A New Call Admission Control scheme for Real-time traffic in Wireless Networks

Maneesh Tewari and H.S. Jamadagni
Center for Electronics Design and Technology,
Indian Institute of Science,
Bangalore, 560012 INDIA
Ph: 080-360-0810
tewari@lycos.com
hsjam@cedt.iisc.ernet.in

Abstract—In this paper, we introduce a framework to explain an adaptive call admission control for real time traffic of a Wireless network in order to improve the system performance in terms of Call Blocking Probabilities. The key contribution of the work explained in this paper constitutes the application of a transcoding scheme of the Real-time voice for a call admission controller. The advantage of employing the adaptive call admission control is illustrated by performance curve showing the reduction in the overall call blocking probability of the system. The proposed scheme attempts to supplement existing call admission control schemes, rather to replace them.

Keywords—Call admission control, Real-time traffic, QoS , Base station, Call blocking probability.

1. INTRODUCTION

With the deployment of packetized wireless networks, the need for Quality of Service (QoS) has become important. For efficiently supporting the QoS, a number of key areas in wired - wireless integration need to be addressed. One key issue is call admission policy for the base station scheduler (BSC).

The call admission problem for a wireless packet switched network supporting homogeneous applications, such as a cellular voice network, has been extensively investigated in the past [1]. The focus of this paper is to investigate the call admission policy for a wireless system supporting heterogeneous real-time constant-bit-rate (CBR) applications with different quality of service (QoS) requirements and traffic characteristics.

While IP network is Packet switched (PS), 1G and 2G cellular networks are typically Circuit switched (CS). Therefore, the wired and wireless network integration will distinguish three different types of service: CS services , PS real-time (RT) and PS non real-time services (NRT). Presently, CS services are voice calls from any CS network [3], while RT and NRT traffic are from IP network. Therefore the bandwidth available in a cell must be shared by calls of these different service classes as depicted in Figure 1, so that the different service requirements can be met.

In this work we focus on the Call Admission Control (CAC) policy for RT traffic, which is necessary to support the QoS guarantees in a BSC. Much work in this area has focused on cellular voice networks predominantly with voice traffic and with fixed constant bandwidth for all calls. This work aims to contribute towards the development and analysis of a control policy for admitting RT flows into a BSC, where each flow has different requirements in terms of bandwidth as quality measure. However, for our work, we consider wireless links as bottlenecks in the domain, and the Service Level Agreement (SLA) is negotiated mainly based on the bandwidth resource availability in the wireless network.

In a cellular cell, before a mobile session starts, the mobile user needs to specify its desired performance requirements by a ‘QoS Profile’. Then an admission controller (AC) decides to accept or reject the user’s request based on its QoS profile and the current network state. Purpose of the AC is to guarantee the QoS requirements of the user who requested admission while not violating the QoS profiles of already admitted users. At AC the call admission criteria is different for each service class. Admission control of RT sessions is based on a QoS profile that specifies guaranteed bit-rate that should be provided to the application to work properly. If the desired bandwidth requirements can not be satisfied by the network, the corresponding admission request is rejected. NRT sessions are admitted based on the current resource available for the NRT traffic [3].

For this work, we develop the notion of QOS for multimedia traffic in terms of maximum call blocking probabilities (CBP) for a given bandwidth partition, as mentioned in [4].

Considering all user classes of equal priority, forced-termination of an ongoing call is less desirable than the blocking of call [5]. From the user’s point of view, it is better to be blocked at the beginning of a call than to be dropped in the middle of one. As a result, handoff calls should be serviced with higher priority than new calls. To further decrease the handoff dropping probability, we use the guard channel scheme to reserve a fixed percentage of each BS’s resources.
for handoff calls [6], as can be seen in Figure 1 that $B_H$ is separately reserved for handover calls.

We partition the available bandwidth in a cell to meet the QoS requirements of various service classes such as voice calls, RT sessions, and NRT sessions. The bandwidth partitioning constitutes the rationale behind the admission control. Figure 1 illustrates the partitioning of the available bandwidth into different areas taking B as the overall bandwidth available in one cell. A portion $b_r$ of the bandwidth B is exclusively reserved for handover calls from neighboring cells in order to reduce handover failures. A portion $b_t$ is reserved for RT and $b_nr$ is reserved for non RT data packets. The remaining bandwidth $(1 - b_r - b_t - b_{nr})B$ is allocated for the CS voice calls. For simplicity, we assume that the complete partitioning policy for scheduling mechanism among the service classes is used, so that the admission control of each class can be considered separately [7].

In a wireless network, since bandwidth availability fluctuates frequently due to user mobility and channel quality variations, an adaptive service is very important to improve resource utilization efficiency.

2. A NEW APPROACH

We propose a scheme, confined to the bandwidth portion which is particularly reserved for the RT-sessions only. This scheme is applicable to any CAC scheme which partitions the total available bandwidth in a cell and reserves separately some portion of it for the RT sessions only $B_{RT}$, and can fit into that area. The RT traffic is more sensitive to the delays and less sensitive to losses [8]. In our scheme we will exploit the robustness of the RT traffic against losses and the method of transcoding to compress a RT traffic stream from one bit rate to another of a lower bit rate.

We assume following parameters:

B is the total bandwidth of the BS, and we assume it to be constant over time.

$B_{RT}$ is the bandwidth allotted to RT traffic, $B_{RT}(t)$, a discrete parameter, is the bandwidth consumed by the RT traffic at any instant of time $t$.

$th_1, th_2, \ldots, th_n$ are the various bandwidth threshold levels where different transcoding with higher compression starts, and $0 < th_1 < th_2 < \ldots < th_n < B_{RT}$

Let there be a total of K different bit rate classes of bit rate requirement $r_k$ for $k = 1, \ldots, K$ respectively and $r_1 > r_2 > \ldots > r_n$, and $bw_{req}$ is the bandwidth required by the newly arrived call.

The Algorithm

1. while ($B_{RT}(t) + bw_{req} <= B_{RT}$), perform steps (2) to (4).
2. else reject all the incoming arrivals
3. while ($0 < B_{RT}(t) + bw_{req} <= th_1$), accept the call of any class and update $B_{RT}(t)$
4. for ($j = 2$ to $j = n$)
   repeat
      if (arrived call rate > $r_j$) transcode the call to $r_j$ and accept it
      else reject all the incoming arrivals
5. while ($th_{j-1} < B_{RT}(t) + bw_{req} <= th_j$).
6. while ($th_n < B_{RT}(t) + bw_{req} <= B_{RT}$), accept only the lowest class calls as it is and update $B_{RT}(t)$; reject all other class calls arriving in the system.

We can define various algorithms of such category depending upon the exact action to be taken. One algorithm of such category is the following, which we implement for our simulation.

Implementation of the Algorithm — For simplicity, we take two kinds of incoming calls $r_1, r_2$ and two threshold parameters $th_1, th_2$.

1. while ($B_{RT}(t) + bw_{req} <= B_{RT}$), perform steps (2) to (4); else reject all the incoming arrivals
2. while ($0 < B_{RT}(t) + bw_{req} <= th_1$), accept all the calls of both classes and update $B_{RT}(t)$;
3. while ($th_1 < B_{RT}(t) + bw_{req} <= th_2$), do not accept the highest data rate call as it is, transcode the highest data rate call to the lower data rate call and then accept it; accept the other class calls as they are and update $B_{RT}(t)$;
4. while ($th_2 < B_{RT}(t) + bw_{req} <= B_{RT}$), accept only the lower data rate calls and update $B_{RT}(t)$; else, reject all other class calls arriving in the system;

According to the algorithm, we define the call admission functions $f_1$ and $f_2$ for data rate 1 calls and data rate 2 calls respectively as following:

$$f_1(n) = \begin{cases} 1 & : \bar{r} \bar{n} + r_2 <= th_2 \\ 0 & : \text{otherwise} \end{cases}$$

$$f_2(n) = \begin{cases} 1 & : \bar{r} \bar{n} + r_2 <= BW_{RT} \\ 0 & : \text{otherwise} \end{cases}$$
where \( \bar{r}, \bar{n} = r_1, n_1 + r_2, n_2 \).

Figure 2 shows the significance of existence of the two thresholds \( th_1 \) and \( th_2 \). We see that \( th_1 \) and \( th_2 \) divides \( BW_{RT} \), the bandwidth available for RT traffic, among 3 zones. Zone 1 gives free access to both kinds of call arrivals. Zone 2 accepts the calls of data rate \( r_1 \) after transcoding this call to data rate \( r_2 \) and accepts calls of data rate \( r_2 \). Zone 3 accepts only calls of data rate \( r_2 \) and rejects calls of data rate \( r_1 \).

![Figure 2. Partition of the RT Bandwidth in the BS](image)

3. THE QUEUING MODEL

We are interested in finding the CBPs of both kinds of arriving calls: \( CBP_1 \) and \( CBP_2 \). To make the system tractable for analysis purpose, we have taken following sample values given in Table 1 as an example:

**Assumptions**

1. Call arrival process is Poisson with mean arrival rate \( \lambda_k \).
2. Service rate of a call is exponentially distributed with mean \( \mu_k \).
3. The call arrival and service processes are independent of each other.
4. Voice Activity Detection (VAD)[9] and statistical multiplexing have not been considered.

**Queuing Analysis**

The system is modeled as an irreversible, continuous time, 3-dimensional Markov Chain, therefore its closed form solution does not exist [10]. Now, we can find the state transition matrix \( Q \) for the Markov chain, and then solve it for the \( \pi \) vector with the help of the following two equations:

\[
\pi \cdot Q = 0 \quad (1)
\]
\[
\sum_{i=1}^{n} \pi(i) = 1 \quad (2)
\]

\( \pi \) is the stationary distribution with ‘i’ as a vector.

The Q matrix is the state transition matrix of the 3-dimensional Markov chain with stationary probability distribution \( \pi(n_1, n_1', n_2) \) where \( n_1 \) is the units of resources (i.e. amount of bandwidth) consumed by the number of users of data rate \( r_1 \), \( n_1' \) is the units of resources consumed by the number of users of data rate \( r_1 \) which are transcoded to data rate \( r_2 \) before accepting by the system, and \( n_2 \) is the units of resources consumed by the number of users of data rate \( r_2 \) in the system at any given instant of time.

By solving above equations 1 and 2, we get the value of all \( \pi(n_1, n_1', n_2) \), now we can find out the respective CBPs for both kind of calls. Therefore, for this system, the overall CBP of the system will be

\[
CBP = \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \cdot CBP_1 + \frac{\lambda_2}{(\lambda_1 + \lambda_2)} \cdot CBP_2 \quad (3)
\]

We can generalize this formula for \( n \) calls. For a system having \( n \) different kinds of call, and all calls generating with equal probability, then the overall CBP can be found by using formula:

\[
CBP = \sum_{i=1}^{n} \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \cdot CBP_i \quad (4)
\]

4. PERFORMANCE IMPROVEMENT DUE TO THE ADAPTIVE FRAMEWORK

**Simulation Environment**

In the simulation, we consider a BS in a cell. We assume that calls arrive at the BS according to Poisson process with rate \( \lambda \). Before a new mobile session can start, it has to pass the AC introduced in earlier sections. The amount of time a mobile user with an ongoing session remains within the cell is called dwell time. If the session is still active after the dwell time, a handover towards an adjacent cell takes place. We assume the duration of RT sessions in a cell to be exponentially distributed with mean \( \mu \). Also, all calls are being generated with equal probability.

The simulation program has been written in C and is implemented on a Linux platform. A session of a mobile user can be completely specified by the following parameters: service class (e.g. RT, NRT, CS-voice), packet arrival process, dwell time, session duration, and QoS profile. In the simulation en-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B&lt;sub&gt;RT&lt;/sub&gt;</th>
<th>th&lt;sub&gt;1&lt;/sub&gt;</th>
<th>th&lt;sub&gt;2&lt;/sub&gt;</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value (units)</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Table 1. System Parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B&lt;sub&gt;RT&lt;/sub&gt;</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( \mu_1 )</th>
<th>( \mu_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Table 2. Simulation Parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
vironment, we specify a session of mobile user with RT service class, Poisson arrival process, exponentially distributed dwell time, and the bandwidth required. The parameters used in the simulation are shown in Table 2.

![Figure 3](image.png)

**Figure 3.** Comparison of Theoretical and Simulation values for CBP of data rate 1 call

![Figure 4](image.png)

**Figure 4.** Comparison of Theoretical and Simulation values for CBP of data rate 2 call

![Figure 5](image.png)

**Figure 5.** Comparison of Theoretical and Simulation values for CBP of the system

**Results and Discussion**

Using simulation experiments, we present the performance of our RT-CAC scheme using our adaptive framework and compare it to the same system without the proposed scheme. We also compare our simulation results with the theoretical results obtained from its queuing analysis. We take the sample values for various system parameters as mentioned in Table 2, and analyze the problem theoretically as mentioned in section 3. Then we compare the results of the theoretical analysis with the results of our simulation program. We find that the results closely match as can be seen in Figures 3, 4 and 5.

Table 3 shows various CBPs of the system without using our scheme. Table 4 shows the same parameters using our scheme in the system for different sets of thresholds. One can compare both the tables and see that performance of the system improves considerably for all sets of thresholds, and at the same time, the overall CBP of the system decreases. Same comparison is also represented pictorially in Figure 6.

Our results are intuitively appealing. We compare Table 3 with Table 4 and observe that by using our policy the performance of the system improves considerably. By using this scheme, CBP of the system decreases from 0.5 up to 0.33 which corresponds to an improvement up to 34% as shown in Figure 6.

Figure 7 shows the values of CBP of rate \( r_1 \) call. We see that without using this scheme the value of CBP1 is 0.63, and with this scheme the value varies from 0.33 to 0.78. For certain sets of thresholds \((th_1, th_2)\) such as \((2,6), (3,5), (3,6), (4,6), (5,6)\), CBP of rate1 call may increase and for rest it decreases showing improvement. Figure 8 also shows improvement in the CBP2 except certain set of thresholds of \( th_1 \) and \( th_2 \) such as \((5,8), (6,8)\) and \((7,8)\).

![Figure 6](image.png)

**Figure 6.** overall CBP of the system with and without using our scheme

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CBP1</th>
<th>CBP2</th>
<th>CBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.63</td>
<td>0.36</td>
<td>0.50</td>
</tr>
</tbody>
</table>

**Table 3.** CBP without using our scheme

<table>
<thead>
<tr>
<th>Threshold1</th>
<th>Threshold2</th>
<th>CBP1</th>
<th>CBP2</th>
<th>CBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>0.66</td>
<td>0.13</td>
<td>0.39</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.52</td>
<td>0.19</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>0.33</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.78</td>
<td>0.10</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0.66</td>
<td>0.13</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.52</td>
<td>0.19</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0.68</td>
<td>0.14</td>
<td>0.41</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.53</td>
<td>0.20</td>
<td>0.36</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.75</td>
<td>0.21</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>0.61</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0.61</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
</tbody>
</table>

**Table 4.** CBP with our scheme
We observe that, the scheme divides $B_{RT}$ into 3 zones as shown in Figure 2. We can interpret the meaning of these zones as follows: **Zone 3** gives exclusive access to 2nd type of calls and rejects all 1st type of calls, therefore CBP of 1st call is directly proportional and CBP of 2nd call is inversely proportional to the size of Zone 3. Thus setting the size of Zone 3 requires a tradeoff between the CBPs of both type of calls. **Zone 1** and **Zone 2** both gives access to both kind of calls but resource consumption is faster in Zone 1 than Zone 2. Therefore size of Zone 1 and Zone 2 should be decided considering the fact that Zone 2 causes the QoS degradation of 1st type of call due to transcoding of it into 2nd type of call.

Setting $th_2$ equal to $B_{RT}$ eliminates the existence of Zone 3, and hence a natural increase in the CBP of 2nd type of call. In this case the performance of 1st type of call increases since now it has whole $B_{RT}$ to access as can be seen in the Table 4. Therefore, the scheme demands a tradeoff among the CBPs of all types of calls. Exact choice of the thresholds depends on the choice of network administrator, in order to get the maximum benefits from the system. For example, we see in Table 4 that, if we have to make a choice for CBP1, CBP2, and CBP, between the triplets (0.66, 0.13, 0.39) and (0.39, 0.38, 0.38) then, to decide the final allocation of bandwidths we need other parameters like the revenue earned by each class of traffic.

**Features of the scheme**

- It performs better than those schemes which do not implement it.
- It can be fit to any of the Call Admission Scheme as a specialized layer for the Real-time traffic.
- It requires an *adaptive multirate transcoder* at BS to convert the audio stream of one data rate to another possible data rate.

- In this scheme, due to transcoding at the base station a small amount of delay is added.
- Control messages are needed to communicate the transcoding scheme between the BS and the user terminal.

### 5. Conclusions

We have introduced an adaptive framework for the CAC of real-time traffic. We have utilized transcoding and given a simple algorithm for CAC of the Real-time traffic in a wired to wireless integration network to increase the system throughput. Simualtion as well as theoretical analysis is conducted to highlight the advantages of the adaptive framework in comparison with non-adaptive framework. In future multimedia applications with adaptive framework will become widespread [5]. Therefore the basic ideas underlying the adaptive framework can also be applied for the adaptive control of wire-line multi-services IP networks also.

### References


