Combined Delay and Rate Differentiation Packet Scheduling for Multimedia Content Delivery in Satellite Broadcast/Multicast Systems

Hongfei Du, Linghang Fan and Barry G. Evans

Center for Communication Systems Research, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Email: H.Du@surrey.ac.uk

Abstract—The design of efficient packet scheduling algorithms is crucial to the radio resource management (RRM) in the satellite digital multimedia broadcasting (SDMB) system, which has emerged as a promising solution to the multimedia content delivery. In order to achieve more efficient quality of service (QoS) provisioning among different multimedia services, a novel packet scheduling algorithm, namely *combined delay and rate differentiation* (CDRD) packet scheduling, is proposed in this paper. This algorithm takes into account key QoS parameters aiming at fairly prioritising and scheduling heterogeneous multimedia contents in satellite environment. Its performance has been evaluated via simulations. The results show that, in comparison with existing scheduling algorithms, CDRD achieves better performance on delay, jitter and channel utilisation.

Index Terms — Packet scheduling, radio resource management, SDMB, MBMS, quality of service.

I. INTRODUCTION

THE PROVISIONING of point-to-multipoint services over the 3rd generation (3G) terrestrial mobile cellular networks is investigated within the third generation partnership project (3GPP) multimedia broadcast/multicast services (MBMS) framework [1]. MBMS data is mapped onto radio network bearers and is transmitted over air in parallel to unicast data. In Europe, much research effort [2] has been devoted to satellite digital multimedia broadcasting (SDMB) system, where satellite is used in the delivery of point-to-multipoint multimedia services to 3G handsets.

The SDMB system implements a satellite based broadcast/multicast layer over unicast terrestrial universal mobile telecommunication system (T-UMTS) infrastructure aiming at the efficient delivery of the interactive MBMS services to a wide range of audience. Due to the unidirectional nature of the system and the point-to-multipoint services it provides, given the absence of the power control mechanism and lack of channel state information (CSI), the design of packet scheduling scheme, which is a key element of radio resource management (RRM) functionalities implemented at the SDMB access layer, is challenging. In fact, an efficient packet scheduling algorithm in SDMB is not only required to satisfy the quality of service (QoS) requirements of different services, but also has to optimise 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 within a satellite beam is fixed.

A large amount of research [4-6] has been taken on developing efficient scheduling scheme for effective QoS

provisioning, in both wired and wireless systems. One interesting research subject foreseen in this context is delay differentiated scheduling, where waiting time and queuing delay are considered in packet scheduling, as waiting time priority (WTP) and proportional delay differentiation (PDD) schemes proposed in [4] for terrestrial networks. Besides, adaptive proportional fairness (APF) scheduling was proposed in high-speed downlink packet access (HSDPA) system [5], considering QoS demands for multimedia applications, where the return channel is in presence and CSI for individual user is available. Nevertheless, all those schemes can not be directly adopted in SDMB system in that there is no real-time interaction between the user and the satellite radio access network (SRAN) in the SDMB baseline architecture, the scheduler has to allocate resources without knowledge of the channel conditions, i.e. channel state dependent scheduling is not possible. Previous studies [6] have systematically addressed the RRM problems in the SDMB system via classical packet scheduling schemes, namely multi-level priority queuing (MLPQ) and weighted fair queuing (WFQ). However, both of them feature major weaknesses in QoS-differentiated multimedia services provisioning with respect to both efficiency and fairness.

MLPQ employs a strict QoS-based prioritisation scheme, in which a lower-priority service may suffer from considerably longer queuing delays. Furthermore, since round-robin is employed amongst queues with the same priority, no differentiation is made for services with the same traffic priority. However, an efficient scheduling algorithm must also consider other essential QoS metrics (e.g. queuing delay).

In the case of WFQ, the serving orders of the queues depend on the time-stamp of the head packet of each queue, queues with the lowest time-stamp will be served first. The weights are set according to the data rate of each queue rather than its priority. The non-priority nature of this scheduling policy leads to unacceptable long queuing delay in higher priority queues.

To overcome drawbacks inherited in MLPQ and WFQ, in this paper, we proposed a novel packet scheduling scheme, namely combined delay and rate differentiation (CDRD) packet scheduling, which is *distinct* from most existing scheduling algorithms in that: 1) it guarantees the prescribed QoS requirements and 2) accounts for queuing behaviours at link layer, while at the same time, 3) preserves the physical layer power/resource constraints. To the best of the authors' knowledge, no results are available to prove this concept under the current RRM framework in SDMB, and this work aims to fill this gap. The proposed methodology is envisaged in the

Figure 1. Proposed packet scheduling framework

SDMB system, but it can be applied adaptively to any WCDMA-based broadcast/multicast networks.

This paper is organised as follows. The proposed CDRD algorithm is presented in Section II. In Section III, the simulation methodology is described. We then proceed in section IV with the performance evaluation of CDRD algorithm, in comparison with existing packet scheduling algorithm. Finally, we draw our conclusion in Section V.

II. PROPOSED SCHEDULING ALGORITHM

The framework of the proposed CDRD scheduling scheme for delivering heterogeneous multimedia services in SDMB is illustrated as Fig. 1. The accepted ongoing traffic can comprise different sessions with different QoS requirements. Each session is assumed to retain an individual queue in the Forward Access CHannel (FACH) transport channel buffers. The types of user service supported within the SDMB system are streaming and download, which correspond to UMTS QoS class streaming and background respectively [7]. Herein the traffic class considered can be divided into two main categories depending on delay constraint. The first category is delay-sensitive streaming class. The second category is the delay-insensitive download class, which can be further sub-categorised into two sub-classes according to its sensitivity to delay, namely “hot download” and “cold download” [3]. Packets in FACH buffers are firstly prioritised by “Service Prioritization” module with their respective QoS metrics as criteria. Consequently, “Dynamic Adaptive Packet Scheduler” module will deal with all the queues in FACH buffers according to their instant priorities instead of their inherent QoS class. Nonetheless, since download service has no explicit delay constraint, the scheduler will restrict those download services aiming at preserving certain QoS level for streaming services if the available resource is rather scarce. On the contrary, as long as there is spare resource remaining in the streaming queues, the scheduler will enable download services to fill the gap left by the streaming services adaptively.

The selected queues will be passed to “Resource Allocation” module and be allocated with the required resources, which

consist of bit rate and transmission power assignments. The scheduled packets will be delivered to Secondary Common Control Physical Channel (S-CCPCH) in the form of transport block (TB) [3]. For each active physical channel, the exact format of Transport Format Combination(TFC), which consists of transport block set (TBS) will be selected (i.e. the amount of data from each transport channel mapped to the physical channel) from the Transport Format Combination Set (TFCS). It is noted that a separate TFCS is provided for each S-CCPCH.

As shown in Fig. 2, packet scheduling strategy can be generally conceptualised into the following two steps:

A. Service prioritisation

Upon receiving the incoming service requests, the scheduler sorts the queues according to some priority criterion. In selecting the respective criteria, the service attributes are considered in a joint judgment function (JJF) to provide dynamic intelligent scheduling task based on several essential QoS factors that have crucial impact on system performance.

CDRD scheme is developed based on proportional delay differentiation (PDD) scheme [4], where only delay is considered for service differentiation. By taking into account several key performance parameters (i.e. required data rate, queueing delay threshold), CDRD is able to balance all service flows with diverse QoS requirements, and thus, achieves best possible QoS satisfaction and efficient resource utilization.

For each FACH queue i at current time slot (i.e. transmission time interval (TTI)) n , the JJF is defined as:

$$P_i(n) = \alpha_i \cdot \delta_i(n) \cdot \lambda_i(n) \cdot \gamma_i(n) \quad (1)$$

where $P_i(n)$ is the priority index for queue i at current time slot n . n is the sequence number of the TTI at current time.

Figure 2. Packet scheduling procedure

α_i is QoS class factor, which is essentially a time-independent parameter designated for queue i . Since CDRD assumes that there are QoS ratios between different QoS priority classes, the value of α_i represents the relative traffic priority of the service carried by queue i , i.e. streaming service has higher priority than hot download service, which has higher priority than cold download service.

$\delta_i(n)$, defined in (2), specifies the delay factor at current time slot n for queue i . This measure describes the delay states of all packets passing through the respective queue, including both the packets which are currently in queue i and the packets which have already been served by queue i till current TTI n .

$$\delta_i(n) = \frac{\sum_{j=1}^{N_q} W_{i,j}^q(n) + \sum_{k=1}^{N_d} W_{i,k}^d(n)}{N_q + N_d} \quad (2)$$

where $W_{i,j}^q(n)$ is the waiting delay for the j^{th} packet currently in queue i ; N_q is the number of packets that are currently in queue i ; $W_{i,k}^d(n)$ is queuing delay for the k^{th} packet which has been served by queue i before TTI n ; N_d is the number of packets that have been served by the queue before TTI n .

$\lambda_i(n)$ represents the data rate factor for queue i at current time slot n . It is based on the ratio of the service required data rate against the average transmitted data rate $\overline{\lambda_i(n)}$ for queue i at current time slot n , which can be expressed as:

$$\overline{\lambda_i(n)} = \frac{\sum_{k=1}^{N_d} S_{i,k}}{(n-1) \times T_{tti}} \quad (3)$$

where $S_{i,k}$ is the packet size for k^{th} packet in queue i ; N_d is the number of packets that have left the queue before this TTI; T_{tti} is the value of TTI (e.g. it is 80ms in our simulation).

Therefore, the data rate factor $\lambda_i(n)$ is defined as follow:

$$\lambda_i(n) = \frac{\lambda_i^{req}}{\lambda_i(n)} \quad (4)$$

where λ_i^{req} is the required/guaranteed data rate specified by the service QoS level. If the average offered data rate is smaller than the required data rate, $\lambda_i(n)$ is larger than 1, thus the priority index for this queue is increased for this under-utilized queue; otherwise, the priority index will be decreased for this over-utilized queue. This factor fine-tunes the priority and lead the offered data rate be approaching to the guaranteed data rate.

$\gamma_i(n)$ is the delay constraint factor for queue i at TTI n , reflecting the current queuing delay status, which is defined as:

$$\gamma_i(n) = \begin{cases} 2 & , \quad \forall n : \frac{\sum_{j=1}^{N_q} W_{i,j}^q(n)}{N_q} \geq W_i^{threshold} \\ 1 & , \quad \forall n : \frac{\sum_{j=1}^{N_q} W_{i,j}^q(n)}{N_q} < W_i^{threshold} \end{cases} \quad (5)$$

where $W_{i,j}^q(n)$ is the waiting delay for the j^{th} packet currently in queue i ; N_q is the number of packets that currently in queue i ; $W_i^{threshold}$ is the delay threshold for the service queue i .

If the average queuing delay for queue i is larger than its delay threshold, the delay constraint factor $\gamma_i(n)$ doubles the priority of this queue for better chance to be processed; otherwise, it remains the same. The weight can be chosen as various ratios reflecting the effectiveness of the delay constraint factor in the overall joint judgement function. It is noted that delay threshold can be chosen as a tuneable parameter, which depends on the maximum tolerable delay of the corresponding service. $\gamma_i(n)$ is only in effect when the average queuing delay is beyond the designated delay threshold, which provides more efficient action to be taken for better QoS provisioning amongst differentiated traffic flows.

In each TTI, the scheduler will sort the FACH queues in descending order, according to their priority indices calculated from the JJF. FACH queues with higher priority indices will be served ahead of their lower priority counterparts.

B. Resource Allocation

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 rate and transmission power assignments within the specific resource allocation interval, which in our case is every TTI.

As shown in Fig. 2, for all S-CCPCHs, the packet scheduler serves the FACHs according to their instant priorities, which are dynamically calculated from the JJF in a particular TTI. The FACH queues with higher priorities will be served ahead of the lower priority FACH queues in non-preemptive order.

For each FACH, the packet scheduler scans the TFCS of its corresponding physical channel to find all the different TBS sizes and then seeks to allocate the maximum TBS size to the selected FACH based on data queued at its buffer [1]. Then the packet scheduler checks the power requirement on the basis of the BLER curve of the active service flow. If the power allocation satisfies the power and load constraints, the scheduler will allocate this FACH and derive a reduced TFCS; otherwise, this service is not scheduled.

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

From the viewpoint of implementation, the proposed CDRD algorithm poses extra computation complexity. With the input size of n (i.e. total number of FACH queues), the computational complexity of the proposed algorithm is derived as $O(n)$, following typical linear statistics.

III. SIMULATION METHODOLOGY

In order to elaborate the performance enhancement of the proposed CDRD packet scheduling algorithm, a system-level

simulator implementing the SDMB system has been developed with the software package *ns2* [8]. The scheduler is implemented in the Satellite Hub (Node-B) employing the SDMB functions, supporting three types of QoS classes, namely: 1) real-time video streaming, 2) hot download, and 3) cold download [3]. The streaming traffic model applies publicly available trace files [9] for video streaming traffics. Traffic characteristics associated with hot- and cold- download (i.e. push-and-store) services follow the *Pareto* distribution, with different traffic priorities. In addition, different guaranteed data rates are selected for individual MBMS session.

Our link budget simulation results provide the E_b/N_o v.s *BLER* look-up curves for each FACH. The radio propagation channel model features either 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 size of packets is 1280 bytes, TTI equals to 0.08 seconds, Turbo coder and QPSK are applied. The simulation period is set to 1000s or 12500 TTIs. In the SDMB system, the queuing delay threshold is assumed to be 20-100ms for video streaming, and 200-2000ms for push-and-store service. Accordingly, various queuing delay threshold values are applied and examined for the specific scenario, showing the range of the performance gain by tuning the delay threshold parameter.

Simulation has been conducted over a wide variety of traffic mixes. Due to the space limitation, an indicative and persuadable scenario is selected to discuss our simulation outcomes. We consider 6 individual MBMS session with diverse QoS profiles in terms of service type, data rate, and delay constraints for broadcast transmission, each of which is carried by a single FACH queue, the considered radio bearer mapping scenario is given as Table I. Since the performance of WFQ-based scheduling algorithm is worse than that of MLPQ-based algorithm in terms of both delay and delay variation [6], MLPQ, which is applied in the existing S-DMB access scheme, is used as the reference in this paper.

IV. PERFORMANCE EVALUATION

A: Queuing delay and delay variation evaluation

Firstly, we evaluate the performance of CDRQ scheduling by investigating the mean queuing delay experienced by packets in each FACH. As illustrated in Fig. 3, download multimedia services (i.e. FACHs 1 and 6) experience much less mean queuing delay in CDRD whilst the mean delay experienced by streaming services features similar performance. Numerically, hot download and cold download classes have a reduction of 32.6% and 23.7% on their mean queuing delay respectively, whilst the maximum increase on the mean queuing delay of

Figure 3. Mean queuing delay for MLPQ and CDRD

Figure 4. Mean queuing jitter for MLPQ and CDRD

streaming service is 7.6%. It is noted that the significant reduction on delay of the lower class service (i.e. download service) does not result in dramatic performance degradation on its higher class counterpart (i.e. FACHs 2 to 5 for streaming service) beyond their guaranteed QoS rank. It is implying that CDRQ enables the download service to efficiently utilize the spare resources of the streaming service without posing significant detrimental affect on satisfying the QoS target of the streaming services.

Fig. 4 shows the mean queuing jitter experienced by each FACH for MLPQ and CDRD. Obviously, the latter features much lower jitter for both streaming and download services. Typically, the average jitter reductions for download services (i.e. FACHs 1 and 6) and streaming services (i.e. FACHs 2 to 5) are 45.5% and 29.1% respectively.

It is worth noticing that the unidirectional streaming service in SDMB system is quite sensitive to delay variation (i.e. jitter), thereby the delay variation of the flows should be limited in order to preserve the time variation between packets of the stream [10]. The results in Fig. 4 shows that the proposed CDRD scheme provides a way to balance all FACH queues in order to get minimum delay variation for streaming services.

B: Queuing delay statistics analysis

Herein the probability density function (PDF) statistic of the queuing delay is used to describe statistical distribution density of queuing delay for each FACH. Fig. 5 shows the PDF curves of queuing delay for each FACH by using MLPQ and CDRD. It is noted that CDRD has better delay distribution characteristics

TABLE I. CHANNEL MULTIPLEXING CONFIGURATION

S-CCPCH	1	2	3
Bit rate(kbps)	384	384	384
Streaming (kbps)	256x1;64x1	256x1;128x1	-
Hot Download (kbps)	64x1	-	-
Cold Download (kbps)	-	-	384x1

Figure 5. PDF queuing delay for MLPQ and CDRD

than MLPQ. Especially, it achieves both lower average delay and delay variation for download service (i.e. FACHs 1 and 6).

Unlike those of download service, the PDF curves of the queuing delays of streaming FACHs (i.e. FACHs 2 to 5) are mixed at the bottom-left corner of Fig. 5. In order to analyse the statistic distribution performance on the queuing delay of streaming service and get better illustration, Fig. 7 plots the cumulative distribution function (CDF) distribution curves of queuing delay for considered streaming FACHs, which is a stronger expression of this effect. The results show that CDRD outperforms MLPQ in that all the streaming FACHs achieve better probability distribution on queuing delay when CDRD-based scheduling is employed.

In fact, the explanation of PDF and CDF queuing delay performance becomes straightforward if the conclusions of mean queuing delay/jitter derived from Figs. 3 and 4 are recalled. In conclusion, simulation results show that the delay and delay variation are greatly improved in the considered multiplexing scenario by adopting CDRD-based scheduling, compared with MLPQ-based counterpart.

C: S-CCPCH and FACH channel utilisation evaluation

Channel utilisation performance is presented in this section, considering utilisation ratios on both code channels(S-CCPCHs) and transport channels (FACHs). Fig. 6 displays average code channel utilisation status when adopting MLPQ and CDRD scheduling algorithms for the given scenario.

Viewable improvements can be noticed from the chart for those code channels which are comprised with transport channel carrying background class services (i.e. S-CCPCHs 1 and 3). Both schedulers managed to achieve throughput close to the maximum. For instance, the average code channel utilisation ratios are: 97.8%, 96.2% and 85.4% for respective S-CCPCHs under MLPQ-based scheduling; whilst they achieve 98.4%, 96.2% and 86.4% respectively when CDRD-based scheduling is adopted. Therefore, CDRD manages to obtain slight channel utilisation improvement on those code channels carrying background traffic (i.e. S-CCPCHs 1 and 3). From the results, it can be inferred that the proposed algorithm not only improves the delay performance among different QoS classes, but also has

Figure 6. Mean code channel utilisation for MLPQ and CDRD

Figure 7. CDF queuing delay for MLPQ and CDRD scheduling

essential impact on increasing the overall physical channel utilisation.

The explanation of the above statement can be seen from Fig. 8, where the comparison of the mean channel utilisation ratio of FACHs is displayed. Contrary to the MLPQ scheduling, in the CDRD scheduling scheme, it appears to achieve higher utilisation score, especially for background class traffic (i.e. download FACHs 1 and 6). These results also coincide with those of Fig. 6, explaining that the main improvement on code channel utilisation is virtually caused by the higher channel utilisation ratio achieved by CDRD-based scheduling on those FACHs carrying background traffic.

D: Delay threshold evaluation

As explained earlier, the queuing delay threshold can be tuned as an adjustable parameter that indicates different how to choose the queuing delay threshold for specific QoS class, we select typical delay threshold values in our simulation in order to illustrate that the delay threshold can be regarded as a tuneable parameter for delay tolerance/sensitivity adjustment.

Fig. 9 presents the performance of CDRD-based packet scheduling algorithm with variable delay threshold values

Figure 8. Mean FACH channel utilisation for MLPQ and CDRD

against specified scenarios. Tuning the delay threshold factor will change the priority index of related FACH queues and therefore change the serving order of all the FACH queues. As a result, the delay threshold factor can be adjusted according to the delay sensitivity/tolerance of differentiated QoS classes among heterogeneous multiplexing traffic flows and can be used to optimise the overall system performance.

As shown in Fig. 9, by tuning the delay threshold value for a specified QoS service class, the CDRD-based packet scheduling algorithm can adjust the delay performance of corresponding FACH queues; this also affects the performance of other QoS classes. For example, in comparison with Scenario 1, cold download FACH 6 suffers from worse delay in Scenario 4 when its delay threshold is increased from 0.8 second to 2.0 second, but this leads to the achievable performance gain on the streaming and hot download FACHs. On the contrary, compared with Scenario 1, more stringent delay threshold makes the streaming FACHs 2 to 5 have better delay performance in Scenario 2, whilst it causes longer delay on the download FACHs 1 and 6.

E: Fairness analysis

Herein the main parameter of interest is normalized throughput, which is obtained by dividing the offered throughput with the guaranteed throughput. The variance of normalized throughput, which compares the fairness scores achieved by MLPQ and CDRD, is plotted in Fig. 10. It can be seen that CDRD achieves the considerably lower variance values with a faster convergence curve, which means that it can provide UEs with better throughput equality in a relatively shorter time. As time elapses, comparing with MLPQ, CDRD is capable of maintaining a fairly low throughput variance, and thus, the long-term fairness amongst all queues are proven to be better guaranteed.

V. CONCLUSIONS

In this paper, a novel packet scheduling algorithm is proposed for the SDMB system. This algorithm takes into account multiple important performance factors reflecting service QoS demands and queuing behaviours in order to optimise the overall system performance. Simulation results show that the proposed packet scheduling scheme achieves better performance on queuing delay, channel utilisation and fairness compared with the existing schemes.

Figure 9. Mean queuing delay performance for different delay thresholds

Figure 10. Variance of normalized throughput for MLPQ and CDRD

ACKNOWLEDGMENT

This work is supported by the SatNEx project, funded by European Commission (EC) under the Research Framework Program 6. The financial contribution of the EC towards this project is greatly appreciated.

REFERENCES

- [1] 3GPP TS 22.146 V6.6.0, "Multimedia Broadcast/Multicast Service: Stage 1", October 2004.
- [2] N.Chubierre et al. "Satellite Digital Multimedia Broadcasting for 3G and beyond 3G systems", 13th IST Mobile & Wireless Communication Summit 2004, Lyon, France, June, 2004.
- [3] 3GPP TS 25.301 V6.2.0, "Radio Interface Protocol Architecture", March 2005.
- [4] C. Dovrolis et al. "Proportional Differentiated Services: Delay Differentiation and Packet Scheduling", IEEE Transaction on Networking, Vol.10, NO.1, February 2002, pp.12-26.
- [5] G. Aniba and S.Aissa, "Adaptive Proportional Fairness for Packet Scheduling in HSDPA", IEEE Globecom 2004, pp. 4033-4037.
- [6] M. Karaliopoulos et al. "Packet Scheduling for the Delivery of Multicast and Broadcast Services over S-UMTS", Int. J. Satell. Commun. Network. 2004; 22:503-532.
- [7] H. Holma, A. Toskala (Eds). WCDMA for UMTS: radio access for third generation mobile communications. Second edition, John Wiley & Sons Ltd, 2002.
- [8] <http://www.isi.edu/nsnam>
- [9] <http://www.tkn.tu-berlin.de/research/trace/ltvt.html>
- [10] 3GPP TS 23.107 V6.3.0, "Quality of Service (QoS) Concept and Architecture", June 2005.