QoS Support for the Reverse Packet Data Channel in Third Generation (3G) Wireless Networks

Patrick Hosein and Tao Wu
Ericsson Wireless Communications Inc.
5012 Wateridge Vista Dr., San Diego, CA 92121
Email: {patrick.hosein, tao.wu}@ericsson.com

Abstract—Third Generation (3G) wireless networks have been designed to provide high speed data services on both forward and reverse links. In addition to increased data capacity they also support QoS (Quality of Service) guarantees for user applications. Users may run multiple applications (service instances) simultaneously and QoS guarantees may be required for one or more of them. In order to maintain these QoS guarantees the radio resources assigned to the mobile must be dynamically adjusted. Since mobile resources are assigned by the Base Station (BS) then each mobile must report relevant information to the BS so that it can make appropriate resource allocation decisions. The BS may either time-share or code share these resources. In this paper we present a simple framework to maintain QoS guarantees for each service instance of each mobile.

I. INTRODUCTION

The Third Generation (3G) of wireless networks, [1], [2], all support high speed data services. In addition, they have the ability to provide Quality of Service (QoS) guarantees for the applications they transport. Users may run multiple applications (service instances or SI) simultaneously and each of these may require different attribute (jitter, delay, throughput, error loss) guarantees. Furthermore, different user subscriptions (e.g. Gold, Silver, Bronze) may be supported and each class may be provided with varying control of these QoS guarantees.

Several papers [3], [4], [5], [6] have been written on scheduling for the Forward Packet Data channels in 3G networks since in general the forward link tends to form the throughput bottleneck. Papers have also been written on reverse link scheduling [7], [8], [9], [10]. Since future applications are predicted to also place heavy demands on the reverse link (video conferencing, multimedia messaging, Voice over IP, etc.), efficient management of the reverse link resources is essential. Therefore recent standards efforts [1] were focused on improving the reverse link performance. Since different groups were assigned to the physical layer enhancements and the MAC layer enhancements then the linkage between the two was not evaluated in detail.

In this paper we address the linkage between the QoS needs of the various SIs of a mobile and the management of the finite resources available on the reverse link air interface. We first present the framework assumed (based on 1xEV-DV Rev D) and show how the various QoS guarantees can be mapped into reverse link rate demands. We then present a BS scheduler for assigning reverse link resources (rate requests) so as to maintain user QoS guarantees.

II. MODEL AND TERMINOLOGY

Assume that the IP packets of each SI are split into multiple RLP (Radio Link Protocol) frames and placed at the tail of its RLP buffer. The RLP frames in this buffer are served in a First In First Out (FIFO) fashion. If the transmission of a RLP frame fails then it is re-transmitted at the physical layer. This is repeated if necessary for some specified number of times. Such RLP frames are placed in a re-transmit buffer and are given strictly higher service priority over new frame transmissions.

The physical layer frame consists of one or more RLP frames. The number of data bits per physical layer frame depends on the reverse link rate of the mobile. If the first transmission of a physical layer frame is unsuccessful then it is re-transmitted up to a maximum of two re-transmissions. These re-transmissions are combined with earlier transmissions (soft combining) to increase its probability of success. The transmission power of each transmission is varied so that the residual FER (after all transmissions) is at most 1%. Note that the 1% failures are retransmitted at the RLP layer but with a fresh rounds of physical layer transmissions.

In the following description the subscript \(i\) is used to identify the SI while the index \(n\) denotes the control period. On a periodic basis (with period \(\tau\) which is assumed to be a multiple of a frame duration), the mobile counts the number of information bits, \(b_i(n)\), that departed the original transmission queue during period \(n\) (the prior period) for SI \(i\). It also determines the size of this queue, \(q_i(n)\) (in units of information bits). The (filtered) information bit departure rate of the original transmission buffer for this service instance is then estimated by

\[
\mu_i(n) = \alpha_i \frac{b_i(n)}{\tau} + (1 - \alpha_i)\mu_i(n-1)
\]

where \(\alpha_i\) is the corresponding filter constant and \(\mu_i(0) = 0\).

Next note that the number of information bits that arrive during this period is given by \(b_i(n) + q_i(n) - q_i(n-1)\). Therefore the (filtered) information bit arrival rate is given by

\[
\lambda_i(n) = \alpha_i \frac{b_i(n) + q_i(n) - q_i(n-1)}{\tau} + (1 - \alpha_i)\lambda_i(n-1)
\]

where \(\alpha_i\) is the corresponding filter time constant and \(\lambda_i(0) = 0\).
At the beginning of control period \( n \) we assume that SI \( i \) requests a rate \( \tilde{r}(n) \) and is assigned (after some delay) a rate \( r_i(n) \) by the BS for the subsequent control period. The difference between \( r \) and \( \tilde{r} \) is due to the finite number of rates supported in the reverse direction.

We assume that once an RLP frame departs the original transmission queue it eventually arrives successfully at the BS. In other words we assume an unlimited number of RLP re-transmissions. Under this assumption, this information bit departure rate, \( \rho_i \), equals the goodput of the channel since it is the rate of successfully received frames. If we denote the overhead incurred (due to physical and RLP layer re-transmissions) in transmitting an RLP frame over the air interface by \( \rho \) then we have

\[
\mu_i = (1 - \rho) r_i
\]

Therefore at the end of period \( n \) we can estimate \( \rho \) by:

\[
\rho(n) = \alpha_r \left( 1 - \frac{\sum_i \mu_i(n)}{R(n)} \right) + (1 - \alpha_r) \rho(n - 1)
\]

where \( \alpha_r \) is the filter constant, \( \rho(0) = 0 \) and \( R(n) = \sum_i r_i(n) \) is the reverse link rate. This overhead estimate will be used by all SIs and hence the goodput of SI \( i \) during period \( n \) is estimated by

\[
g_i(n) = (1 - \rho(n)) r_i(n)
\]

At each decision point (i.e., every \( \tau \) seconds) the objective is to determine the service rate, \( \tilde{r}(n+1) \), that should be applied in the subsequent interval so that at the end of the interval the expected value of the concerned QoS attribute equals the desired value. Note that this is computed for each SI and hence the request that is sent to the BS is obtained by summing over all SIs.

### III. RATE DETERMINATION FOR EACH QoS ATTRIBUTE

In [11], the above model was used to determine the rate that is required in the subsequent control period in order to maintain the corresponding QoS attribute near its desired value. More precisely, the rate is chosen so that the expected value of the attribute at the end of the subsequent period equals the desired value. Whenever a single SI is being analyzed we drop the subscript \( i \) for simplicity. In [11] it was found that the goodput guarantee can be maintained if the rate in the subsequent period was set to

\[
\tilde{r}(n+1) = \frac{g}{1 - \rho(n)}.
\]

For jitter guarantees the corresponding rate is given by

\[
\tilde{r}(n+1) = \frac{q(n) + \lambda(n) \tau}{(1 - \rho(n))(\tau + d_T)}.
\]

Finally for guarantees on the maximum delay the corresponding rate is given by

\[
\tilde{r}(n+1) = \frac{q(n^2) + \lambda(n) \mu(n) d_{max}}{(1 - \rho(n)) d_{max} \mu(n)}.
\]

### IV. THE SCHEDULER

On a periodic basis (every frame duration) each mobile sends a \( N \)-bit report to the base station. In the 1xEV-DV standard these bits contain queue length and power headroom (the amount of power available for potential rate increases). We propose a simpler scheme in which only two bits of information is reported. These two bits indicate whether the mobile wants to decrease its rate by one level, keep it as its present level, increase its rate by one level or increase its rate by two levels. The base station then uses these rate requests, together with the radio conditions (geometry) of each mobile to determine what rates to assign for the subsequent control period.

If we denote the mobile’s desired rate by \( \tilde{r} \) and the present rate by \( r \) then if \( \tilde{r} \) is two or more rate levels above \( r \) then “11” is reported else if it is at least one level above \( r \) the “10” is reported else if it is at least one rate level below \( r \) then “00” is reported otherwise “01” is reported. Let \( b \) denote the value of the reported bits (i.e., \( b \in \{0,1,2,3\} \)). The base station computes a smoothed average \( \tilde{b} \) of this report as follows:

\[
\tilde{b} = \alpha b + (1 - \alpha) \tilde{b}
\]

where \( \alpha > 0 \) is the smoothing factor.

We assume that the base station can estimate the maximum rate that each mobile can support in the subsequent control period. This is determined based on the geometry of the mobile (which is estimated from forward link channel quality reports made by the mobile for scheduling on the forward link). Let \( d_i(n) \) denote the maximum achievable rate of mobile \( i \) during period \( n \). The scheduler then determines the priority of each mobile using the priority function

\[
P_i(n) = \frac{d_i(n)}{3 - b_i(n)}
\]

This means that users in good radio conditions and/or with urgent rate requests (\( \tilde{b} \) close to 3) are given high priority.

We assume that, given the geometry of each user, the base station can compute the total reverse link loading for a given set of mobile rates. This reverse link loading must be kept below a specified value. Once user priorities have been determined the base station assigns rates as follows. First it tentatively reduces all mobile rates by one level. It picks the highest priority user and checks to see if satisfying that user’s instantaneous rate request (based on \( b(n) \)) maintains the reverse link loading below capacity. If this is the case then the rate for that user is adjusted accordingly and the process is repeated for the next highest priority user. If this is not the case then the user is assigned the highest possible rate that maintains the reverse link loading below capacity and the scheduling process is stopped. The remaining mobiles (those not specifically scheduled) are assigned their tentative rates.

This scheduling procedure attempts to serve users when they are in good radio conditions. However, those users that have been consistently asking for higher rates (in order to maintain its QoS guarantees) will be given ever increasing priority because its value of \( b \) will approach 3 which causes an
exponential increase in its priority. This approach (commonly called a barrier function approach since it prevents motion into unfeasible regions) has been used in other QoS aware schedulers [6].

V. SIMULATION METHODOLOGY AND RESULTS

The performance of streaming video over the Reverse PDCH is simulated in this section for the 1xEV-DV system. We first summarize our simulation assumptions, traffic model and performance criteria. We then present simulation results with the proposed scheduling algorithm. These results show an enhanced performance on QoS support compared with using Reverse Supplemental Channels (R_SCH).

A. Simulation Assumptions

The simulation environment and assumptions comply with the 1xEV-DV Evaluation Methodology [12] documents defined in 3GPP2, with the following parameters. We assume a layout of 19 cells with 3 sectors per cell. We use the channel models provided in Table I, a frame size of 10ms and 4 H-ARQ channels with a maximum of 3 transmissions with Chase combining. The overhead from the reverse link channel quality indicator channel (used for forward link transmissions) is also simulated. Data rates of 19.2, 40.8, 79.2, 156, 309.6, 463.2 and 616.8 kbps are used with a typical autonomous rate of 19.2 kbps. Inner and outer loop power control as well as soft hand-off is simulated. We use the Dedicated Rate Control algorithm for reverse link resource management.

B. Near Real Time Video Model

A near real time video model [12] is adopted in the simulation for video streaming traffic on the reverse link. The video packet arrives at a deterministic interval of 100ms. Each frame contains 8 slices (packets). The size of these packets is distributed as a Truncated Pareto with a maximum of 125 bytes and a mean value of 50 bytes. This corresponds to an average source video rate of 32 kbps (8 packets with an average of 50 bytes every 100ms). A 5 seconds de-jitter buffer window is used for this video streaming service to ensure a continuous display of video streaming data. Ideally, the mobile should maintain the buffer level in the middle so that the buffer will not overflow or run dry. This ensures continuous display of the stream at the destination. The base station scheduler (rate controller) should let the mobile transmit an entire video frame within 5 second after receiving it at the buffer. If a video frame exceeds the 5 second requirement, the remaining portion will be dropped. The fraction of dropped video packets is defined as the packet loss rate (PLR).

C. Performance Criteria

In packet data transmission systems, throughput and delay are the most important performance numbers. These as well as other metrics are defined as follows:

- Rise over Thermal (ROT) defined as the received power normalized by the thermal noise level. The probability that this exceeds 7dB should be at most 1%.
- System load $L_j$ of Base Station $j$ is defined as

$$L_j = \frac{\sum_{i: j \in A_i} \gamma_i}{1 + \gamma_i}$$

where $A_i$ is the active set of mobile $i$, and $\gamma_i$ is the Signal to Interference and Noise Ratio (SINR) of the signal from mobile $i$ to base station $j$ averaged over the most recently received frame.
- Data throughput of a mobile defined as the ratio of the total number of correctly received bits and the total simulation time.
- Video packet drop rate is defined as the fraction of video frames that are not completely transmitted within 5 seconds of their arrival at the mobile. It should be less than 2% for each individual mobile.
- Service coverage is defined as the percentage of mobiles with a drop rate less than 2% given the uniformly distributed users. The system should be engineered for a cell coverage of at least 96%.

D. Simulations Results with Dedicated Rate Control

In Section V we described a scheduling algorithm that can be used to maintain the QoS guarantees for each mobile service instance. In this particular case of video streaming we make one small modification to the scheduling scheme. Since the average mobile throughout equals the arrival rate which is constant then, instead of each mobile sending a rate request, we assume that it periodically reports a quantized version of the mobility report. In particular, if the expected delay is less than 2 it reports “-1”, else if it is less than 3 it reports “0” else if it is less than 4 it reports “1” else it reports 2. The BS filters these reports and uses the same priority function and scheduler as previously described. One other difference is that the BS adjusts the mobile’s rate by at most one level. Hence if the instantaneous value of $b$ is 2 then the mobile’s rate is increased by one (instead of two) levels. Note that all mobiles can independently transmit at any rate up to the Autonomous rate. Rates above this value must be assigned by the BS.

We assume that each mobile has a single service instance consisting of a video stream. These streams are modeled as previously described. We simulate 15 users per sector. We experimented with two autonomous rates, 19.2 kbps and 79.2 kbps. 19.2 kbps is the typical autonomous rate since it is

<table>
<thead>
<tr>
<th>Channel Model(%)</th>
<th>Multi-path Model</th>
<th>No. of Fingers</th>
<th>Speed (kmph)</th>
<th>Fading</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (30%)</td>
<td>Ped A</td>
<td>1</td>
<td>3</td>
<td>Jakes</td>
</tr>
<tr>
<td>B (30%)</td>
<td>Ped B</td>
<td>3</td>
<td>10</td>
<td>Jakes</td>
</tr>
<tr>
<td>C (20%)</td>
<td>Veh A</td>
<td>2</td>
<td>30</td>
<td>Jakes</td>
</tr>
<tr>
<td>D (10%)</td>
<td>Ped A</td>
<td>1</td>
<td>120</td>
<td>Jakes</td>
</tr>
<tr>
<td>E (10%)</td>
<td>Single Path</td>
<td>1</td>
<td>0</td>
<td>Rician</td>
</tr>
</tbody>
</table>

TABLE I
CHANNEL MODELS
<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VIDEO STREAMING PERFORMANCE RESULTS</strong></td>
</tr>
<tr>
<td>Autonomous Rate (kbps)</td>
</tr>
<tr>
<td>System Load</td>
</tr>
<tr>
<td>Ave ROT (dB)</td>
</tr>
<tr>
<td>Pr(ROT &gt; 7 dB)</td>
</tr>
<tr>
<td>RLP Throughput (kbps)</td>
</tr>
<tr>
<td>Average Drop Rate</td>
</tr>
<tr>
<td>Service Coverage</td>
</tr>
</tbody>
</table>

The least reverse link rate. We also tried 79.2 kbps since it closely matches the video streaming rate (after H-ARQ retransmissions). With the same target load 0.6, the average ROT and system load are all higher with 79.2 kbps autonomous rate. On the other hand, the 7dB overshoot with 79.2 kbps is lower than for the 19.2 kbps case. 79.2 kbps also achieves higher throughput but with the same service coverage. Both cases achieve 96% coverage.

The histograms of the data rates are plotted in Figure 1. With autonomous rate set to 19.2 kbps, the mobiles rates have a wide variance. With 79.2 kbps, the data rate stays close to the autonomous rate most of the time and only goes higher occasionally. In both cases the probability of using very high rates is negligible.

This difference is also reflected in the histograms for sector load shown in Figure 2. The smaller rate variation with the 79.2 kbps case leads to a lower load variance. Under the same ROT 7DB percentile constraint, this means that a higher average ROT can be used resulting in a higher user capacity.

For the 19.2 kbps case, a sample buffer evolution is plotted in Figure 3. We can see that most of the time the buffer size oscillates between 40% and 60% (corresponding to 2 and 3 seconds delay) which is the desired range for low jitter. The corresponding transmit data rate over the entire simulation run is plotted in Figure 4. Note that the mobile is scheduled to transmit periodically. Once it is allowed to transmit, the rate is relatively stable and oscillates between two or three consecutive rates. When the queue drops to a lower (satisfied) level, the mobile will stay at the autonomous rate so that others will in turn use the reverse link resource to transmit.

The buffer evolution for the 79.2 kbps case is plotted in Figure 5. The oscillation is eliminated and changes are more gradual. Figure 6 plots the corresponding transmit data rate. The data rate is very stable and only moves between 79.2 kbps and the next rate level, 156 kbps. This reduced rate oscillation is the reason for the narrower system load distribution.

In conclusion, with a low autonomous rate (e.g., 19.2 kbps), the scheduler rotates service among the users (TDM control). On the other hand, with high autonomous rates (e.g., 79.2 kbps), all users maintain a roughly constant rate (CDM control).
control). From the simulations, the later scheme exhibits better performance.

VI. CONCLUSIONS

In this paper we presented a simplified framework for supporting QoS services on the reverse link of a CDMA network. For each QoS attribute we show how a corresponding desired rate can be computed for each service instance. The rate that the mobile requests from the BS is then the sum of these rates over all SIs. This results in very low signaling overhead since two bits of signaling per control period are required. We then showed how one can try and achieve this total rate per mobile by an appropriate choice of the utility function used by the scheduler. This utility function performs a barrier to buffer overflow thus reducing the FER of bad radio condition users.

We then considered an example in which all users are running streaming services in the reverse link direction. Using a slightly modified mobile report scheme we show that, for suitable performance criteria (at least 96% of the mobiles achieve a packet loss rate no greater than 2%) the system can support as many as 15 of these users. Note that in 2G networks, the reverse link data is transported over power controlled synchronous channels. A channel rate of 4X rate (38.4 kbps without the fundamental channel) would be required to support a 32 kbps stream. However, typically no more than about six such channels can be supported simultaneously. This implies that the R_PDCH has twice the user capacity than CDM channels. This capacity gain is similar to the gains obtained with other traffic types.

REFERENCES