A Cross-Layer Delay Differentiation Packet Scheduling Scheme for Multimedia Content Delivery in 3G Satellite Multimedia Systems

Linghang Fan, Hongfei Du, Member, IEEE, Upendra Mudugamuwa, Student Member, IEEE, and Barry G. Evans, Senior Member, IEEE

Abstract—The design of efficient packet scheduling algorithms, which play a key role in the radio resource management (RRM), is crucial for the multimedia delivery in the satellite digital multimedia broadcasting (SDMB) system. In this paper, a novel packet scheduling scheme, which uses the cross-layer approach in its design, is proposed. This scheme comprises a new service prioritization algorithm and a dynamic rate matching based resource allocation algorithm, aimed at utilizing both the applications’ QoS attributes and the physical layer data rate information. The performance of the proposed scheme has been evaluated via simulation. In comparison with existing schemes, the proposed scheme achieves significant performance gain on delay, delay variation and physical channel utilization.

Index Terms—Cross-layer design, DDQ, dynamic rate matching, packet scheduling, RRM, SDMB.

I. INTRODUCTION

The EUROPEAN satellite digital multimedia broadcasting (SDMB) [1] system implements a satellite based broadcast layer over 2.5G and 3G terrestrial mobile cellular networks aimed at the efficient delivery of the multimedia broadcast multicast service (MBMS) [2], [3]. Due to the unidirectional nature of the SDMB system and the point-to-multipoint services it provides, the design of packet scheduling algorithms in SDMB is challenging. The packet scheduling schemes in SDMB are not only required to satisfy the quality of service (QoS) requirements of different service flows, but also have to optimize the transmission power setting of each physical channel on the basis of the required reception QoS level and under the constraint that the total available transmission power for all the physical channels within a satellite beam is fixed. Given the absence of a power control mechanism and the lack of channel-state-information (CSI) via the return channels, existing packet scheduling algorithms [4]–[8] used in general mobile networks cannot be used directly in the SDMB system.

Previous studies [9] have addressed the packet scheduling problems in the SDMB system via adaptation of two well-known packet scheduling schemes, namely multi-level priority queuing (MLPQ) and weighted fair queuing (WFQ). However, both feature major weaknesses in provisioning QoS differentiated multimedia services with respect to both efficiency and fairness. Furthermore, the power allocation algorithm used in previous packet scheduling schemes is based on static rate matching (SRM), where the transmit power setting for the physical channel is based on the required reception quality of the most demanding $E_b/No$ requirement against target block error rate (BLER) of all the active multiplexed service flows in the physical channel. No cross-layer information is exchanged between the MAC layer scheduling algorithm and the physical layer functions. Relevant issues in relation to the reliability mechanisms for the SDMB system are evaluated in [10], where the cross-layer approach is investigated for combining redundancy introduction across protocol stacks.

To this end, based on our previous research [11], a novel cross-layer packet scheduling scheme, called cross-layer delay differentiation queuing (CL-DDQ), is proposed in this paper. Two main algorithms can be identified in the proposed packet scheduling scheme: an efficient service prioritization algorithm, namely delay differentiation queuing (DDQ), and dynamic resource allocation (DRA) algorithm based on a new technique called dynamic rate matching [12]. As the first step, DDQ performs service prioritization based on the applications’ QoS demands and delay status of competing flows. Thereafter, the DRA scheme, which operates in coordination with the dynamic rate matching, further optimizes the resource allocation based on the physical layer data rate. Dynamic rate matching is proposed against the conventional downlink SRM in order to minimize the number of discontinuous transmission (DTX) bits required for the chosen transport format combination in a transmission time interval (TTI) scale given the available physical layer resources. Unlike its SRM counterpart, by utilizing the cross-layer information (i.e. physical layer instantaneous data rate), dynamic rate matching achieves better resource utilization at the MAC layer, and thereby improves the overall system performance.

This paper is organized as follows. An outline of the radio resource management (RRM) concept in the SDMB system is presented in Section II. The proposed RRM strategy is then presented in Section III. In Section IV, the simulation scenarios are described. We then proceed in Section V with the performance evaluation of the proposed scheme, in comparison with existing schemes. We summarize our proposal and conclude this paper in Section VI.

II. RRM IN SDMB

A. System Review

As illustrated in Fig. 1, the SDMB system defines a hybrid satellite-terrestrial communication system, featuring a unidi-
rectional geostationary satellite component that is responsible for the delivery of the point-to-multipoint MBMS services and provides a European coverage by multiple umbrella cells. The SDMB radio interface employs an adaptation of the wide-band code-division multiple access (WCDMA) \cite{13}, with the satellite gateway (Sat-GW) hosting both the Radio Network Controller (RNC) and the Node B functional entities of the UMTS Radio Access Network. The user equipment (UE) applies the standard 3G terminal enriched with SDMB-enabling functions, which, given the unidirectional nature, are very limited. The terrestrial gap-fillers, identified as intermediate module repeater (IMR), are co-installed physically at the terrestrial base stations to enhance the signal reception quality and provide adequate coverage in urban, built-up areas. The SDMB-enabled Broadcast/Multicast Service Center (BMSC), is enhanced with SDMB-specific functions from the standard 3GPP MBMS BMSC \cite{3}.

It is noteworthy that no direct satellite return link is envisaged under the baseline SDMB infrastructure, the return path is rather provided via the terrestrial link if needed. It is assumed that in SDMB, MBMS services are intended for transmission to UEs in either broadcast or multicast mode. In the latter case service is only delivered to the UEs within a specific multicast group. Packets from the BMSC are firstly buffered at the Sat-Hub (or Node B) in a FIFO manner before being scheduled for transmission over the satellite link. Being closely integrated into the baseline architecture of 2.5 G/3 G mobile cellular networks, the system enjoys maximum reusing of technology and infrastructure and minimum system development cost \cite{14}. The hybrid system takes advantage of the satellite inherent broadcast capability to provide efficient delivery of MBMS contents to the extensive mass mobile market.

Being defined by the European Telecommunications Standards Institute (ETSI), the SDMB system provides datacast capacity for various mobile operators. Based on a broadcast nature, the SDMB system offers extensive coverage, low transmission cost for large numbers of terminals as well as high QoS guarantees for real time multimedia applications. By employing the WCDMA with frequency division duplexing (FDD), the system can be closely integrated with existing mobile cellular networks, and minimize potential cost impacts on both 3 G cellular terminals and network operators. The successful validation and demonstration of the innovative SDMB concept carried out within the EU IST project MoDiS \cite{15} has pushed the system towards an operational stage. The whole range of issues pertinent to the SDMB system, from system definition to standardization, is addressed in the EU IST project MAESTRO \cite{16}.

In the S-DMB system, the service types are considered as: “streaming”, “hot download”, and “cold download”:

- **Streaming** allows multimedia to be stored temporarily in the receiver buffer and displayed continuously even before the completion of transmission. Streaming service requires explicit upper bound on queuing delay/ delay variation.
- **Hot download** allows the data to be stored at the receiver for their offline access. Compared with streaming, the hot download service has more tolerant demand on delay and delay variation but more stringent demand on packet loss.
- **Cold download** requires the least demand on delay/ delay variation but the most stringent demand on packet loss, services in this category are often transmitted as individual file, such as software package, video/images, and text messages.

### B. RRM in SDMB

The Radio Resource Management (RRM) functionalities implemented at the SDMB access layer comprise three main separated but cooperated parts: packet scheduling, radio resource allocation (RRA), and admission control. Due to the unidirectional nature of the SDMB system and the point-to-multipoint services it provides, the design of an efficient RRM scheme is challenging.

The physical channels are multiplexed in the satellite gateway through a two-level RRA procedure \cite{17}. The RRA is responsible for the radio bearer configuration at the time of the admission for each session, which includes the estimation of the required number of logical/transport/physical
channels and their mapping from logical channels to the transport/physical channels [18]. As seen from Fig. 2, each service is mapped, one-to-one, onto an MBMS point-to-multipoint Traffic CHannel (MTCH) logical channel [19], which is then mapped onto the Forward Access Channel (FACH) transport channel. At the physical layer, the Secondary Common Control Physical CHannel (S-CCPCHs) can carry one or more FACH(s) via transport channel multiplexing [18]–[22]. The implementation on the radio bearer mapping and allocation are out of scope of this paper, interested readers may refer to [23] for more details.

The role of the scheduler is to time-multiplex competing QoS-differentiated service flows and adjust the transmit power setting for each physical channel accordingly.

The admittance decision of each incoming requested session is handled by the admission control function. An appropriate transport format combination set (TFCS) for each physical channel must be derived for the packet scheduler performing the short-term resource allocation work.

Fig. 3 illustrates an example of data exchange through the SDMB access layer. A transport block is the minimum element that can be accepted by the physical layer to be jointly encoded, and a set of transport blocks that are exchanged between MAC and physical layer at the same TTI using the same transport channel is defined as a transport block set (TBS). For each TTI, the packet scheduler in the MAC will choose an appropriate transport format from each multiplexed transport channel. In the presence of the transport channel multiplexing, the combination of the selected transport formats/transport format sets for all the transport channels forms a transport format combination. All the transport format combinations that an UE is permitted to transmit in each physical channel during each TTI are included in a list called the TFCS. One of the main functions of the packet scheduler is to select the appropriate transport format combination from the TFCS list in each TTI.

Given the absence of a power control mechanism in SDMB, the packet scheduler becomes the main mechanism of fast resource allocation and performs the priority handling and transport format selection tasks. More specifically, in SDMB, the packet scheduler is responsible for two important tasks that are executed in each TTI interval of the radio bearers:

1) Time-multiplexing of service flows with different QoS requirements into physical channels with fixed spreading factor, so as to satisfy these requirements.

2) Adjusting the transmit power of the physical channel carrying the data flows on the basis of the required reception quality of the service (i.e. the target BLER) under the constraint that the total available power for all the physical channels within a satellite beam is fixed.

As shown in Fig. 4, time multiplexing from transport channels to physical channels is preformed independently from other physical layer functions, such as CRC, turbo coding and interleaving. However, the performance of rate matching, which is responsible for matching input data rates to the allocated data rate by performing puncturing or repeating the bits, is highly relevant to the transport channel multiplexing.

It is noted that extensive research [24], [25] has been done on the upper layer Turbo coder and first interleaving, but limited research has been carried out with respect to the rate matching technique. In traditional SRM schemes, the allocated data rate is based on the maximum data rate supportable for each physical channel. This strategy can only influence long term resource allocation, whilst the short term physical layer data rate variation can waste system capacity. Since the rate matching functionality is performed at the physical layer in accordance with other physical layer procedures, cross-layer interactions between physical layer and MAC layer can lead to beneficial impact on resource utilization/allocation.

C. Problem Analysis

According to the specific features in the SDMB system, the packet scheduling scheme has the following requirements to consider:

1) Time multiplexing;
2) Power constrains;
3) Absence of return channel(s).

As aforementioned, previous studies [9] have systematically addressed the packet scheduling problems in the SDMB system via adaptations of two well-known packet scheduling schemes, namely multi-level priority queuing (MLPQ) and weighted fair queuing (WFQ). These two scheduling schemes were designed based on the terrestrial scheme according to specific features in the SDMB system. However, both feature major weaknesses in provisioning QoS-differentiated multimedia services with respect to efficiency and fairness.

Firstly, MLPQ always processes packets from those non-empty queues with the highest QoS rank; as a result, packets waiting in lower-priority queues may suffer from considerably longer queuing delays. Furthermore, it is generally agreed that background applications do not have strict delay constraint, and the only requirement for applications in this category is that information should be delivered to the user error free. In fact, background applications still need a delay constraint as there will always be an upper limit for any service category. Finally, MLPQ deals with queues having the same priority in a round-robin fashion. As a consequence, no differentiation is made between queues within the same QoS rank. However, this is not an efficient tactic, since other essential QoS metrics (e.g. queuing delay) should also be taken into account when scheduling.

In the WFQ case, the weights are set according to each service flow’s data rate instead of priority. The serving order of all the
queues is based on the data rate and the time-stamp of packets at the head of each queue. This means that the higher the data rate of a service, the higher possibility it will be served. If two services have the same data rate, the service with the lower time-stamp on its head packet will be served first. Since non-priority policy is applied in this scheme, this may lead to unacceptably long delay in higher priority queues.

The SRM scheme has been used within the existing RRM scheme [9]. The objective of the SRM is to minimize the number of DTX bits required for the maximum data rate supported by the TFCS for a given physical channel. The rate matching ratio is calculated according to the maximum data rate at the beginning of each session. However, when the instantaneous data rate is lower than the maximum data rate, the rate matching ratio is kept constant, which will result in unnecessary puncturing and inefficient resource/power utilization in the physical layer.

III. PROPOSED PACKET SCHEDULING SCHEME

To overcome the problems discussed in the above section, an efficient packet scheduling scheme not only needs to consider power constraints and the existing time multiplexing scheme, but also needs to consider the following main requirements:

1) Both fairness and priority;

2) Optimizing the power usage.

In this paper, a novel cross-layer packet scheduling scheme is proposed to meet these requirements. This scheme consists of two new algorithms: Delay Differentiation Queuing (DDQ) and Dynamic Resource Allocation (DRA).

The packet scheduling scheme in the SDMB system is shown in Fig. 5, which can be conceptualized into the following two parts: service prioritization and resource allocation.

DDQ, considering both fairness and priority, performs the service prioritization task. DRA is developed based on the novel dynamic rate matching technique and performs the resource allocation task at the MAC sublayer according to the cross-layer data rate information from the physical layer. This new technique is more resource-friendly in that the radio resource is allocated dynamically based on the instantaneous data rate for each application class in a short-term TTI-scale.

A. Service Prioritization Algorithm

The incoming service requests are first ordered according to a priority criterion. In order to select the respective criteria, the service attributes are considered, which are normally mapped onto the traffic handling priorities, as defined by the SDMB QoS classes (i.e. streaming, hot download and cold download.
classes). It is noted that the prioritization can be more or less dynamic; in a more dynamic prioritization, the relative priorities of channels may change in each resource allocation interval (i.e., one TTI), depending for example on the QoS rank or average queuing delay status of competing flows.

The proposed DDQ service prioritization algorithm is developed based on the concept from the Hybrid Proportional Delay (HPD) scheduling scheme [4], which is widely used in service differentiated networks. In the HPD, the hybrid delay consists of two separate parts: average queuing delay and head waiting time. The head waiting time is the waiting time of the packet at the head of each class. This algorithm was modified as follows: 1) The waiting time used in our algorithm is the average waiting time of all the packets in the queue of each class instead of the waiting time of the packet at the head of each class; 2) instead of separating the average queuing delay and head waiting time, both queuing delay and waiting delay have been considered together in our algorithm and have also been assigned to the same weight in order to obtain the overall delay performance. DDQ performs service prioritization dynamically depending on the QoS and the waiting time/queuing delay experienced by packets in each FACH. This assumes that each MBMS session maintains a separate FACH queue and that there are QoS ratios between different QoS priority classes. In each TTI, the serving indices will be calculated for each queue. These serving indices are obtained based on two metrics: the QoS factor and the fairness factor.

The QoS factor $\alpha_i$ indicates the QoS priority of the MBMS services. The fairness factor $\delta_i$ indicates the fairness among the MBMS services, and is expressed by the average waiting delay for all the packets currently in the queue and the average queuing delay for all the packets that have left the queue before the current TTI. Important notations used in this article are summarized in Table I.

The mathematical formulation of DDQ can be expressed as follows.

Let $\delta_i(n)$ be the fairness factor at the current time slot $n$ for each queue $i$. This measure describes the delay states of all packets passing through the respective queue, including both the packets which are currently in the queue and those packets which have already left the queue (are served). It is calculated for each queue $i$ in each TTI as (1):

$$
\delta_i(n) = \frac{\sum_{j=0}^{N_q} W_{i,j}^q(n) + \sum_{j=0}^{N_q} W_{i,j}^d(n)}{(N_q + N_d)} \quad (1)
$$

where $\delta_i(n)$ is the fairness factor for queue $i$, $N_q$ is the number of packets that are currently in the queue, $W_{i,j}^q(n)$ is the waiting delay for packet $j$ currently in the queue $i$, $N_d$ is the number of packets that have left the queue before this TTI, $W_{i,j}^d(n)$ is queuing delay for packet $j$, which has left the queue $i$ before this TTI.

Let $\alpha_i$ be the QoS priority factor for the service flow at FACH queue $i$; the priority for queue $i$ in TTI $n$ can be defined as (2):

$$
P_i(n) = \alpha_i \times \delta_i(n) \quad (2)
$$
Consequently, the serving orders are calculated and assigned to each FACH by (2) at the beginning of each TTI. FACH queues with higher priority will be served ahead of the lower priority counterparts in non-preemptive order.

With the above approach of dynamic service prioritization in mind, the instant priorities of the FACH queue indicate the serving order for each TTI by the scheduler.

B. Dynamic Resource Allocation Algorithm

Once all the services to be transmitted are prioritized, the next step is the allocation of the resources to these services, which consists of bit rates and transmission power assignments within the specific resource allocation interval (i.e. one TTI).

Previous research on the rate matching and power allocation separates their optimizations [9]. By using the cross-layer design, a new power allocation algorithm is proposed to work in coordination with the rate matching technique to further optimize the scarce radio resources in the SDMB system.

As shown in Fig. 4, this new power allocation algorithm uses dynamic rate matching to select the required transmission power for all physical channels according to their instantaneous data rate requirements. The proposed dynamic rate matching relies on instantaneous data rate instead of maximum data rate used in the SRM. The rate matching ratio is calculated for every TTI and corresponds to the instantaneous data rate. Therefore, it has two main advantages: firstly, it allows better DTX minimization, and secondly its power requirement will be less for lower data rates (i.e. it requires less power when the instantaneous data rate is lower than the maximum data rate). The disadvantage of dynamic rate matching is that it needs more processing and memory compared to SRM.

The procedure for the proposed power allocation algorithm based on dynamic rate matching is described in the following.

For each FACH, the packet scheduler scans the TFCs of the physical channel to find all the different TBS sizes that could be used, namely to serve the whole or part of the queued data at the FACH buffer. A sorted list of all candidate TBS sizes, in decreasing order, is created. The scheduler first selects to allocate the maximum TBS size to the first FACH.

As shown in Fig. 5, the scheduler first checks the rate matching ratio based on (3). It calculates the RM ratios for all the subsets of transport format combinations from the full transport format combination set, and these values are stored against each transport format combination set. Then the scheduler checks which subset the chosen transport format combination belongs to, and based on this information the rate matching ratio obtained from the stored data, and a tentative value is determined according to the selected transport format combination. Following this, the scheduler performs tentative value correction and then rate matching patterns are generated.

\[
R_{p,q}^{T} = \frac{N_{\text{Data}}}{\sum_{k=1}^{q} \left( (I_k + I_{C_{RC}}) \times CR + I_{\text{ainibds}} \times N_k \right)} \quad \text{for } 1 \leq q \leq \omega \quad (3)
\]

where, \( p \) is the S-CCPCH ID, \( q \) is the transport format combination ID, and \( \omega \) is the transport format combination subset that its TBS that has been calculated. \( N_{\text{Data}} \) is the allowed data. \( I_k \) is the transport block length of FACH \( k \) in this transport format combination. \( N_k \) is the number of TB with allocated TBS. \( CR \) is the coding rate.

According to the calculated RM ratio values, the scheduler determines the required \( E_{b}/N_{0} \) value according to each session BLER requirement. The scheduler then checks whether the selected TBS size for the new session satisfies the total transmit power criteria.

- If this is not satisfied, it will check whether all the possible TBS sizes have been checked for total transmit power criteria. For the next TBS size (less than the previous one), the above procedure is performed. If none of the TBS sizes satisfy the power criteria, the scheduler assigns TBS to zero.

- If the power criteria is met, based on each session and RM combinations, the transmit power for each session is calculated separately according to (4) and the highest power requirement assigned as the physical channel transmit power.

\[
P_{n}(n) = \frac{P_{N}}{R_{s} \times CR \times RF} \quad (4)
\]

where \( P_{n}(n) \) is the required transmit power at time slot \( n \), \( P_{N} \) is the thermal noise, \( \rho \) is the \( E_{b}/N_{0} \) requirement, \( L \) is path loss, and \( RF \) is the rate matching ratio, \( R_{s} \) is the modulation scheme, and \( SF \) is the spread factor.

These procedures are repeated recursively until all the FACHs mapped to each S-CCPCH are assigned.

IV. SIMULATION SCENARIOS

Simulation has been carried out for the proposed RRM scheme using the software package ns2 [26], where different traffic mixes and physical channel capacities are evaluated. Three types of SDMB QoS categories, namely streaming, hot download, and cold download, are considered in the simulation. These services correspond to UMTS QoS classes streaming and background respectively [13]. In addition, we choose different guaranteed data rates in order to examine the performance between users with different rates. The performance of our proposed strategy and that of previous studies are compared via simulation metrics, such as queuing delay, queuing delay variation, and system utilization.

The performance of our proposed CL-DDQ scheme is examined over a wide variety of traffic mix scenarios. We consider individual MBMS sessions with diverse QoS profiles in terms of service type, data rate, and QoS constraints for broadcast transmission, each of which is carried by a single FACH queue. Multiple S-CCPCHs are used for carrying heterogeneous multimedia services and the radio bearer mapping scenarios given in Table II are considered as representative for performance evaluations.

Scenario I is formed by 5 FACHs and 2 S-CCPCHs carrying homogenous multimedia service, i.e., streaming service, where no QoS-differentiation is envisaged in this scenario, herein the parameter of interest is the different data rates of these streaming services.

1) S-CCPCH 1: Has three FACHs. Two streaming FACHs have data rates at 64 kbps, and one has a data rate at 128 kbps.
TABLE II

<table>
<thead>
<tr>
<th>S-CCPCH</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit rate (kbps)</td>
<td>384</td>
<td>384</td>
<td>384</td>
</tr>
<tr>
<td>Streaming (kbps)</td>
<td>256x1;128x1</td>
<td>256x1;128x1</td>
<td>-</td>
</tr>
<tr>
<td>Hot download (kbps)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cold download (kbps)</td>
<td>-</td>
<td>-</td>
<td>256x1;128x1</td>
</tr>
</tbody>
</table>

SCENARIO II:

<table>
<thead>
<tr>
<th>S-CCPCH</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit rate (kbps)</td>
<td>384</td>
<td>384</td>
<td>384</td>
</tr>
<tr>
<td>Streaming (kbps)</td>
<td>256x1;128x1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hot download (kbps)</td>
<td>-</td>
<td>256x1;128x1</td>
<td>-</td>
</tr>
<tr>
<td>Cold download (kbps)</td>
<td>-</td>
<td>256x1;128x1</td>
<td>256x1;128x1</td>
</tr>
</tbody>
</table>

SCENARIO III:

<table>
<thead>
<tr>
<th>S-CCPCH</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit rate (kbps)</td>
<td>384</td>
<td>384</td>
<td>384</td>
</tr>
<tr>
<td>Streaming (kbps)</td>
<td>256x1;64x1</td>
<td>256x1;128x1</td>
<td>-</td>
</tr>
<tr>
<td>Hot download (kbps)</td>
<td>64x1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cold download (kbps)</td>
<td>-</td>
<td>-</td>
<td>384x1</td>
</tr>
</tbody>
</table>

2) S-CCPCH 2: Has two streaming FACHs, whose data rates are 256 kbps and 128 kbps respectively;
   Scenario II considers two additional lower priority traffic classes: hot download and cold download. However, the scenario is defined as only a single service type is carried by each S-CCPCH.
   1) S-CCPCH 1: Has the two streaming FACHs have data rates at 256 kbps and 128 kbps respectively;
   2) S-CCPCH 2: Has two hot download FACHs, whose data rates are 256 kbps and 128 kbps respectively;
   3) S-CCPCH 3: Has two cold download FACHs, whose data rates are 256 kbps and 128 kbps respectively;
   Scenario III describes a more complex situation; where heterogeneous traffic types are carried by arbitrary S-CCPCHs. The task of the packet scheduler therein not only includes the differentiation of the session within a single S-CCPCH, but also embraces the traffic differentiation between FACHs which are carried by different S-CCPCHs.
   1) S-CCPCH 1: Has three FACHs. The hot download FACH (i.e. FACH 1) has a data rate at 64 kbps, and the two streaming FACHs (i.e. FACH 2, 3) have data rates at 256 kbps and 64 kbps respectively;
   2) S-CCPCH 2: Has two streaming FACHs (i.e. FACH 4, 5), whose data rates are 256 kbps and 128 kbps respectively;
   3) S-CCPCH 3: Has one download FACH (i.e. FACH 6), and its data rate is 384 kbps;
   Our link budget simulation results provide the $E_b/N_0$ v.s. BLER look-up curves of each FACH, where the radio propagation channel model features either classical Ricean characteristics for satellite-associated path, or Rayleigh multipath fading channel for UE-associated path with the consideration of both Doppler effect and propagation impairments. The maximum SDU size is 1500 bytes, TTI is 0.08 seconds, Turbo coder and QPSK is applied. The simulation period is set to 1000 s or 12500 TTIs. We use publicly available trace files [27] for video streaming traffic flow. Hot download and cold download traffic characteristics follow the typical ns2 Pareto distribution, with different QoS factors assigned [28].

V. PERFORMANCE EVALUATION

Since the selection of rate matching techniques features major changes between the proposed scheduling algorithm and that of the previous research, we identify the scheduling algorithm scenarios as combinations of scheduling algorithms and rate matching techniques:
1) Weighted Fair Queue scheduling algorithm with Static Rate Matching technique: WFQ-SRM;
2) Multi-level Priority Queue scheduling algorithm with Static Rate Matching technique: MLPQ-SRM;
3) Delay Differentiation Queue scheduling algorithm with Static Rate Matching technique: DDQ-SRM;
4) Delay Differentiation Queue scheduling algorithm with Dynamic Rate Matching technique: DDQ-DRM. This is our proposed CL-DDQ scheduling scheme.

The following major parameters, which have significant impact upon the overall system performance, have been examined in our evaluation:
1) Queuing delay;
2) Queuing delay variation;
3) Physical channel utilization.

A. Queuing Delay Evaluation

The mean queuing delay experienced by packets of each FACH queue in the buffer is an important parameter to evaluate the performance of packet scheduling algorithms. Fig. 6 shows the queuing delay of different scheduling schemes under different traffic mix scenarios.

As illustrated in Fig. 6(a), only streaming sessions are envisaged in scenario I, we set the simulation parameters to heavy traffic load scenario, i.e., all queues are non-empty during a session transmission time. It is shown that, for homogenous traffic mix, the performance of WFQ mainly depends on the service data rate. As shown, the MLPQ achieves the best delay performance for streaming services and the worst delay performance for download services because of its strict-priority based scheduling mechanism. DDQ, with its delay-differentiation nature, provides improved performance for low class services by sacrificing the performance of high class sessions.
In scenario III, compared with MLPQ-SRM and DDQ-SRM, WFQ-SRM has much lower delay for the high data rate services (i.e., FACH 6 at 384 kbps), but has much longer delay for the relative low data rate services (i.e., FACHs 1, 3 and 5). However, it is noted that FACH 6 carries the download services and FACHs 3 and 5 carry the streaming services. This means the predominance of high data rate download services causes long delays on lower data rate streaming services since the priority between service classes is almost ignored by WFQ-SRM according to its data-rate-based nature.

Numerically, for instance in scenario III, hot and cold download classes have a reduction of 51.3% and 74.0% on their mean queuing delay, respectively. It is noted that, by using DDQ-DRM, the significant reduction on delay of the lower class service has been achieved. This indicates that, unlike the SRM scheme, dynamic rate matching scheme is more suitable to DDQ.

As stated above, both DDQ and dynamic rate matching achieve better queuing delay than previous approaches. When the performance of DDQ with dynamic rate matching is compared with those of MLPQ and WFQ with SRM, significant performance enhancement are seen.

### B. Queuing Delay Variation

The “queuing delay variation” is defined as the variation of the instantaneous queuing delay from their mean value during a specified transmission period. Fig. 7 shows the mean queuing variation experienced by each individual FACH for three different packet scheduling schemes under different traffic mixes.

As shown in Fig. 7, by using SRM, compared MLPQ and DDQ, the WFQ has much lower queuing variation for the high data rate services, but has much longer queuing variation for the relative low data rate services. It is noted that the unidirectional streaming service in the SDMB system is quite sensitive to delay-variation. The delay variation of the streaming service should be limited in order to preserve the time variation between packets of the stream [29]. Although WFQ assigns higher priority to high data rate services, there is no differentiation made with respect of QoS rank, therefore WFQ makes delay variation-tolerant high data rate download services have higher priority than the delay variation sensitive lower data rate streaming services.

As shown in Fig. 7, in the case of SRM presence, DDQ has better performance on queuing delay variation than both MLPQ and WFQ. If the dynamic rate matching is used, DDQ-DRM features much lower delay variation for both streaming and download services. Typically, the average delay variation reduction for download services in scenario III is as much as 73.0%, while the average delay variation reduction for the streaming service reaches 38.2%.

### C. Physical Channel Utilization Evaluation

The mean S-CCPCH channel utilization for different schemes is studied in this section. As shown in Fig. 8(a), in Scenario I, where the traffic mix is limited to single streaming traffic class, i.e., no QoS-differentiation is envisaged for the scheduling plan, the WFQ behaves as a rate-based differentiation scheme while the MLPQ can be regarded as a simple round robin discipline; therefore the discrepancy featured in the physical channel utilization is mainly caused by the incoming traffic dynamics. DDQ-SRM is able to provide a primary differentiation in terms of queuing delay behaviors and the impact of traffic dynamics on the channel utilization is largely eliminated, i.e., smaller differences on S-CCPCH channel utilization. The proposed DDQ-DRM scheme, achieves better overall resource utilization; the mean physical channel utilization for all S-CCPCHs...
reaches 96.1%, compared with 86.4%, 91.2% and 92.3% in WFQ-SRM, MLPQ-SRM and DDQ-SRM, respectively.

From the result obtained for Scenario II, where single QoS traffic classes are carried within a single S-CCPCH, it indicates that by adaptively utilizing the resource in an intelligent way, the DDQ-DRM achieves higher resource utilization than other schemes in that it either utilizes the wasted resources or re-utilizes/shares the resources in a more efficient way. The overall channel utilization has been improved by 56.7%, 43.1% and 23.5% from WFQ-SRM, MLPQ-SRM and DDQ-SRM respectively.

In scenario III, DDQ-DRM achieves much higher physical channel throughput than for all the other schemes. It reaches 378.2kbps on S-CCPCH 1 and 378.9kbps on S-CCPCH 2 and 328.1 kbps on S-CCPCH 3, which are equivalent to 98.4%, 98.6% and 85.4% of the S-CCPCH capacity, respectively. Its nearest candidate, DDQ-SRM, reaches 366.4kbps on S-CCPCH 1, and 363.4kbps on S-CCPCH 2 and 328.1 kbps on S-CCPCH 3, which are equivalent to 95.5%, 94.7% and 85.4% channel utilization, respectively.
Hence, DDQ-DRM manages to obtain channel utilization improvement on those physical channels carrying background traffic. From the results, it can be inferred that the proposed DDQ-DRM scheme not only improves the delay performance among different QoS classes, but also increases the throughput and physical channel utilization.

As a summary, the proposed cross-layer scheduling algorithm has the following advantages over the existing packet scheduling scheme:

- Achieves dynamic proportional delay-driven prioritization;
- Balances all FACHs irrespective of their QoS class;
- Has significant improvement on delay and delay variation performance;
- Achieves more efficient resource allocation;
- Achieves higher throughput and channel utilization.

VI. CONCLUSIONS

The SDMB system intends to implement a commercial satellite based broadcast-multicast (BC/MC) layer over the unicast terrestrial 3 G mobile telecommunication system infrastructure aimed at the efficient delivery of the interactive MBMS services to mobile terminals. The design of efficient packet scheduling algorithms plays a key role in radio resource management (RRM) for the multimedia content delivery in the SDMB system.

In this paper, a novel cross-layer packet scheduling scheme, namely CL-DDQ, is proposed for the SDMB system. The proposed scheme not only balances the priority and fairness among QoS-differentiated traffic flows, but also considers physical channel data rate factor.

Performance evaluation of the proposed approach has been carried out via simulation studies. The results show that, compared with the existing packet scheduling schemes, the proposed cross-layer packet scheduling scheme achieves better performance for delay, delay variation, throughput and channel utilization. Therefore, it better optimizes the overall system performance of the SDMB system.

In future work, more factors reflecting differentiated service QoS attributes will be taken into account in resource allocation algorithm in order to achieve even better QoS provisioning of multimedia content delivery in the SDMB system.

ACKNOWLEDGMENT

The authors also thank Dr. Atta Quddus for reviewing this paper.

REFERENCES


[26] [Online]. Available: http://www.ist.edu/nsnam

[27] [Online]. Available: http://www.tkn.tu-berlin.de/research/hltv.txt
