Delay-Prioritized Scheduling (DPS) for Real Time Traffic in 3GPP LTE System

Kumbesan Sandrasegaran, Huda Adibah Mohd Ramli and Riyaj Basukala
Faculty of Engineering and Information Technology
University of Technology
Sydney, Australia
[kumbes, HudaAdibah.MohdRamli]@eng.uts.edu.au
Riyaj.Basukala@uts.edu.au

Abstract—Given that the co-existence of multimedia applications will be a norm in the future wireless systems, their quality of service (QoS) requirements need to be guaranteed. This has imposed new challenges in the design of packet scheduling algorithms in these systems. To address those challenges, a new packet scheduling algorithm for real time (RT) traffic in downlink third generation partnership project long term evolution (3GPP LTE) system is proposed in this paper. The proposed algorithm utilizes each user’s packet delay information and its instantaneous downlink channel conditions when making scheduling decisions. Simulation results show that the proposed algorithm outperforms opportunistic scheduling and maximum-largest weighted delay first algorithms by maximizing system throughput and satisfying the QoS requirements of the RT traffic.

Keywords-3GPP LTE; Radio Resource Management; Packet Scheduling; Quality of Service; Real Time Traffic

I. INTRODUCTION

Third generation partnership project long term evolution (3GPP LTE) is a new radio access technology that enhances the legacy systems by providing higher peak data rates, fully packet switched core network and supporting diverse quality of service (QoS) requirements of multimedia applications. In the downlink, 3GPP LTE adopts orthogonal frequency division multiple access (OFDMA) technology due to its immunity to frequency selective fading of the radio channels and its robustness to the inter-symbol interference. The 3GPP LTE radio network architecture consists of only one node between the user and the core network known as eNodeB which is responsible to perform all radio resource management (RRM) functions. Packet scheduling is one of the RRM functions and it is responsible for intelligent selections of users and transmissions of their packets such that the radio resources are efficiently utilized and the users’ QoS requirements are satisfied.

Video streaming is one of the real time (RT) applications that need to be supported in the 3GPP LTE system. To ensure that the QoS requirements of video streaming users are satisfied; the packet loss rate has to be minimized by keeping it below a threshold. The packets of a video streaming user should be received by the user within its delay threshold; otherwise the packets will be discarded and hence considered as lost packets.

There are many packet scheduling algorithms developed to support the RT traffic in the multi-carrier wireless systems. For example, the algorithms proposed in [1-3] extend the well-known maximum-largest weighted delay first (M-LWDF) algorithm [4] into the OFDMA systems. The M-LWDF algorithm that utilizes instantaneous downlink channel conditions (e.g. signal-to-noise ratio, SNR), average throughput and packet information is considered as the best packet scheduling algorithm for the RT traffic in single carrier wireless systems. A method that decomposed packet scheduling of RT traffic in the multi-carrier wireless systems into resource allocation and resource assignment techniques is proposed in [5, 6]. The resource allocation technique determines the amount of radio resources that is to be allocated to each user and the resource assignment technique maps the available radio resources to the users. Other methods for scheduling RT traffic discussed in the literature include utilizing the mathematical model [7] and the use of utility functions [8].

Though, the packet scheduling algorithms discussed earlier achieve a good performance in the multi-carrier wireless system supporting RT traffic, they cannot be directly implemented into the downlink 3GPP LTE system that support video streaming applications due the following reasons: (i) the video streaming application is very sensitive to the packet loss and (ii) packet scheduling in the downlink 3GPP LTE system is performed at 1 ms interval (transmit time interval, TTI) and the radio resources that are available to be shared among users consist of groups of sub-carriers which are known as resource blocks (RBs). Therefore, a new packet scheduling algorithm known as delay-prioritized scheduling (DPS) is proposed in this paper for the downlink 3GPP LTE system supporting the video streaming applications. DPS algorithm aims to maximize system throughput while satisfying the QoS requirements of video streaming users by utilizing each user’s instantaneous downlink SNR values as well as its packet delay information.

The rest of this paper is organized as follows: Section II describes the downlink 3GPP LTE system model followed by detailed explanations on the proposed algorithm in Section III. Simulation environments are given in Section IV while Section
V contains results of the simulation. Finally, conclusions are given in Section VI.

II. SYSTEM MODEL

In this paper a cell of 5 MHz bandwidth made up of 25 RBs and 2 GHz carrier frequency is used. The serving eNodeB has a fixed location at the centre of the cell and it controls all the available RBs. These RBs are to be shared by all users within the cell. The downlink 3GPP LTE system parameters used in this paper are given in Table I.

In the system, users report their instantaneous downlink SNR values to the serving eNodeB at each TTI. Each user experiences different instantaneous downlink SNR value at each TTI and on each RB due to the frequency-selective fading nature of multi-path propagation and time-selective fading nature due to user movement respectively [5].

The reported instantaneous downlink SNR value is used to determine the data rate and hence the number of bits in two consecutive RBs that a user can support at each TTI. The approach proposed in [9] is used to compute the number of bits per symbol of user \(i\) at time \(t\) on a sub-carrier within RB \(j\) \((nbits_{ij}(t)/symbol)\). The user’s achievable data rate \((d_{r}(t))\) at time \(t\) on two consecutive RBs are determined using:

\[
d_{r}(t) = \frac{nbits_{ij}(t)}{symbol} \times \frac{nsymbols}{slot} \times \frac{nslots}{TTI} \times \frac{nsc}{RB}.
\]

where \(nsymbols/slot\) is the number of symbols per slot, \(nslots/TTI\) is the number of slots per TTI and \(nsc/RB\) is the number of sub-carriers per RB. The minimum instantaneous downlink SNR values and their associated achievable data rates used in this paper are given in Table II.

Each user is assigned a buffer at the serving eNodeB. Packets of a user arriving to the serving eNodeB are time stamped and queued in its buffer for transmission based on a first-in-first-out (FIFO) basis. For each packet in the queue at the serving eNode buffer, the head of line (HOL) packet delay (time difference between the current time and the arrival time of a packet) is computed. Packet is discarded if the HOL packet delay exceeds the delay threshold, as discussed in Section I.

III. DELAY-PRIORITY-SCHEDULED SCHEDULING

The video streaming application requires that the packet loss rate be kept below a threshold. This can be achieved by monitoring the values of the HOL packet delay of each user and schedules the packets that are approaching the delay threshold. Besides the packet delay information (the HOL packet delay and the delay threshold), the instantaneous downlink SNR value also plays an important role in packet scheduling. According to [10], the radio resources can be fully utilized if the packet scheduler exploits the current instantaneous downlink channel conditions when making scheduling decisions.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>5 MHz</td>
</tr>
<tr>
<td>Number of Sub-carriers</td>
<td>300</td>
</tr>
<tr>
<td>Number of RBs</td>
<td>25</td>
</tr>
<tr>
<td>Number of Sub-carriers per RB</td>
<td>12</td>
</tr>
<tr>
<td>Sub-Carrier Spacing</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Slot Duration</td>
<td>0.5 ms</td>
</tr>
<tr>
<td>Scheduling Time (TTI)</td>
<td>1 ms</td>
</tr>
<tr>
<td>Number of OFDM Symbols per Slot</td>
<td>7</td>
</tr>
</tbody>
</table>

The DPS algorithm is proposed to satisfy the QoS requirements of video streaming applications and maximize system throughput in the downlink 3GPP LTE system. When compared with the existing packet scheduling algorithms (discussed in Section I) that select each user based on packet delay information in addition to other scheduling criteria (e.g. instantaneous downlink channel conditions, average throughput, packet loss rate and etc.), this algorithm has a very low implementation complexity as only the packet delay information is utilized when selecting the users in each TTI.

The operation of the DPS algorithm can be divided into three steps. In Step 1, the algorithm computes for each user the remaining time of the HOL packet delay to approach the delay threshold \((d_{r}(t))\) described in (2). This step is performed in each scheduling instant.

\[
d_{r}(t) = T_{i} - W_{i}(t) \quad \forall \text{user}.
\]

where \(T_{i}\) is the delay threshold of user \(i\) and \(W_{i}(t)\) is the HOL packet delay of user \(i\) at time \(t\).

Thereafter, a user with the lowest \(d_{r}(t)\) (metric \(k\) in (3)) is selected in Step 2.

\[
k = \arg \min d_{r}(t) \quad \forall \text{user}.
\]

Once user \(k\) is determined in Step 2, Step 3 selects the user’s best RB (the RB with the highest instantaneous downlink SNR value reported by the user) from a list of available RBs, transmits the user’s packets on the selected RB and updates the user’s new \(d_{r}(t)\). The selected RB is then removed from the list of available RBs. Step 2 and 3 are repeated until the list of available RBs becomes empty.

### Table I. DOWNLINK 3GPP LTE SYSTEM PARAMETERS

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>5 MHz</td>
</tr>
<tr>
<td>Number of Sub-carriers</td>
<td>300</td>
</tr>
<tr>
<td>Number of RBs</td>
<td>25</td>
</tr>
<tr>
<td>Number of Sub-carriers per RB</td>
<td>12</td>
</tr>
<tr>
<td>Sub-Carrier Spacing</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Slot Duration</td>
<td>0.5 ms</td>
</tr>
<tr>
<td>Scheduling Time (TTI)</td>
<td>1 ms</td>
</tr>
<tr>
<td>Number of OFDM Symbols per Slot</td>
<td>7</td>
</tr>
</tbody>
</table>

### Table II. INSTANTANEOUS DOWNLINK SNR TO DATA RATE MAPPING TABLE

<table>
<thead>
<tr>
<th>Minimum Instantaneous Downlink SNR Value (dB)</th>
<th>Modulation and Coding</th>
<th>Data Rate (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>QPSK (1/2)</td>
<td>168</td>
</tr>
<tr>
<td>3.7</td>
<td>QPSK (2/3)</td>
<td>224</td>
</tr>
<tr>
<td>4.5</td>
<td>QPSK (3/4)</td>
<td>252</td>
</tr>
<tr>
<td>7.2</td>
<td>16 QAM (1/2)</td>
<td>336</td>
</tr>
<tr>
<td>9.5</td>
<td>16 QAM (2/3)</td>
<td>448</td>
</tr>
<tr>
<td>10.7</td>
<td>16 QAM (3/4)</td>
<td>504</td>
</tr>
<tr>
<td>14.8</td>
<td>64 QAM (2/3)</td>
<td>672</td>
</tr>
<tr>
<td>16.1</td>
<td>64 QAM (3/4)</td>
<td>756</td>
</tr>
</tbody>
</table>
Fig. 1 shows the flow chart of the DPS algorithm and a generalized model of the DPS algorithm in the downlink 3GPP LTE system is given in Fig. 2. In each TTI, the packet scheduler decides which users are to be scheduled based on the packet delay information obtained from each user’s buffer. The instantaneous downlink SNR values reported by the user are utilized in order to determine which RB is to be used for packet transmissions. In the DPS algorithm, there is a possibility that a user may be allocated zero, one or more than one RBs at each scheduling instant as shown in Fig. 2.

IV. SIMULATION ENVIRONMENTS

A computer simulation using C++ platform is conducted to evaluate the performance of the DPS algorithm. The system parameters, traffic parameters and performance metrics used in the simulation are described in the following sub-sections.

A. System Parameters

In this paper, a single hexagonal cell scenario with the downlink 3GPP LTE parameters as discussed in Section II is used. There are $K$ (20 to 100) video streaming users within the cell and they are uniformly located within the serving eNodeB. Users are constantly moving at speeds between 1-100 km/h in random directions. A wrap-around method [11] is employed at the cell boundary to ensure users always remain within the simulation area.

Several assumptions are made in this paper due to time limitations as well as in order to reduce complexity of the system simulation. These assumptions are consistent with other works [1, 12]. It is assumed that each user reports its instantaneous downlink SNR values on each RB and at each TTI to the serving eNodeB and this reporting is assumed to be error-free and delay-free. It is also assumed that equal downlink transmit power is allocated on each RB and all transmitted packets are received correctly by the receiver.

Pathloss [13], shadow fading [14] and multi-path fading [15] are used to determine the channel gain and hence the instantaneous downlink SNR value of each user on each RB and at each TTI. It is assumed in this paper that at any time instant pathloss and shadow fading are fixed on each RB. The channel gain ($Gain_{ij}(t)$) of user $i$ on RB $j$ at time $t$ is computed using:

$$Gain_{ij}(t) = 10^{\left(\frac{pl_i(t)}{10}\right)} \cdot 10^{\left(\frac{\xi_i(t)}{10}\right)} \cdot 10^{\left(\frac{mpath_{ij}(t)}{10}\right)}.$$  (4)

where $pl_i(t)$ and $\xi_i(t)$ are the pathloss (in dB) and shadow fading (in dB) of user $i$ at time $t$, respectively and $mpath_{ij}(t)$ is the multi-path fading (in dB) of user $i$ on RB $j$ at time $t$.

From the computed channel gain, the instantaneous downlink SNR value of user $i$ on RB $j$ at time $t$ ($\gamma_{ij}(t)$) is computed using the approach proposed in [16] given as follows:

$$\gamma_{ij}(t) = \frac{P_{tdd} \cdot Gain_{ij}(t)}{N(i + N_s)}.$$  (5)

where $P_{tdd}$ is the total eNodeB downlink transmit power, $N$ is the available number of RBs, $N_s$ is the thermal noise and $I$ is the inter-cell interference. Since only one cell is used in the simulation, the inter-cell interference is assumed to be a constant.

B. Traffic Parameters

The video streaming packets, arriving from the network, as seen by the serving eNodeB buffer are modeled using parameters given in Table III. The video frames arrive at a regular interval and consists of fixed number of packets. The packet sizes are of variable in length and are based on a
Truncated Pareto distribution. Similarly, the inter-arrival time of each packet in a frame follows the Truncated Pareto distribution. An example of video streaming model is shown in Fig. 3.

The packets are streamed into users’ buffers from variable bit rate (VBR) source encoders running at 128 kbps in average. The video streaming applications are assumed to be “played” as the packets are being streamed through the air interface, instead of being downloaded first and then played. In this paper, the threshold for HOL packet delay of each user is set to 20 ms which is the maximum waiting time of a video streaming packet at the serving eNodeB buffer [17]. Furthermore, less than 1% threshold for packet loss ratio (PLR) is recommended in [18] in order to satisfy the QoS requirement of this application. The buffer of each user is assumed to be infinite and a packet is considered lost when it is discarded.

C. Performance Metrics

The performance of the DPS algorithm is evaluated based on system throughput, average system delay, PLR and fairness as defined as follows:

The system throughput is defined as the total transmitted packets per second and can be mathematically expressed as:

\[ \text{system throughput} = \frac{1}{T} \sum_{i=1}^{K} \sum_{t=1}^{T} p_{\text{transmit}}(t). \]  (6)

where \( p_{\text{transmit}}(t) \) is the size of transmitted packets of user \( i \) at time \( t \), \( K \) is the total number of users and \( T \) is the total simulation time.

The average system delay is defined as the average HOL packet delay for the whole simulation time. It has the following mathematical expression:

\[ \text{average system delay} = \frac{1}{T} \sum_{i=1}^{K} \sum_{t=1}^{T} \frac{1}{K} \sum_{i=1}^{K} W_i(t). \]  (7)

where \( W_i(t) \) is the HOL packet delay of user \( i \) at time \( t \), \( K \) is the total number of users and \( T \) is the total simulation time.

<table>
<thead>
<tr>
<th>Information types</th>
<th>Distribution</th>
<th>Distribution Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-arrival time between the beginning of</td>
<td>Deterministic (Based on 20fps)</td>
<td>50ms</td>
</tr>
<tr>
<td>successive frames</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of packets (slices) in a frame</td>
<td>Deterministic</td>
<td>8</td>
</tr>
<tr>
<td>Packet (slice) size</td>
<td>Truncated Pareto (Mean=100bytes, max=125bytes)</td>
<td>K=40bytes, ( \alpha=1.2 )</td>
</tr>
<tr>
<td>Inter-arrival time between packets (slices)</td>
<td>Truncated Pareto (Mean=6ms, Max=12.5ms)</td>
<td>K=2.5ms, ( \alpha=1.2 )</td>
</tr>
<tr>
<td>in a frame</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III. PARAMETERS OF A VIDEO STREAMING APPLICATION [19]

Finally, fairness is defined as the difference in service levels (total sizes of the transmitted packets) between the most and the least served users over a given time frame [20] and is given below:

\[ \text{fairness} = 1 - \frac{\text{ptotaltransmit}_{\text{max}} - \text{ptotaltransmit}_{\text{min}}}{\sum_{i=1}^{K} \sum_{t=1}^{T} \text{psize}_{i}(t)}. \]  (9)

where \( \text{ptotaltransmit}_{\text{max}} \) and \( \text{ptotaltransmit}_{\text{min}} \) are the total size of the transmitted packets of the most and the least served users, respectively and \( \text{psize}_{i}(t) \) is the size of all packets that have arrived into eNodeB buffer of user \( i \) at time \( t \).

V. RESULTS

In this section, the simulation results of the DPS algorithm are presented and compared with the opportunistic scheduling (OS) [5] algorithm. The OS algorithm is chosen as it is one of the algorithms proposed to support the RT traffic in OFDMA systems, which is similar to our work. This algorithm is more complex compared to the DPS algorithm as it divides packet scheduling into the resource allocation and resource assignment techniques, as previously discussed in Section I. The radio resources are allocated and assigned to each user based on the user’s average instantaneous downlink channel conditions, packet delay information and packet loss rate.

The PLR is defined as the ratio of total size of discarded packets to the total size of all packets arriving into the serving eNodeB buffer and is given as:

\[ \text{PLR} = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T} p_{\text{discard}}(t)}{\sum_{i=1}^{K} \sum_{t=1}^{T} p_{\text{size}}(t)}. \]  (8)

where \( p_{\text{discard}}(t) \) and \( p_{\text{size}}(t) \) are the size of discarded packets and the size of all packets that have arrived into eNodeB buffer of user \( i \) at time \( t \) respectively, \( K \) is the total number of users and \( T \) is the total simulation time.
Moreover, the simulation results of the M-LWDF [4] algorithm is also included in this paper. Since M-LWDF algorithm is developed for the single carrier wireless systems, some modifications are made such that it can support packet scheduling in the downlink 3GPP LTE system. In this paper, the M-LWDF algorithm selects a user in each TTI based on (10) and allocates all the available radio resources to the selected user.

\[ k = \arg \max \alpha_i \cdot \frac{W_i(t) \cdot \frac{1}{N} \sum_{j=1}^{N} \text{SNR}_{i,j}(t)}{R_i(t)} \quad \forall \text{user}. \]  

\[ R_i(t) = (1 - \frac{1}{t_c}) \cdot \frac{r_i(t-1)}{t_c} + \frac{1}{t_c} \cdot r_i(t-1). \quad \forall \text{user}. \]  

where \( W_i(t) \) and \( R_i(t) \) are the average throughput and the HOL packet delay of user \( i \) at time \( t \), \( r_i(t-1) \) is the total amount of transmitted bits of user \( i \) at time \( t-1 \). \( \text{SNR}_{i,j}(t) \) is the instantaneous downlink SNR value of user \( i \) on RB \( j \) at time \( t \) and \( N \) is the available number of RBs. Variable \( \alpha_i \) is set to 1 as all users have similar applications which is video streaming.

The system throughput performance of the three algorithms with increasing number of users is shown in Fig. 4. Besides these algorithms, another line is plotted representing the maximum theoretical throughput that the system can achieve assuming all users have a very good instantaneous downlink SNR values and a fixed 128 kbps source video data rate is assigned to all users. From the figure it can be seen that the system throughput performance of the DPS algorithm is closer to the theoretical maximum and it has a higher system throughput when compared with the OS algorithm as the number of users increases above 50. A higher system throughput performance is achieved in the DPS algorithm since the algorithm always transmits the selected users’ packets on their best RBs. On the other hand, the system throughput performance of M-LWDF algorithm is significantly lower compared to the two algorithms as the radio resources being allocated to a selected user in each TTI is not being efficiently utilized due to the limited data in the selected user’s buffer.

Fig. 5 shows the average system delay of the DPS, OS and M-LWDF algorithms with increasing number of users. From the figure, it can be seen that the M-LWDF algorithm has the worst average system delay compared to the DPS and OS algorithms as the algorithm gives scheduling opportunity to only one user in each TTI and hence resulting in accumulation of HOL packet delays of the other users. It is also can be observed in the figure that the DPS algorithm is able to maintain a low average system delay when the number of users is below 80. However, when the number of users is above 80, its performance degrades compared to the OS algorithm.

As discussed in Section III, the DPS algorithm always selects the users with the lowest \( d_i(t) \), or in other words, in each scheduling opportunity, it selects the user with the highest HOL packet delay (as long as the HOL packet delay does not exceed the delay threshold). With increasing number of users, there will be more video streaming packets waiting for downlink transmission at the serving eNodeB buffer. Since there are insufficient RBs to transmit all the packets, the DPS algorithm only transmits the packets of the users having the highest HOL packet delays. The HOL packet delays of the other users will continue to increase in the consecutive TTIs. These packets are either selected for transmission by the DPS algorithm once their HOL packet delays are the highest or are discarded if the HOL packet delays have reached the delay threshold. These have resulted in a significant increase of the average system delay in the DPS algorithm when the number of users is above 80.

Since the PLR performance of the M-LWDF algorithm is very poor (already above 0.2 when the number of users is 20), only the PLR performance of the DPS and OS algorithms are shown in Fig. 6. As expected, with increasing average system delay (as shown in Fig. 5), there will be more packets being discarded since there are insufficient RBs to transmit all the packets whose HOL packet delays are approaching the delay threshold. However, when compared with the OS algorithm, the DPS algorithm achieves better PLR performance by supporting a higher number of users at the recommended PLR threshold (as discussed in Section IV-B). In the figure, it can be seen that, at 1% PLR threshold, approximately 30 and 80 users can be supported by the OS and the DPS algorithms respectively.

![Figure 4. System throughput vs. number of users.](image1.png)

![Figure 5. Average system delay vs. number of users.](image2.png)
Finally, the fairness performance of the three algorithms is shown in Fig. 7. It can be seen from the figure that, the DPS algorithm has a better fairness performance compared to the OS and M-LWDF algorithms. However, as the number of users increases above 80, the fairness performance of the DPS algorithm degrades since there is a larger difference between the total size of the transmitted packets between the most and the least served users (as defined in (9)). One possibility is that, when the number of users increases above 80, most of the packets of the least served user are being discarded while most of the packets of the best served users are being transmitted on the best RBs reported by the user.

VI. CONCLUSIONS

In this paper, a new packet scheduling algorithm known as delay-prioritized scheduling is proposed to support the real time traffic in the downlink 3GPP LTE system. The DPS algorithm aims to maximize system throughput while satisfying the QoS requirements of the RT users by utilizing the instantaneous downlink SNR values and packet delay information of each user. It is shown in the simulation results that, when compared with the opportunistic scheduling and maximum-largest weighted delay first algorithms, the DPS algorithm is able to improve system performance by (a) improving system throughput, (b) maintaining lower average system delay and packet loss ratio at a higher number of users and (c) distributes more fairly the available radio resources to all RT users.

REFERENCES