call in the network is assigned a utility function to reflect the relationship between bandwidth allocation and the end-user’s satisfaction.

Bandwidth adaptation is performed in each BS and it consists of two processes – bandwidth degrades and bandwidth upgrades, which are triggered by call arrival events and call departure events, respectively (see Figure 5.2)

![Figure 5.5 Bandwidth Adaptation](image)
Call arrival events include new call arrival events (a new call is generated within the cell) and handoff call arrival events (a handoff call arrives to the cell). Call departure events include call completion events (a call within the cell completes) and outgoing handoff events (a call leaves its current cell).

5.4.1 Bandwidth Degrade

When a new or handoff call arrives to a cell of the network, if the cell has enough bandwidth available, the new or handoff call is admitted at its maximum bandwidth requirement. If the cell is overloaded, the bandwidth of adaptive ongoing calls can be degraded to smaller values to accommodate the new or handoff call.

5.4.2 Bandwidth Upgrade

When an ongoing call is terminated due to its completion or outgoing handoff from its current cell to another, if all calls in the current cell have received their maximum bandwidth, the released bandwidth is saved for future use. Otherwise, the released bandwidth can be utilized to upgrade the adaptive ongoing calls that have not received their maximum bandwidth. With the proposed scheme, each call in the network is assigned a utility function.

5.5 CROSS-LAYER SCHEDULING ARCHITECTURE

The scheduling algorithm at the MAC layer is modeled as an optimization problem with respect to some physical layer constraints and application QoS constraints.
At every timeslot, the scheduling algorithm has to produce rate allocation and power for all the k users, which is based on the observation of the current channel state information (CSI) from the physical layer and the queue state information (QSI) from the application layer. Rate allocation and power allocation are selected so that they optimize the system objectives.
5.5.1 Uplink Scheduler

The uplink (UL) scheduler comprises three modules: information module, database module and service assignment.

![Diagram of Uplink Scheduler]

Figure 5.7 Uplink Scheduler
**Information Module**

This module performs the following functions and passes them to the scheduling data base module.

- Categorizing the packets on the MS (Mobile Station) basis.
- Extracting the queue size information (number of waiting packets and size of each packet of each connection from the BW (Bandwidth-Request messages)).
- Deciding the arrival time.

The process determines the arrival time of packets that arrived during the previous frame i.e. decides time bound. The time bound is given by the sum of the packet’s arrival time and the packet’s maximum delay requirement (as determined by the connection QoS Parameters).

The type field in the bandwidth request header indicates whether the request is incremental or aggregate. Since piggybacked Bandwidth Requests do not have a type field, PBR shall always be incremental.

When the BS (Base Station) receives an incremental bandwidth request, it shall add the quantity of bandwidth request to its current perception of bandwidth needed for the connection. When the BS receives an aggregate bandwidth request it shall substitute its perception of bandwidth needs of the connection with the quantity of bandwidth requested.
**Scheduling Database Module:**

The scheduling database module serves as the information database for all the MS in the network.

**Service Assignment Module:**

The service assignment module determines the uplink sub-frame allocation in terms of the number of bits per MS. The number of bits will finally be converted to the number of time slots, i.e. the units used in the information elements (IE) of the Up Link Map (UL-MAP) generation. The number of bits per time-slot is determined by the physical layer of the wireless network. The surplus bandwidth is distributed among all the connections according to their instantaneous bandwidth requirements. The sharing of the surplus bandwidth among different classes follows priority logic, from highest to lowest: UGS, rtPS, ertPS, nrtPS and BE. In another word the method boils down to the following two stages:

- Reserved bandwidth of each priority class is distributed among the admitted connections of that class, according to the desired scheduling policy.

- The excess bandwidth would be allocated to those connections that have not been granted a part or total of their requested bandwidth.
5.5.2 Downlink Scheduler

In this structure, the scheduling scheme is based on a principle called Grant Per Type-of-Service (GPTS). The delay performance is differentiated for each queue and respective time slots are assigned with varying priorities and QoS guarantee. The priority function is given below for queue $i$ which is defined as

$$
\mu_i(t) = \frac{r_i(t)}{R_i(t)/C_i(t)M_i(t)}
$$

(5.7)

where $M$ is the minimum rate requirement, $C_i(t)$ is the number of connections of the $i$ queue and $R_i(t)$ is the $i$th transmission capacity at time $t$, which is determined by the channel quality. Each queue corresponds to one QoS requirement class, respectively.

The queue having the highest value of $\mu_i(t)$ is scheduled. The estimated average $I$ throughput of scheduled queue $i$ (denoted as $R_i(t)$) is updated by the following simple exponential smoothing model which is given below,

$$
R_i(t+1) = (1 - \frac{1}{T_i})R_i(t) + \frac{1}{T_i} \sum_{n=\Delta_i(t)} r_i(t, n)
$$

(5.8)

Where $r_i(t)$ is the supportable data rate for user $i$ in sub carrier $n$ at time slot $t$ and $R_i(t)$ is the average throughput of $i$ and $i^{th}$ is received by the $i$ user up to slot $t$ that indicate the set of sub-carriers in which user $i$ is scheduled for transmission at time slot $t$. In this case, the system throughput is enhanced and efficient bandwidth utilization is achieved.
To evaluate the performance of the proposed utility-maximization bandwidth adaptation scheme, a multimedia wireless network simulation model has been developed.
Table 5.1 Simulation Parameters

<table>
<thead>
<tr>
<th>App. Group</th>
<th>Traffic Class</th>
<th>Bandwidth Requirement (Mbps)</th>
<th>Average Connection Duration</th>
<th>Example</th>
<th>Utility Function ((b) is Mbps)</th>
</tr>
</thead>
</table>
| 0          | I (Hard Real-Time) | \(b_{\text{max}} = 0.03\) \(b_{\text{des}} = 0.03\) | 3 minutes | Voice Service & Audio Phone | \[1, b \geq 0.03\]
|            |               |                              |                           |         | \[0, b < 0.03\] \(u_{\text{max}} = 1\) |
| 1          | I (Hard Real-Time) | \(b_{\text{max}} = 0.25\) \(b_{\text{des}} = 0.25\) | 5 minutes | Video Phone & Video Conference | \[1, b \geq 0.25\]
|            |               |                              |                           |         | \[0, b < 0.25\] \(u_{\text{max}} = 1\) |
| 2          | I (Adaptive Real-Time) | \(b_{\text{min}} = 1\) \(b_{\text{intr}} = 1.5\) \(b_{\text{des}} = 2\) \(b_{\text{max}} = 6\) | 10 minutes | Interact. Multimedia & Video on Demand | \(1 - e^{-\frac{1.38}{8.3+b}}\)
|            |               |                              |                           |         | \(u_{\text{max}} = 0.99\) |
| 3          | II (Non-Real-Time) | \(b_{\text{max}} = 0\) \(b_{\text{des}} = 0.003\) \(b_{\text{max}} = 0.02\) | 30 seconds | E-mail, Paging & Fax | \(1 - e^{-\frac{4.6b}{0.02}}\)
|            |               |                              |                           |         | \(u_{\text{max}} = 0.99\) |
| 4          | II (Non-Real-Time) | \(b_{\text{max}} = 0\) \(b_{\text{des}} = 0.1\) \(b_{\text{max}} = 0.5\) | 3 minutes | Remote Login & Data on Demand | \(1 - e^{-\frac{4.6b}{0.5}}\)
|            |               |                              |                           |         | \(u_{\text{max}} = 0.99\) |
| 5          | II (Non-Real-Time) | \(b_{\text{max}} = 0\) \(b_{\text{des}} = 1.5\) \(b_{\text{max}} = 10\) | 2 minutes | File Transfer & Retrieval Service | \(1 - e^{-\frac{4.6b}{10}}\)
|            |               |                              |                           |         | \(u_{\text{max}} = 0.99\) |
5.7 SIMULATION RESULTS

The simulator employs discrete-event simulation to model the traffic and its management in wireless networks. A discrete-event simulation is one in which the state of the model changes at only a discrete set of simulated time points. With discrete event simulation a real network system is decomposed into a set of separate components. The fundamental element of the simulation is event and the operation of the system is represented as a chronological sequence of events. Each event is assigned a time stamp and takes place on a specific component.

The result of this event can be a message passed to one or more other components. On arrival at the other components, the content of the message may result in the generation of new events to be processed at some future logical time. The simulation parameters are assumed as follows and the results are given in Figure 5.6, 5.7 and in Table 5.1 and 5.2.

![Bandwidth Adaptation](image)

**Figure 5.9 Effect of Utilized Bandwidth**
Figure 5.10 Efficiency of Bandwidth Adaptation
Table 5.2 Comparison of Bandwidth Utilization between proposed and existing method

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Required BW (Mbps)</th>
<th>Available BW (Mbps)</th>
<th>Utilized Bandwidth Without Adaptation (Mbps)</th>
<th>Utilized Bandwidth With Adaptation (Mbps)</th>
<th>Efficiency % Without Adaptation</th>
<th>Efficiency % With Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All calls are Adaptive real time traffic (ART)</td>
<td>30</td>
<td>10</td>
<td>6</td>
<td>20</td>
<td>20</td>
<td>33.33</td>
</tr>
<tr>
<td>60% calls (ART), 20% calls hard real time (HRT), 20% calls Non-Real time (NRT)</td>
<td>18.75</td>
<td>10</td>
<td>6.75</td>
<td>9.75</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>40% calls (ART), 40% calls (HRT), 20% calls (NRT)</td>
<td>13.25</td>
<td>10</td>
<td>7.25</td>
<td>9.95</td>
<td>54.72</td>
<td>75.09</td>
</tr>
<tr>
<td>20% calls (ART), 60% calls (HRT), 20% calls (NRT)</td>
<td>7.75</td>
<td>10</td>
<td>7.75</td>
<td>7.75</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Thus, the proposed hybrid approach of assigning Bandwidth Adaptation scheme based on application utility function and PWA to data packets which are transferred from MAC layer efficiently reduces the delay between node to node data transmission. Simulation results have clearly demonstrated the superior performance of the proposed joint cross-layer schemes such as Bandwidth Adaptation with Prioritized Weight Assignment (PWA) by comparing them with some existing ones.