PACKET SCHEDULING FOR THE DELIVERY OF BROADCAST/MULTICAST SERVICES VIA S-UMTS

M. Karaliopoulos(1), P. Henrio(2), K. Narenthiran(1), E. Angelou(3), B. Evans(1)

(1)Centre for Communication Systems Research, University of Surrey
Guildford, GU2 7XH, Surrey, UK, M.Karaliopoulos@surrey.ac.uk
(2)Alcatel Space, Nanterre, Paris, Pierre.Henrio@space.alcatel.fr

(3)Space Hellas S.A, Athens, Greece, eangelou@space.gr

ABSTRACT

The deployment of satellite systems as an overlay multicast/broadcast layer complementing the mobile terrestrial networks in delivering point-to-multipoint services is a concept attracting more and more interest the last years within the satellite community. The adoption of WCDMA technology for the radio interface of these systems allows their close integration with the terrestrial 3G mobile networks and more efficient handset implementations. We focus our attention on the radio resource management task and -in particular- on the packet scheduling function within an interface featuring maximum commonalties with the FDD mode of terrestrial UMTS. We describe the role of the packet scheduler, formulate its task into mathematical notation and investigate adaptations of two well-known scheduling disciplines, the multilevel priority queuing and weighted fair queuing schemes, as candidates for the time multiplexing task of the scheduler. Simulation results confirm the significance of the Transport Format Combination Set (TFCS) with respect to both the power efficiency of the scheduler and the performance obtained by the flows at packet-level. The performance gap of the two schemes regarding the fairness guaranteed to competing flows can be narrowed via appropriate selection of the TFCS, while absolute jitter scores are dependent on the packet-level dynamics of individual flows.

1. INTRODUCTION

The delivery of multimedia broadcast and multicast services to mobile users via satellite has been one of the main areas of the satellite communications industry activities in the last years. The inherent broadcast capabilities of satellites favours them with respect to the delivery of these services, in particular those featuring large and widely distributed audience. Moreover, the limited success of the stand-alone system model, providing a similar set of services with the terrestrial mobile networks, in capturing the consumer market has motivated the re-consideration of the satellite role in service provision. Within this context the close integration of the S-UMTS component with the terrestrial 3G architecture is regarded as a key factor for the success of the system. The level of this integration may vary. From the terminal point of view, use of the same waveform over the satellite radio interface enables maximum re-use of terminal hardware for both T-and S- UMTS modes with significant advantages in terms if its size, power consumption and –eventually- cost [1].

The potential of stand-alone UTRA FDD downlink carriers in order to deliver unidirectional point-to-multipoint (p-t-mp) services has been investigated within the framework of the IST SATIN project[1]. Such services are currently under investigation within the 3GPP Multimedia Multicast/Broadcast Services (MBMS) framework [2]. These stand-alone carriers convey several p-t-mp services in either broadcast or multicast mode. The MBMS data are mapped onto appropriate radio bearers and transmitted from the Node B in parallel to the UTRA FDD carriers. Intermediate module repeaters co-located physically with Node Bs boost the signal and ensure coverage in-door and in densely built-up areas [3].

This paper focuses on one of the radio resource management blocks of this system, the packet scheduling function. First, we briefly review the radio interface engineered within SATIN and present the proposed radio resource management (RRM) strategy fitting its constraints. We then proceed with the description of the packet scheduler role, insisting mainly on the differences from the respective function in T-UMTS. The formulation of the problem in mathematical notation follows, before the two scheduling disciplines, adaptations of well-known schemes with broad use in wired networks, are presented. The two schemes are evaluated via simulations, addressing mainly their efficiency to satisfy the packet-level QoS requirements of individual services under worst-case (high-load) scenarios, but also their potential to achieve efficient use of the system power resources.

---

[1] Satellite UMTS IP-based Network (SATIN), www.ist-satin.org

2. PROPOSED S-UMTS RADIO INTERFACE AND RRM STRATEGY

The main feature of the proposed access scheme is the lack of satellite return link[2], prohibiting the real time interaction between user equipments (UEs) and Satellite-Radio Network Controller (S-RNC), physically located within the satellite Gateway node. Connections over the air as regards the satellite radio interface, Packet Data Protocol (PDP) messaging between User Equipment (UE) and Satellite Radio Network Controller (Satellite-RNC) and feedback/loop mechanisms (power control, ARQ) are apparently not possible, whereas the configuration of lower layers is based on unidirectional signalling from S-RNC to UEs, carried over common channels.

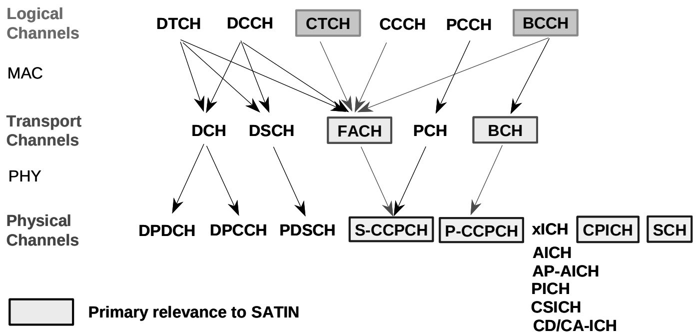


Fig. 1. UTRA FDD logical / transport / physical channels retained in the SATIN radio interface

Fig. 1 shows the common channels and their mapping foreseen in SATIN, in accordance with the FDD mode of UTRA. The proposed access scheme combines regular broadcasting features with Packet Switched domain concepts. For the multimedia data transport, there is one-to-one correspondence between services and logical channels (common traffic channel - CTCH). The logical channels are then mapped, again one-to-one, to transport channels; the *forward access channel (FACH)* was selected within SATIN, since it necessitates the least changes in the standard UTRA FDD interface. At the physical level, *secondary common control channels (S-CCPCHs)* carry (multiplex) one or more FACHs.

Most of the signalling is carried by the broadcast control channel (BCCH mapped to BCH at MAC level): first System Information Broadcast messages allow reception of in-band service signalling, carried in Broadcast Multicast Control (BMC) messages, onto a *system master CTCH/FACH* (to remain available to all UEs in the cell). The latter then provides service-to-channel mapping, hence allows protocol stack configuration for reception of any other CTCH of interest. Discontinuous reception functionality is combined with BMC resource allocation and scheduling.

From RRM point of view, the lack of real-time interaction between the user and the network renders (downlink) power control irrelevant. It also implies that the satellite RNC has to do without the assistance of user-side measurements regarding the quality of the downlink; the use of such measurements (even if available) would anyway have to be different than conventional unicast T-UMTS. In short, such a unidirectional system imposes some hard limitations to the shorter-term functions of RRM, respectively increasing the weight of the longer-term functions, like dimensioning and the design of broadcast schedules.

Additional requirements for RRM stem from the target service set and more generally from the overall service delivery paradigm. At high-level, one of the fundamental requirements is compliance with the current core requirements of MBMS as specified within the ongoing 3GPP Release 6. The system can be envisaged as a Content Delivery Network (CDN), primarily oriented towards *data streaming* (e.g. audio, video broadcasting, alert and emergency announcements) and *push & store* applications (e.g. infotainment, entertainment, software delivery, webcasting). In the first case the multimedia content is played directly upon reception at the user terminal, while in the second the multimedia contents are stored in a local cache for later processing (pre-stored).

The main RRM functions relevant to data streaming services are the Admission Control (AC), Load Control (LC), Packet Scheduling (PS) and the radio bearer allocation and mapping (RBAM) function. Push & store services are mainly handled by the Broadcast Scheduler (BS). The RBAM module is responsible for the Radio Bearer (RB) configuration, i.e. the estimation of the required number of transport/physical channels and their mapping together with the actual Transport Format Combination Set (TFCS) for each physical channel. This block lies functionally within RRC and is in close connection with the other RRM functional blocks (Fig. 2).

---

[2] A return link via T-UMTS is available but may be used for specific functions (access to content decoding keys, retrieval of multimedia content blocks corrupted on the satellite forward link, fetching content "advertised" in the broadcast/multicast channel but not stored in the local cache)

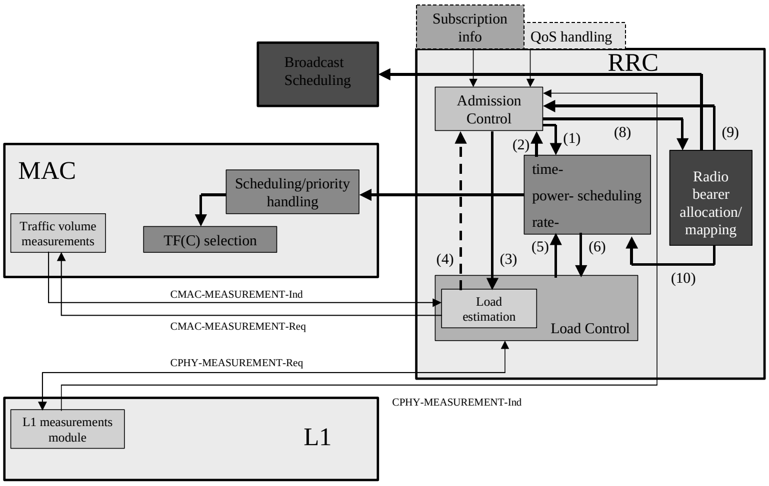


Fig. 2. Model of the RRM entities in SATIN

Two RRM operational modes are possible, each one implying different functionality for some of the aforementioned RRM blocks: in *Mode A* the RBAM dimensions the system for some interval of time, over which the traffic mix remains the same (for example in the order of 1h). Over this interval, the channel mapping and the RB configuration remains fixed. The AC function has to consider the availability of FACHs as this arises from the fixed RB configuration, when deciding on acceptance of a new service. In *Mode B* the RB mapping is drawn in an ad-hoc manner by the AC without any prior configuration. The AC decision upon the acceptance or rejection of a service request is made only on the basis of power or load constraints, without the additional constraints coming from FACH multiplexing [4].

In both modes, the packet scheduler will have to cope with a number of S-CCPCHs, whereupon a set of CTCH / FACHs are mapped. In mode A, the mappings are drawn once from the RBAM function and what changes over time is the status of the individual pre-configured FACHs, which may be idle or active, i.e. a service flow is running over them or not. In mode B, what changes over time is the actual channel mapping. At each time instant, the S-CCPCH configuration features only active channels. In both modes, there are two possibilities for mapping services onto physical channels: in the standard *bin-packing* based method, the aim is simply to fill the capacity of the physical channels. In the *power-aware packing* based method, services with similar Eb/No requirements are mapped on the same physical channel, in an attempt to make a more efficient use of power resources [5].

3. PACKET SCHEDULING

The role of the scheduler in the proposed radio interface is apparently different than in the standard T-UMTS case. In that case the packet scheduler allocates the radio resource in short-term having as a significant criterion for its allocations the state of the individual links (channel state). In the case of the SATIN access scheme, information regarding channel state is not available at the scheduler- in any case such information would have to be exploited in a manner accounting for the point-to-multipoint nature of services; decisions about the scheduling of a single service (flow) should consider the state of several links corresponding to the users of each group.

Therefore the role of the packet scheduler is not that dominant in determining the system throughput as it might prove to be in the terrestrial UMTS [6]. The throughput, in terms of number of flows and respective rates, is mainly determined by the admission control function. Nevertheless the scheduler carries the following important tasks:

§ Time-multiplexes flows with different QoS requirements into fixed SF physical channels, in a way that can satisfy these requirements. The latter have to do with the delay jitter and the guaranteed rate for the data-streaming services and the rate only for the broadcast schedules carrying push & store services. Increase of the delay jitter values calls for a respective increase of the buffer allocated for the playout of the stream at the mobile terminal.

§ Adjusts the transmit powers of the flows, not on the basis of channel information but rather on the basis of the packet/transport block size to be served or knowledge of the expected audience distribution within the beam. This power adjustment is not of the same granularity of the power control mechanism but rather limited to a small set of values (section 5.1.3)

3.1. Time scheduling task formulation

The scheduler treats independently at TTI level each physical channel. The exact number of physical channels at a specific time instance and the corresponding mapping of transport channels onto the code channels are defined by the RBAM and/or AC, depending on the RRM operational mode (ref. Section 2).

Let $TBS\ size_{ij}(k)$ denote the k[th] Transport Block Set (TBS) size of the j^{th} FACH, $1 \le j \le N_i$, mapped to the i^{th} S-CCPCH, $1 \le i \le M$. N_i is the number of FACHs mapped to the i^{th} S-CCPCH, while K_{ij} is the TF size of the j^{th} FACH mapped to the i[th] S-CCPCH. We assume that the TBS sizes corresponding to the TFs of each FACH are sorted in increasing order, namely:

$$TBS\ size_{ij}(k) \le TBS\ size_{ij}(k+1),\ 1 \le k < K_{ij} \qquad (1)$$

Each Transport Format Combination (TFC) corresponds to a certain number of bits R_l passed from the scheduler to the Layer 1, upper-limited by the maximum allowed data rate of the physical channel. The scheduler is given L TFCs per S-CCPCH, obeying the limitations of 3GPP standards[3]. The task of the scheduler is to select every TTI and for each S-CCPCH i some "appropriate" TFC l, $1 \le l \le L$, featuring a certain TBS size -*TBS (l,m)*, $1 \le m \le N_i$ - for each one of the N_i FACH channels mapped to it. The actual context of the term "appropriate" is dictated by several factors, like the service QoS requirements and the physical channel utilization efficiency, and differentiates the one scheduler from the other. This differentiation is summarized in the term scheduling discipline, i.e. in the way the semi-statically fixed capacity of S-CCPCHs is time-shared among the different FACHs.

In the following, two of the possible disciplines are analyzed further. Both of them are adaptations of well-known scheduling disciplines that have been used for years in the context of wired networks.

3.2. Multi-level priority queuing (MLPQ) - based scheme

This is effectively the adaptation of the multi-level, non-pre-emptive priority discipline to the WCDMA context. In our case a CTCH queue at the RLC level may carry one service/flow or a broadcast schedule carrying multiple services (content types). The original scheme favours by default the high priority classes, being able to assure minimum delay for their packets, while it provides no guarantees for lower priority classes.

The N_i FACHs mapped to a single code channel i are ordered from 1 to N, according to their priority. The usual convention is followed, i.e. a lower order number implies a higher priority. The choice of the proper TFC for a given TTI includes some search over the possible TFCs envisaged within the TFC set (TFCS) of the code channel.

The scheduler first seeks to allocate the maximum TBS size to the first FACH. If the queued data q_1^i (in bits) are more than the maximum supported TBS size (in bits) for this FACH in the TFCS, the selected TBS size will be the maximum one available in the TFCS. Otherwise, the selected TBS size is the minimum available in the TFCS that can serve the queued bits, padding being applied when the match between queued data and TBS size is not exact:

$$\begin{aligned} &if: \quad q_1^i > TBS\ size(K_{i1}) \quad TBS\ size(i,1)^* = TBS\ size_{i1}(K_{i1}) \\ &else:\ TBS\ size(i,1)^* = TBS\ size_{i1}(n'), \qquad (2) \\ &\text{with } n' = \min_z \{ \mathrm{z : TBS\ size(1,z)} \ge \mathrm{q_1^i} \} \end{aligned}$$

Out of the whole TFCS, a reduced TFCS, $TFCS_R^1$, is derived for each physical channel:

$$TFCS_R^1 = \left\{ \bigcup_{l \in TFCS} TFC_l : TBS(l,1) = TBS\ size(i,1)^* \right\}$$

---

[3] 3GPP TS 25.302, 25.133/ 25.306

The procedure is repeated recursively for each one of the N_i -1 remaining FACHs, namely for each FACH *j*:

$$
\begin{gathered}
if:\ q_j^i > TBS\ size(K_{ij})\ \ TBS\,(i,j)^* = TBS\ size_{ij}(K_{ij}) \\
else:\ TBS\ size(i,j)^* = TBS\ size_{ij}(n') \\
n' = \min_z \{ \mathrm{z : TBS\ size(j,z)} \geq \mathrm{q}_\mathrm{j}^\mathrm{i} \}
\end{gathered}
\qquad (3)
$$

where the search is now over the reduced TFCS that came out of the previous step:

$$
TFCS_R^j = \left\{ \bigcup_{l \in TFCS_R^{j-1}} TFC_l : TBS(l,j) = TBS\,\mathrm{size}(i,j)^* \right\} \qquad (4)
$$

When more than one CTCHs, of the same priority, are multiplexed –at transport or physical layer (via FACHs)- on a single S-CCPCH, the channels may be served in round-robin mode.

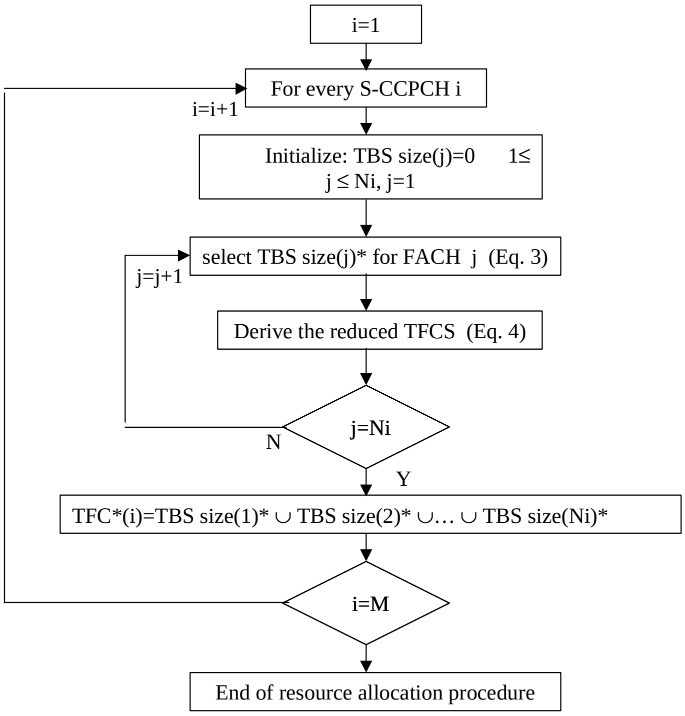


Fig. 3. Outline of the SATIN-adapted version of Multilevel Priority Queueing-based discipline

3.3. Weighted Fair Queuing (WFQ)-based scheduling

This scheme was motivated by the well-known features of WFQ: capability to guarantee a minimum bandwidth per bearer/flux or per set of bearers grouped together for traffic handling purposes, feasibility of upper-bounded queueing delays, fairness in bandwidth-sharing among flows in accordance with the weights assigned to them.

The proposed WFQ scheduler is more specifically based on the *Virtual Spacing* policy that uses the notion of *Virtual Time* [7] and involves the following parameters:

r_i spacing rate or "weight" associated with bearer / CTCH *i* ($T_i = 1/ r_i$): corresponds to the share of the multiplex capacity allocated to the RLC queue

$TSTP_i$ Time Stamp associated with bearer / CTCH *i*: tags the packets at their arrival and is used to order the scheduling

TV Virtual Spacing Time of the system: the time-stamp of the last packet sent out of the queues, i.e. last packet served or being served.

The weights are primarily set according to the rates of the services multiplexed (with respect to the bandwidth-sharing nature of WFQ). Such a distribution can be adapted whenever necessary (e.g. as a response to a new service admission or change of channel mapping configuration). The modification of spacing rates is applied to the packets not served (i.e. new packets as well as packets stored).

The virtual spacing scheme is applied simultaneously to all active radio bearers of the multiplex and consists of the following three processes:

3.3.1. Reception process

When a packet arrives at the queue of CTCH *i*, it is stored in the packet memory buffer, as illustrated in Fig. 4. The new value for $TSTP_i$ of the packet is computed - $TSTP_i$=min (TV, $TSTP_i+T_i$) - and the address of the packet is written in the corresponding line of the scheduler (Fig. 4). If no other packet address is written in the corresponding line, the occupancy bit is set to one.

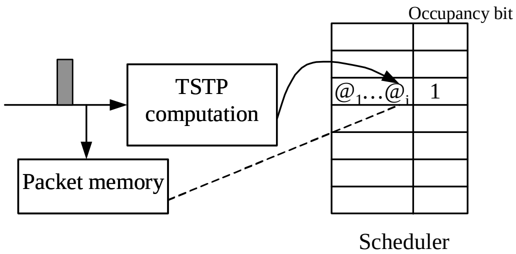


Fig. 4. Virtual Spacing: Arrival of a packet

3.3.2. Scheduler emission process

The scheduler emission process is executed at each TTI, i.e. the global Linked List of addresses of packets to be re-emitted (LLr) is scanned, considering the status of the LLr: if the LLr is empty (or about to become empty), the lowest time (stamp) in the scheduler where there are packets to re-emit is "scanned". The linked list of packet addresses to be re-emitted in that line of the scheduler is added to the global re-emission Linked List associated to the scheduler. After emission from the scheduler the occupancy bits are reset accordingly and the Virtual spacing time of the system (*TV*) is set to the Time Stamp of the scheduler line that has been delivered ($TV = TSTP_{new} = TSTP_{lowest}$) - Fig. 5.

The so-called scanning process or equivalently the Time Stamp sorting problem (see [7] for details) is the heart of the implementation issues as regard Weighted Fair Queueing policies in general, as the time constraint becomes stringent for high data rates (Mbits/s). In our context, the problem is less dramatic due to the lower rates considered (< 384kbps per code channel).

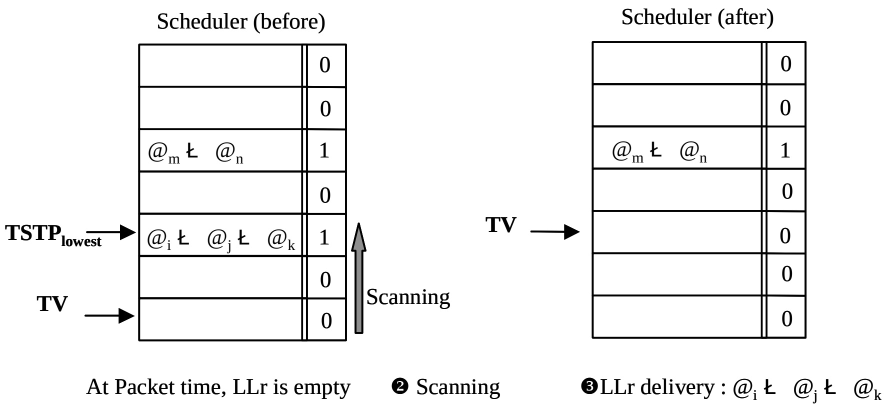


Fig. 5. Scheduler emission process

3.3.3. Re-emission process

There is a unique global re-emission Linked List (LLr) associated to a scheduler. The re-emission process includes the extraction of packets from the memory according to the contents of this List and their emission.

The delivery process has to be done in such a way that no service slot (transmission interval) be ever lost due to any lack of synchronisation between the processes. At each packet time (of the transmission channel), it is important that the processes be executed in strict order: reception, scheduler emission, re-emission.

4. SIMULATION SET-UP AND METHODOLOGY

The simulations with the scheduler focus mainly on the efficiency of the time-multiplexing function. The simulations address the worst-case scenarios for each S-CCPCH, i.e. the intervals over which all FACHs mapped to a physical channel are active. This is the least favourable scenario from the services point of view, i.e. regarding the potential of the PS to guarantee the rate and the jitter requirements of the streams. In other words, these cases can be regarded as a lower bound on the PS performance, as perceived by the flows, and correspond to maximum buffer requirements at the terminal side, in order to absorb the jitter generated at the RLC queues.

The services considered in the simulations are given in Table 1. Services under streaming category correspond to separate streams, mapped one-to-one to logical channels (CTCHs). Push & Store services are organized into broadcast schedules [5] and are treated with lower priority. There is no guaranteed rate attribute for individual services but some mean rate should be maintained for the whole schedule (one CTCH).

Regarding traffic modelling, the main problem we had to face was the lack of models for streaming applications. Few studies are available in the literature mainly because of the proprietary protocols used for such applications and the limited insight to their code. To our knowledge, only one protocol has been proposed for streaming audio, the RealAudio model in [9]. For our simulations, the exponential ON-OFF model with high activity factor (0.8) was retained as the reference model for audio streaming. Additional simulations have been performed with the RealAudio model in order to show the impact of packet-level dynamics (i.e. different short-term burstiness characteristics) on the obtained delay and jitter values. For video streaming services we made use of trace files, available on-line [10].

The channel mapping and the respective RB configuration are the outcome of an elaborate procedure that has as starting point the user profiles. On the basis of these profiles, estimations about the number of the system subscribers and their evolution as well as assumptions about the audience (popularity) of individual services –to account for the point-to-multipoint nature of the services- the traffic load at the system level[4] may be computed. Five different traffic mixes resulted from considering the system at different stages of its life (earlier, later featuring different subscriber numbers), in different hours of the day (morning/ afternoon busy hours) and under different assumptions about the service offer in its early operational years. In the following we maintain the first traffic mix for demonstration purposes [4]. This traffic mix is input to the dimensioning procedure (RRM mode A) and / or AC to provide the mappings depicted in Table 3 and Table 4 for the bin-packing based and power-aware packing based methods.

Table 1. Services considered for simulations and their QoS characterization (class, traffic handling priority)

Service category	UMTS QOS class	Service type	Guaranteed rate (kbps)	Traffic handling priority
Data streaming services	Streaming	Audio Streaming	32 / 64	1
	Streaming	Video Streaming	64 / 128 / 256	1
	Interactive[5]	Location Based Services	16	2
Push & Store services	Background	Webcasting	N/A	Normal
		Rich audio/video info	N/A	High
		Pre-stored movie on demand	N/A	Low
		Pre-stored video clips on demand	N/A	Low
		Pre-stored radio on demand	N/A	Low
		Pre-stored music on demand	N/A	Low
		Software download	N/A	Normal

[4] The system level in this case is a single satellite beam served by an S-RNC. However, extension to different scales is straightforward.

[5] The use of the word interactive in this table deviates a little from the original context of the word interactive within 3GPP. LCS are provided via the streaming service delivery mechanism, i.e. they are not cached. However they do not impose the strict per-packet requirements of audio/video streaming but rather a requirement for the total delivery time of the content.

Table 2. Traffic models considered in simulations

Service	Traffic models	Packet size (bytes)	Model parameters / info
Audio streaming Real Audio	Exponential On-Off Structural model-empirical cdfs	500	Activity factor=0.8 Mean off duration=100ms Idle intervals: multiple of 1.1 secs
Video Streaming	Video trace	500	H.263 files at 64/128/256 kbps target bit rate
LCS	CBR	120	-

Table 3. Mapping for traffic mix 1 (power-aware based, amended)

S-CCPCH	**1-2**	**3**	**4**	**5**	**6**	**7-8**
SF	16	16	16	8	8	8
Data Streaming (DS) FACHs	1x32 4x16	1x64 2x32	2x64	1x128 2x64	2x128	1x256
Sum	96	128	128	256	256	256
Push & Store FACHs	1x48	1x20	1x20	1x48	1x48	1x52

Table 4. Mapping for traffic mix 1 (bin-packing based, amended)

S-CCPCH	**1**	**2**	**3**	**4**	**5**	**6**
SF	8	8	8	8	8	8
Data Streaming (DS) FACHs	1x256 1x32 1x16	1x256 1x32 1x16	2x128 1x32 1x16	1x128 2x64 1x16	2x64 2x16	1x64 1x32 2x16
DS Sum	304	304	304	272	160	128
Push & Store FACHs				1x32	1x144	1x176

For the simulations presented in the next sections, the BLER=f(E_b/N_0) for the combined SAT-IMR [8] channel are considered.

5. PERFORMANCE EVALUATION

Illustration of all metrics and aspects addressed in [8] within the space limitations of this paper is clearly impossible. Nevertheless, we select indicative illustrations that illuminate the main outcome of this study and support the discussion that follows.

5.1.1. Delay and Jitter at S-RNC

The general trend in the MLPQ-based scheme is in agreement with the well-reported behaviour of the ancestor scheme in wired networks: low-priority channels feature high delays values per packet and also high variation of these values (Table 5 and Table 6) Although data carried over these FACHs come on broadcast schedules with a CBR rate, their emission over the air follows less regular patterns: data are buffered for some interval and then are sent in bursts, when the channel is freed by higher priority services. The over-the-air rate of these FACHs effectively oscillates between zero and higher rates, to the extent allowed by the TF of these channels. For example, in the case of FACH 4 @ S-CCPCH 5 the transport channel is not allowed to send for 70% of the simulation time (equal to the mean duration of data streaming services) and when it transmits, it does so using TBS sizes much higher than the ones corresponding to its mean rate or its packets (Fig. 7). Nevertheless, from the service point of view, this does not introduce problems, since there are no packet-level QoS requirements for push & store services (elastic), apart for the provision of a constant rate over longer-than-TTI time intervals (Fig. 6).

The additional flexibility in comparison with the standard MLPQ scheme with exhaustive service is provided by means of the TFCS, which allows compromise of the privileged service higher-priority flows obtain with better support of lower-priority ones. However the penalty is related to the larger TFCS size that in some cases approaches the upper constraints defined by 3GPP (128 TFCs per code channel).

Preliminary results from the Weighted Fair Queuing-based PS showed that fairness between the multiplexed flows implies significant delay/jitter degradation performance for flows with the highest QoS requirements (streaming), when the sharing relies only on the bearer weights. Therefore the WFQ-based PS per S-CCPCH proceeds as follows: bearers are served from highest to lowest QoS, and the actual WFQ scheme (using Timestamps, see section 3.3) is eventually applied to flows of the same QoS class. The jitter values are then similar to the ones obtained by the MLPQ-based scheme. The TF set determination was driven by minimum padding ratio, while a minimum TBS size threshold was introduced in the derivation of TFC Set (threshold set to 83 bytes in the following assessment). in any case WFQ behaviour being a priori affected/curbed by the TFCS constraint. Experimentation with smaller TFCS sizes resulted in lower physical channel throughput and even worst performance by the MLPQ-based scheme.

5.1.2. Fairness

In general, some notion of fairness is desirable among flows of the same traffic handling priority, i.e. two streaming flows should not see much difference in the obtained performance. In the case of MLPQ, an enhancement was made to the scheduler so that the order of serving FACHs of the same traffic handling (TH) priority alternates cyclically on a TTI basis. Apparently, this is a way to smoothen performance discrepancies, although the actual delay and jitter values are still heavily dependent on the behaviour of the flow at TTI-scale. The scores of the scheduler regarding fairness are better in the power-aware mapping case than in the bin-mapping case. The reason for this is that the former tends to map flows of similar rate onto the same code channel. Then the combination of similar TFs with the cyclic alternation of the serving order of FACHs of the same TH priority can lead to similar performance for the flows, i.e. tuning of the TFCS is easier. Nevertheless, the potential to achieve similar performance in terms of jitter is hard-limited by the actual packet-level dynamics of the flows, namely burstier flows feature longer upper tails in the jitter cumulative distribution function (similar to what is shown later in section 5.2).

Contrary to the MLPQ-based scheme, in the WFQ-based scheme it makes little difference whether the underlying mapping is power-aware or not. In general, the fairness scores of the two schemes are similar, although in the case of MLPQ-based more sensitive to the actual TFCS selection.

5.1.3. Power considerations

The link-level simulations [8] indicated that performance, in terms of the required Eb/No in order to achieve a target BLER, improves with higher Transport Block (TB) size[6], since larger TBs interpret into larger turbo-coded packets, hence larger internal interleaver and better Turbo-Code performance.

In light of this result, there are several approaches in setting the power per physical channel. More ‘static’ approaches may set the power according to the Eb/No requirements of the minimum TB appearing in the S-CCPCH TFCS or the TB size corresponding to the mean rate of the most demanding service and keep it fixed till a new flow is mapped on this S-CCPCH or one of the existing one is terminated. In both cases the power setting may change if these events generate a change of the aforementioned parameters. In a more dynamic -and demanding in terms of processing-approach the scheduler can make use of lookup tables and change the power setting at TTI level according to the highest Eb/No requirement arising from the particular TFC choice of the scheduler in the respective TTI.

Fig. 9 compares the Eb/No requirements arising from the PS decisions in the case of MLPQ-based scheme against more static settings of AC, on the basis of the lowest-guaranteed-rate service Eb/No requirement. It comes out that the decisive factor from a power saving point of view is the TFCS and the respective TB sizes that are made available to the scheduler rather than the way the mapping is performed (power-aware or unaware). In fact, the proposed Packet Schedulers (PS) confirm the challenging task of determining optimized TF/TFCs, and more particularly the relevance of TBS size thresholds: the TF/TFCs should not include TBS sizes smaller than some minimum threshold, so as to relax Eb/No requirements and spare system power but in the same time should provide small enough values to achieve better physical channel utilization and the same time meet the rate requirements of the services of the lowest QoS priority.

5.2. <u>Impact of packet-level dynamics</u>

The measured statistics were obtained under certain modelling assumptions. It was deemed interesting to investigate the sensitivity of these assumptions upon these modelling assumptions.There was no specific input from scientific literature favouring the choice of the ON-OFF model; it was selected mainly due to its simplicity and the associated easiness in controlling the traffic dynamics on the basis of a couple of parameters.

The single model we found in literature for streaming is the one in [9]; it has taken a structural modeling approach and relies on measurements of RealAudio traffic. It is reported there that the flows appear to have a constant bit rate when the latter is measured over scales of seconds or tens of seconds but exhibit burstiness over shorter intervals. In fact,

[6] The performance improves monotonically for TBS sizes up to 5000 bits –maximum TB size allowed by 3GPP specifications. Larger TBS sizes yield two transport blocks of smaller size, hence performance will equal that of TBS size/2

Table 5. Delay experienced by packets at the RLC queues (S-CCPCH 6)

S-CCPCH 6	Mean	Standard Dev.	Median	Percentile_90	Percentile_99
FACH 3	1.884743	1.019141	1.9933	3.1066	3.7333
FACH 2	0.076676	0.045593	0.0666	0.14	0.2066
FACH 1	0.079672	0.046208	0.07	0.1457	0.2066

Table 6. Delay variation experienced by packets at the RLC queues (S-CCPCH 6)

S-CCPCH 6	Mean	Standard Dev.	Median	Percentile_90	Percentile_99
FACH 2	0.029012	0.031186	0.0143	0.0533	0.1666
FACH 1	0.028459	0.030892	0.0143	0.0533	0.1666

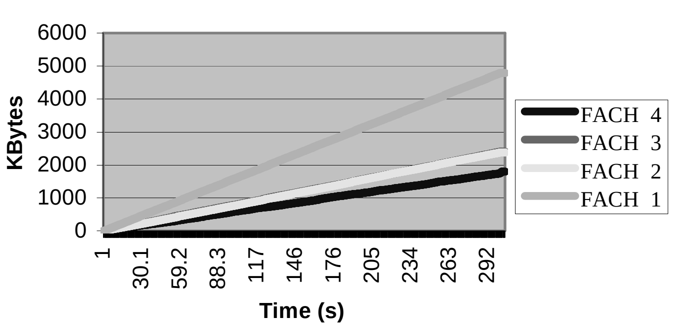


Fig. 6. Per FACH cumulative transferred data on S-CCPCH 5

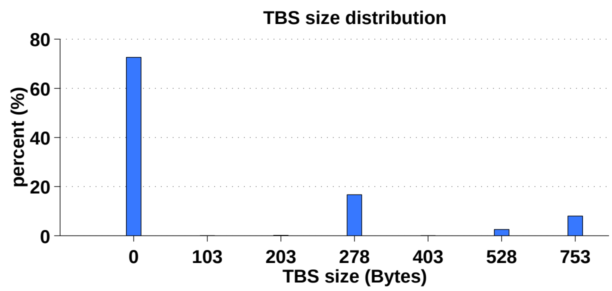
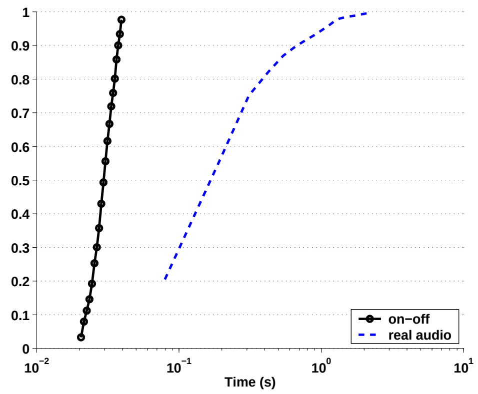


Fig. 7. Transport block set size distribution for FACH 4 (right) – power aware based mapping



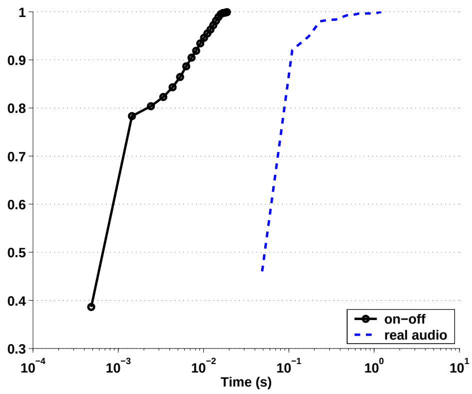
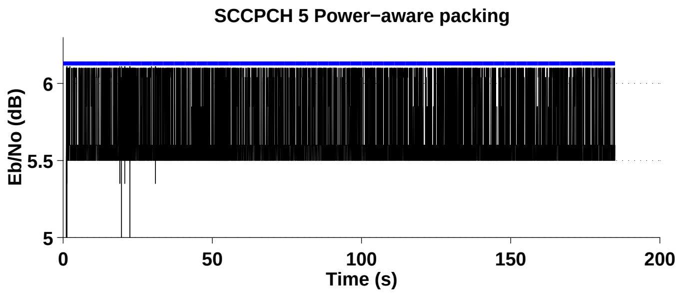


Fig. 8. Comparison of delay and delay jitter between the exponential on-off model and the RealAudio model



AC estimation = 6.13dB
Effective E_b/N_0 =5.6dB
Max E_b/N_0 = 6.13dB

Fig. 9. Eb/No requirements arising from scheduler versus Eb/No considered by AC

packets are sent in bursts following idle intervals of the order of second(s). These intervals were found integer multiples of a specific time interval, whose value was attributed to the multitasking nature of the operating system (context switches). On the basis of these measurements, a model was implemented and incorporated in ns2, reproducing this behaviour with the help of lookup files providing the empirical cdfs of a number of parameters (flow duration, length of idle periods etc).

It becomes obvious from Fig. 8 that the impact of the model is anything but negligible. There are differences approaching two orders of magnitude. These differences are the results of the radically different packet-level dynamics of the two flows (models) rather than the outcome of inefficient, non-optimized scheduling.

6. CONCLUSIONS

In this work, we addressed the packet scheduler role in the delivery of point-to-multipoint services within a satellite radio interface relying on the common channels of the UTRA FDD interface. Given the transport channel choice (FACH), the time scheduling of the different FACHs that have been mapped to a single S-CCPCH is the major task of the packet scheduler. We proposed adaptations of two schemes that are well-known from the context of wired networks for this task, an MLPQ-based scheme and a WFQ-based scheme.

The evaluation of the two proposed Packet Schedulers (PS) confirmed the challenging task of determining optimised TF/TFCs, and more particularly the relevance of TBS size thresholds in meeting the rate requirements of all services and achieving high physical channel utilization, whilst preserving the power resources of the system.

Regarding fairness between flows of same QoS class or physical channel utilization it makes little difference for each PS scheme whether the mapping is power-aware or not, although the tuning of the TFCS appears to be easier in the pure bin-packing case with the MLPQ-based PS. The nature of the WFQ-based PS is curbed by the TFCS selection process and the packet-level dynamics of the flows in many multiplexing cases have not allowed the significant reduction of TF range, which may be foreseen with WFQ, compared to MLPQ. The trade-off between small TFCS and flexibility regarding the fixed-SF channel utilization appears to be a demanding exercise.

The TFCS also carries significant weight on the power consumption of the system. On the contrary, the actual way to perform the mapping (power-aware or not) seems to be less significant. In short, irrespective of the PS scheme the optimization of the system capacity/throughput depends on the cautious trade-off between power saving and multiplexing effectiveness.

REFERENCES

1 ASMS-TF, The Vision of the Advanced Satellite Mobile Systems Task Force, September 2001

2 3GPP TR 23.846, “Multimedia Broadcast/Multicast Service; Architecture and Functional Description”.

3 T. Severijns et al “The intermediate module concept within the SATIN proposal for the S-UMTS air interface”, IST Mobile Summit 2002, Greece

4 E. Angelou, N. Koutsokeras, I. Andrikopoulos, I. Mertzanis, M. Karaliopoulos, P. Henrio, Admission and preventive load control for delivery of multicast and broadcast services via S-UMTS, First International Conference on Advanced Satellite Mobile Systems, Frascati, Italy, July 2003

5 M. Karaliopoulos, E. Angelou, I. Andrikopoulos, P. Henrio, B.G. Evans, Radio Resource Management Strategy in SATIN, 12th IST Mobile & Wireless Communication Summit, Aveiro, Portugal, June 2003

6 H. Holma and A. Toskala, *WCDMA for UMTS radio access for third generation mobile communications*, John Willey & Sons Ltd

7 J.W. Roberts, Virtual Spacing for flexible Traffic Control, *International Journal of Communication Systems*, Vol.7, 1994

8 SATIN IST 2000-25030 Deliverable No. 7, Simulation Results and Evaluation.

9 K-C Lan, J. Heidemann, Structural modeling of Real Audio traffic, ISI-TR-544, September 2001

10 F. Fitzek, M. Reisslein, MPEG-4 and H.263 Video Traces for Network Performance Evaluation, TKN Technical report, October 2000, available on-line: http://www-tkn.ee.tu-berlin.de/research/trace/trace.htm