

Joint Time Slot Optimization and Fair Bandwidth Allocation for DVB-RCS Systems

Antoni Morell, Gonzalo Seco-Granados and María Ángeles Vázquez-Castro
Universitat Autònoma de Barcelona (UAB)
Telecommunications and System Engineering Department (TES)
e-mail: {Antoni.Morell,Gonzalo.Seco,Angeles.Vazquez}@uab.es

Abstract—This paper introduces a novel operational framework for the problem of time slot assignment in a Digital Video Broadcast-Return Channel via Satellite (DVB-RCS) system. The approach is compliant with the latest technical specifications emitted by the European Telecommunications Standards Institute (ETSI) about Quality of Service (QoS) in Satellite Earth Stations and Systems (SES). It is a cross-layer MAC-PHY optimization approach sustained by the powerful framework of convex optimization. The paper proposes a hierarchical dynamic bandwidth allocation approach, which is motivated by the computational complexity of the single-step solution. More specifically, we obtain and analyze the optimal time duration of the time slots and jointly, we make a fair allocation of slots to areas, which is the highest level in the bandwidth allocation hierarchy. Results show up to a 10% increase in transported capacity.

I. INTRODUCTION

The actual evolution of the telecommunications market is clearly demanding broadband systems. Examples can be found in the broadband Internet (including WLAN [1]) or the new cellular or digital television standards. Among them, the Digital Video Broadcasting - Satellite (DVB-S) is a widely accepted standard in broadband satellite communications.

The second generation (DVB-S2) includes the transmission of multimedia traffic and a variety of connection types (from broadcast to unicast or even multicast). Applications such as voice over IP, teleconference, video streaming, web browsing, ftp, etc. will also be embedded in a satellite terminal giving a huge potential package of services to the end-user. We focus on unicast interactive applications that exploit the rate flexibility granted by an adaptive physical layer.

These broad required system potentialities ask for interaction and thus, a return link over the satellite is mandatory. Besides using terrestrial networks for interactive purposes, these can be reached using Digital Video Broadcast-Return Channel via Satellite (DVB-RCS) systems [2].

Users request satellite capacity function of their needs and hence, it is a Bandwidth on Demand (BoD) solution. A centralized algorithm computes then the amount of resources assigned to each user. This Demand Assigned Multiple Access (DAMA) approach is supported at the physical (PHY) layer by a Multi Frequency Time Division Multiple Access (MF-TDMA)scheme. The goal of this paper is to present a

⁰This work was supported in part by the IST-507052 SatNex Network of Excellence, MEC project ESP2005-03403 and by an ESA project.

scheduling framework to deal with the multiple access and to give the first performance analysis and design rules.

II. TIME SLOT ASSIGNMENT: SYSTEM OVERVIEW

We consider a transparent satellite network as depicted in Figure 1. It focuses the DVB-RCS situation. Many RCS terminals (RCSTs) emit their capacity requests to the Network Control Center (NCC), which depend on the queued traffic at each terminal and hence they are normally related to link and/or network layers.

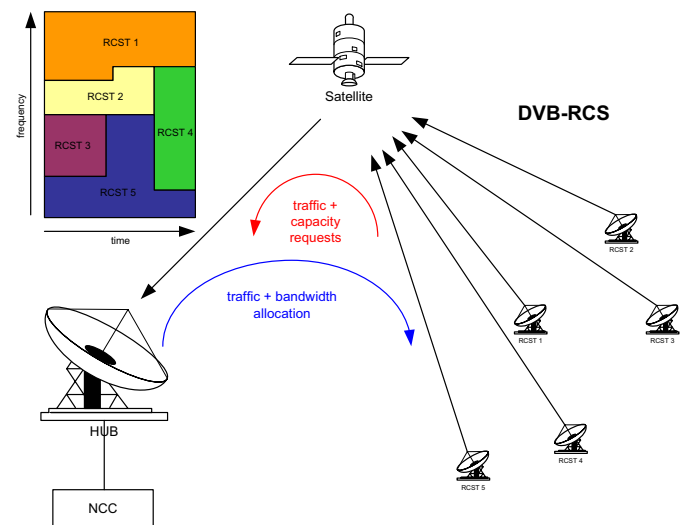


Fig. 1. System overview.

In the DVB-RCS standard, three types of capacity request, from highest to lowest priority, are considered (we obviate the Free Capacity Assignment or FCA [3], which may be granted by the NCC, but not requested):

- Constant Rate Assignment (CRA): the RCST requires a constant rate all the time.
- Rate Based Dynamic Capacity (RBDC): a bandwidth request (in rate capacity) remains effective until it is updated or timed out. In contrast to CRA, RBDC strategy allows for statistical multiplexing among many terminals, resulting in a more efficient use of satellite bandwidth.
- Volume Based Dynamic Capacity (VBDC): it requires for a certain amount of volume capacity to transmit

information regardless the way it is done (no constant rate is needed).

The requests generated by all terminals in a beam constitute the inputs of our assignment problem and we do not need to consider how terminals generate them. For each bandwidth allocation update the NCC signals a Terminal Burst Time Plan (TBTP) to the RCSTs. It points out in what frequencies and time zones each of the RCSTs transmits (see Figure 1).

In sections III and IV we deal with the assignment of time-frequency zones to the RCSTs.

III. PROPOSED FRAMEWORK

In the DVB-RCS, the TBTP is updated and transmitted every Superframe (SF) and it is composed of several Frames (F) of duration T_F . If the total system bandwidth is BW , we assume that the scheduler solves an allocation problem for each $BW \times T_F$ block, although it can be applied without loss of generality to the whole SF. The system bandwidth is usually divided into different carrier types (of different bandwidth) to accommodate different users accounting for different Service Level Agreements (SLAs) and terminal equipment, location, etc.

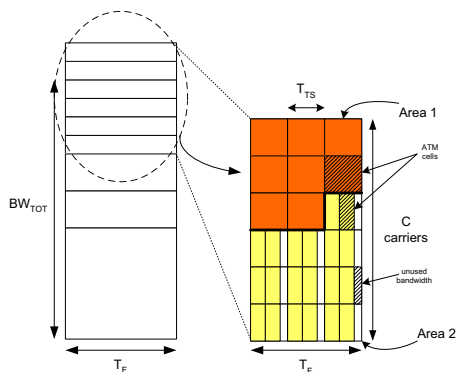


Fig. 2. Scheduling (bandwidth allocation) problem.

We now impose some problem constraints, extracted from the ETSI technical specification [4]:

- The total transmission capacity in the satellite beam is divided in areas (described later).
- All carriers in one area must have the same symbol rate and slot timing.
- A given RCST belongs to one, and only one, area.
- A given RCST can use only one carrier at a given time.

So the problem can be divided into K subproblems, one for each group of carriers of the same type. We illustrate all this in Figure 2. Without loss of generality, the problem to be treated consists in multiplexing N users into C carriers of BW_i bandwidth that transmit during T_F seconds. We can consider that the users in one area are the ones that, while transmitting in a common carrier type, use the same transmission rate. Recall that the DVB-RCS standard yet defines an adaptive-coding PHY layer with several possible coding rates, so the mapping of users to areas is defined by the physical quality of

the transmission (channel conditions). We also consider that the minimum transmission unit is a layer-2 (MAC) packet, either it is an Asynchronous Transfer Mode (ATM) cell (53 bytes) or a Moving Picture Experts Group (MPEG) container (188 bytes). Without loss of generality, we will consider ATM cells in the rest of the paper.

As a result of the previous discussion, the goal in our proposal is to cope with TBTP reduced signalling at the expenses of a reduction in bandwidth efficiency (in contraposition to other proposals such as [5]). We impose a time slot whose duration is common to all areas. A very simple assignment procedure is defined (once the number of time slots per area is known) from left to right and from top to bottom (the reading order). Then, depending on the area rate, one or more ATM cells can be transmitted in a single time slot. The same assigning order as with areas is used, assuring that a given RCST can transmit up to the maximum possible ATM cells (this number corresponds to the number of cells that would fit a fully used carrier). Furthermore, we assume that this threshold value is not exceeded. See in Figure 2 a possible time slot and ATM cell assignment. The problem of how to assign time slots to areas and ATM cells to RCSTs is discussed later.

We now introduce the scheduling hierarchy concept proposed by the ETSI in [4] and include it in our framework.

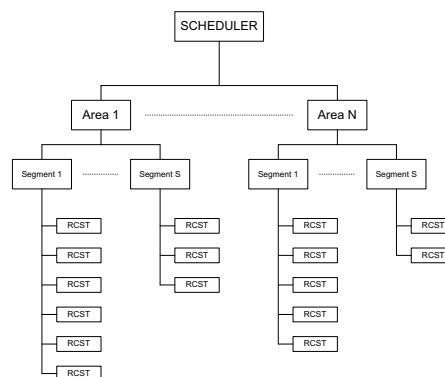


Fig. 3. Scheduling (bandwidth allocation) hierarchy in DVB-RCS.

A. Scheduling hierarchy

Our framework simplifies the general scheduling problem (with maybe thousands or more RCSTs) to some smaller problems by imposing some known structure (it facilitates signalling as well). It is a similar idea the authors in [6] suggest. In concordance with [4], we must guarantee some minimum resources to the Service Providers (SPs). As SPs can have RCSTs distributed over different areas, the scheduling hierarchy in [6] introduces the segment concept, which is a grouping of RCSTs inside a given area along with a minimum predefined amount of resources associated (see Figure 3).

Then, the scheduling strategy we propose is (you can think of resources as ATM cells):

- Aggregate the users' requests in each area and the minimum assigned resources of the segments in the area.

- Assign the minimums to the areas.
- Distribute the remaining resources with a “fair” algorithm, as discussed next.
- For each area:
 - Aggregate the users’ requests per segment.
 - Assign the minimums to the segments.
 - Distribute the remaining resources among segments (with a “fair” algorithm).
 - Finally, within each segment:
 - * Distribute the resources among users depending only on their CRA requests (most priority).
 - * Distribute the remaining resources depending only on the RBDC requests.
 - * Distribute the remaining resources depending on the VBDC requests.

Next, we present the proposed fair allocation solution.

B. Fair resource allocation

It is shown in [7] that a fair distribution of P resources among N entities (areas, segments or terminals) is achieved by the following maximization problem (x_i is the amount of resource assigned to entity i):

$$\begin{aligned} \max_{x_1, \dots, x_N} \quad & \prod_{k=1}^N x_k \\ \text{s.t.} \quad & \sum_{k=1}^N x_k \leq P \\ & d_{\min_i} \leq x_i \leq d_{\max_i} \end{aligned} \quad (1)$$

where d_{\min_i} is the amount of resource guaranteed to i and d_{\max_i} is the demand of i .

The main difference with [6] is that we must consider a minimum resource allocation to each entity. This fact changes the solution. We realize that (1) can be easily converted to a convex optimization problem [8] by just applying the logarithm function to the objective. Furthermore, the resulting problem is analytically solvable using the Karush-Kuhn-Tucker (KKT) conditions [8], which imposes the solution

$$x_i = \begin{cases} \frac{1}{\lambda}, & d_{\min_i} \leq \frac{1}{\lambda} \leq d_{\max_i} \\ d_{\min_i}, & \frac{1}{\lambda} \leq d_{\min_i} \\ d_{\max_i}, & \frac{1}{\lambda} \geq d_{\max_i} \end{cases}, \quad (2)$$

where λ is a positive value such that $\sum_{k=1}^N x_k \leq P$.

Graphically, the solution is found by filling a recipient shaped accordingly with the demands and guaranteed resources with a quantity P of water (see Figure 4). Assuming (1) is feasible, the solution first assigns the minimums (“pale water”) and “equally” distributes the rest (“strong water”).

Note that although the solution is computed for a real-valued problem, the particularization to the integer case (i.e., with integer variables x_i) is straightforward. It consists basically in giving one extra resource (round up) to a subgroup of the group of users that share the same number of resources and round down the rest (see the dotted line in Figure 4).

We get now back to the first problem in the scheduling hierarchy of Figure 3, that is, the time slot distribution among areas. In a first design option, it may seem reasonable to set

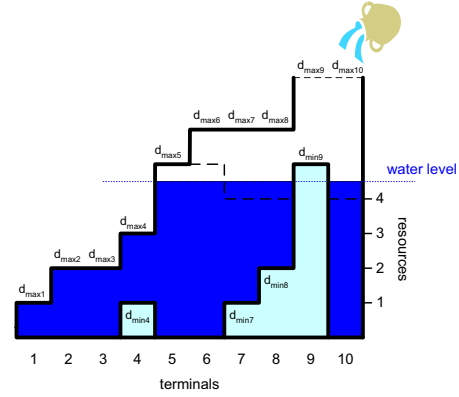


Fig. 4. Fair resource distribution solution.

the time slot duration T_{TS} to exactly fit an ATM cell of the area that is set up with the lowest transmitting rate. Higher rates may transmit more than one ATM cell per time slot. This idea is depicted in Figure 2 (right). Yet with this simple example, we realize that some bandwidth remains unused. It is obvious then to ask: Can we set up T_{TS} so as to reduce this inefficiency? This is discussed in the next section.

IV. AREA SIZE SELECTION AND TIME SLOT OPTIMIZATION

Keeping the fair resource allocation in (1), we optimize also over T_{TS} . Without loss of generality, we do not consider here the minimum guaranteed resources. Being N_{ATM_i} the number of ATM cells assigned to area i , N_i the number of time slots assigned to area i , and T_{TS} the time slot duration, the problem under study maximizes $\prod_{i=1}^N N_{ATM_i}$:

$$\begin{aligned} \max_{T_{TS}, N_1, \dots, N_N} \quad & \prod_{i=1}^N N_i \cdot K_i(T_{TS}, t_i) \\ \text{s.t.} \quad & \sum_{i=1}^N N_i \leq N_{TOT}(C, T_F, T_{TS}) \\ & 0 \leq N_i \cdot K_i(T_{TS}, t_i) \leq d_i \\ & T_{\min} \leq T_{TS} \leq T_{\max} \end{aligned} \quad (3)$$

where t_i is the time duration of an ATM cell transmitted at rate r_i , K_i is the number of ATM cells fitting in a time slot (that depends both on the ATM cell duration and T_{TS}) and N_{TOT} is the total number of time slots (it depends on the number of carriers, the frame duration and the time slot duration).

Note that it is a cross-layer optimization scheme. The input data d_i comes from the MAC layer, although we can also consider some network (NET) cross-layer influence. Furthermore, we require PHY cross-layer information (the area rates r_i) and we affect to the PHY layer of the RCSTs (adjusting T_{TS}).

We can solve the problem by first fixing T_{TS} and optimizing over the N_i 's. We obtain the $N_{i_{opt}}^1$'s. Fixing these values, we can now optimize over T_{TS} and obtain $T_{TS_{opt}}^1$. Iteration of this mechanism would drive into the optimal joint solution if the problem were jointly convex.

Fixing T_{TS} , the problem is convex, and we can rewrite it

as:

$$\begin{aligned} \max_{N_1, \dots, N_N} \quad & \prod_{k=1}^N N_i \cdot \prod_{k=1}^N K_i(T_{TS}, t_i) \\ \text{s.t.} \quad & \sum_{k=1}^N N_i \leq N_{TOT}(C, T_F, T_{TS}) \\ & 0 \leq N_i \leq \lceil \frac{d_i}{K_i(T_{TS}, t_i)} \rceil \end{aligned} \quad (4)$$

where the ceil function ($\lceil \cdot \rceil$) is necessary in the integer case to prevent the counterproductive situation of having one area requesting some ATM cells but receiving 0 time slots. With these simple manipulations, the problem in (4) is equivalent to the integer version of (1) and thus, we know the solution.

Fixing the N_i 's gives the following problem for the time slot optimization (developing expressions for the K_i 's and N_{TOT}):

$$\begin{aligned} \max_{T_{TS}} \quad & \prod_{k=1}^N N_i \cdot \lfloor \frac{T_{TS}}{t_i} \rfloor \\ \text{s.t.} \quad & \sum_{k=1}^N N_i \leq C \cdot \lfloor \frac{T_F}{T_{TS}} \rfloor \\ & 0 \leq N_i \leq \lceil \frac{d_i}{\lfloor \frac{T_{TS}}{t_i} \rfloor} \rceil \\ & T_{min} \leq T_{TS} \leq T_{max} \end{aligned} \quad (5)$$

Clearly, the floor function ($\lfloor \cdot \rfloor$) converts this particular problem into non-convex and thus, the joint problem too. However, we exploit the ‘‘integrality’’ that the function introduces with the following lemma.

Lemma 1: Departing from a feasible value of T_{TS} and increasing it, it can only reduce the objective value unless a multiple value of some of the t_i 's is reached.

In other words, the only meaningful values of T_{TS} are the ones that are multiple of the t_i values and reside inside the interval $[T_{min}, T_{max}]$. The values between any of these special values do not allow to place an extra ATM cell inside any time slot at the expenses of a potential decrease in N_{TOT} . In the case we are considering, with few areas and the same amount of t_i 's, the list of possible T_{TS} values is small and thus (5) can be easily solved via exhaustive (but small) search.

The optimization procedure for the joint problem is then:

- Construct the list of possible values of T_{TS} .
- Reduce the list by suppressing equal values coming from multiples of different t_i 's.
- Optimize the N_i 's for each possible value.
- Finally, get $\{T_{TS}, N_i\}$ with best objective value in (3).

Note that joint convexity is not necessary to guarantee the optimal solution. In the next section, we give some results for the joint optimization and we reflect the importance of taking a good choice of T_{TS} .

V. RESULTS

We assume an scenario with $C = 111$ carriers of $540kHz$ bandwidth and a $T_F = 26.5ms$. We explore in this section a possible DVB-RCS situation, where the RCSTs transmit via 7 different coding rates and thus we have 7 different ATM cells durations. Accordingly, one area per coding rate is defined. See in Table I the relation between areas, coding rates and ATM cells duration. It is assumed a Quadrature Phase Shift Keying (QPSK) modulation transmitted through a raised cosine pulse with a roll-off factor of 0.35, $T_{min} = t_1$ and $T_{max} = 3t_1$.

TABLE I
AREAS DEFINITION.

Area identifier	Coding rate	ATM cell duration
1	$r_1 = 1/3$	$t_1 = 1.59ms$
2	$r_2 = 2/5$	$t_2 = 1.325ms$
3	$r_3 = 1/2$	$t_3 = 1.06ms$
4	$r_4 = 2/3$	$t_4 = 0.795ms$
5	$r_5 = 3/4$	$t_5 = 0.706ms$
6	$r_6 = 4/5$	$t_6 = 0.6625ms$
7	$r_7 = 6/7$	$t_7 = 0.6183ms$

The RCSTs aggregated demand (number of requested ATM cells) per area is computed as follows. We define the Aggregated System Demand (ASD) as the mean of the sum of all demands in all areas. This demand is distributed among the areas using some fixed distribution p . In our case, it makes sense that areas with higher rates accumulate more requests as it is expected that most of the RCSTs are assigned to these areas. Note that low rate areas are designed to fulfill the transmission requirements of areas affected by rain. We use the following distribution $p = [1/28, 2/28, 3/28, 4/28, 5/28, 6/28, 7/28]$. Once the ASD *per area* is known (it is a mean value), we compute a realization of demand in each area using a uniform probability density function (pdf) with the given mean ASD.

We have also defined a reference value for the ASD, which corresponds to the transported capacity by the system when only the highest rate transmits and $T_{TS} = t_7$ (the maximum possible transported capacity). In our case $ASD_{ref} = 4662 ATM_{cell}/frame$, which corresponds to 74.6 Mbps. Note that the ASD can be over that value. Imagine the feasible scenario where the highest rate area asks the reference ASD while the other areas ask their own ‘‘maximum’’ transport capacity (which depends on the area rate and, of course, is less than the reference ASD).

In our results, computed via the Monte Carlo method, we have studied the fair allocation algorithm versus an opportunistic design, both when T_{TS} is optimized (for both designs) and when we set $T_{TS} = t_1$. The opportunistic design optimizes the sum value instead of the product value in (3). The solution consists in allocating all the demand (until there are resources left) to the highest rate area and iterating this procedure for each area (ordered by rate) until the lowest rate area is reached.

Our first analysis in Figure 5 studies the Bandwidth Occu-

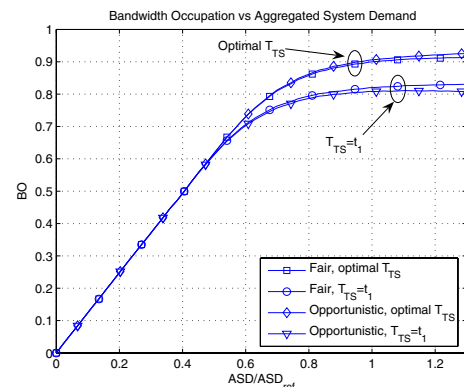


Fig. 5. Bandwidth occupation.

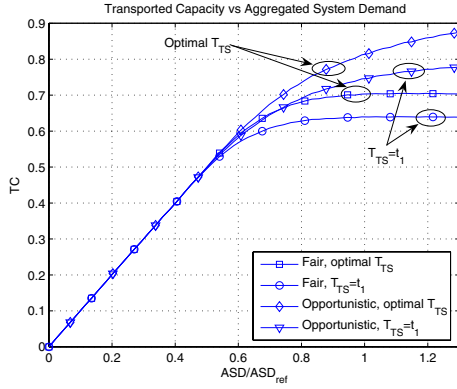


Fig. 6. Transported capacity.

pation (BO), defined as $BO = \frac{\sum_{i=1}^7 N_i \cdot K_i \cdot t_i}{C \cdot T_F}$. It is shown that optimizing T_{TS} improves significantly the occupation for both fair and opportunistic strategies while it reduces the bandwidth occupation differences between the two designs.

We also study the transported capacity defined as $TC = \frac{\sum_{i=1}^7 N_i \cdot K_i}{ASD_{ref}}$. Figure 6 plots the sum of the assigned ATM cells in all areas normalized by the reference ASD value (it is in fact a maximum transport capacity value). Again, optimizing over T_{TS} significantly improves the transported capacity (over a 6% more capacity in the fair case and near a 10% increase in the opportunistic design). This shows that the increase in BO thanks to T_{TS} optimization, shown in the previous figure, effectively translates into an increase in TC. Note that the opportunistic design would reach the maximum TC value as ASD increases (independently of the requests distribution), whereas the fair algorithm will generally saturate in a lower value (between 0.62 and 0.71 in the studied case).

We analyze next the fairness differences between the solutions, using the fairness index definition in [9]. For a given solution $N_{ATM_1}, \dots, N_{ATM_7}$, we define a new solution set $y_1 = \frac{N_{ATM_1}}{N_{ATM_1}^*}, \dots, y_7 = \frac{N_{ATM_7}}{N_{ATM_7}^*}$ and we compute the Fairness Index as

$$FI = \frac{(\sum_{i=1}^7 y_i)^2}{7 \cdot \sum_{i=1}^7 y_i^2} \quad (6)$$

where $N_{ATM_i}^*$ is the most "FAIR" solution obtained with the fair algorithm with optimal T_{TS} (definition). Then we compute the fairness index obtained by the following 2 solutions:

- 1) the fair solution with $T_{TS} = t_1$
- 2) the opportunistic solution with optimal T_{TS}

when compared with the "FAIR" one. The results are shown in Figure 8. Note that whereas solution 1 exhibits good fairness performance, solution 2 reduces fairness significantly.

		Bandwidth Occupation															
small T_{TS}	t_1	$3t_1$	$3t_2$	$3t_3$	$3t_4$	$4t_1$	$4t_2$	$4t_3$	$5t_1$	$4t_4$	$5t_2$	$5t_3$	$6t_1$	$6t_2$	$6t_3$	$7t_1$	$7t_2$
1	0.98	0.80	0.80	0.74	0.61	0.61	0.55	0.55	0.49	0.98	0.86	0.86	0.74	0.74	0.74	0.61	0.61
2	0.82	0.66	0.66	0.61	0.51	0.51	0.92	0.92	0.82	0.82	0.72	0.72	0.61	0.92	0.92	0.77	0.77
3	0.65	0.53	0.53	0.98	0.82	0.82	0.74	0.74	0.65	0.98	0.86	0.86	0.74	0.74	0.98	0.82	0.82
4	0.98	0.80	0.80	0.74	0.92	0.92	0.83	0.83	0.74	0.98	0.86	0.86	0.74	0.92	0.92	0.77	0.77
5	0.87	0.71	0.71	0.98	0.82	0.82	0.74	0.98	0.87	0.87	0.76	0.95	0.82	0.82	0.98	0.82	0.82
6	0.82	0.66	1.00	0.92	0.77	0.77	0.92	0.92	0.82	0.82	0.89	0.89	0.77	0.92	0.92	0.77	0.89
7	0.76	0.93	0.93	0.86	0.72	0.95	0.86	0.86	0.95	0.95	0.83	0.83	0.86	0.86	0.86	0.83	0.83
mean	0.84	0.73	0.78	0.83	0.74	0.77	0.79	0.83	0.76	0.91	0.83	0.85	0.75	0.84	0.90	0.77	0.79

Fig. 7. System occupation analysis.

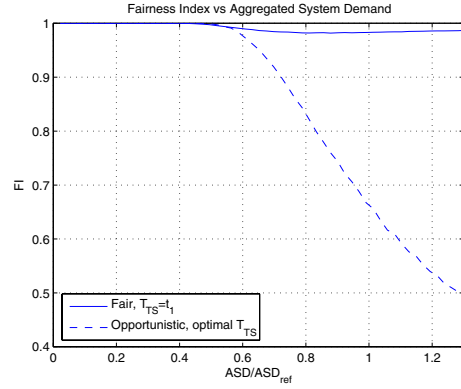


Fig. 8. Fairness study.

Finally, we study the occupation efficiency for the different significant values of T_{TS} when only one area is requesting resources (we assume a very high demand to transmit to the highest extent). The results can be seen in Figure 7. Note that some T_{TS} values exploit better the system occupancy than others depending on which areas we consider active. We have marked the $T_{TS} = 4t_4$ as the configuration that gives better results in the general case (when we assume all areas active), that is, the most robust choice.

VI. CONCLUSIONS

We have introduced a scheduling framework compliant with the latest standards and technical specifications for the DVB-RCS system. In the hierarchical procedure proposed, we have analyzed and optimized the performance of the first bandwidth allocation phase, highly related with the choice of the time slot duration. We have compared both the opportunistic (the best RCST gets the transmission) and the fair (fair transmission distribution) approaches and we have shown the importance of a good time slot choice. A method to choose the T_{TS} is addressed, either it can be optimal depending on the demand or a fixed value is preferred.

REFERENCES

- [1] K. Ahmavaara, H. Haverinen and R. Pichna, "Interworking architecture between 3GPP and WLAN systems," *IEEE Communications Magazine*, Vol. 41, No. 11, pp. 74-81, Nov 2003.
- [2] ETSI, "Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems," *ETSI EN 301 790*, Apr 2005.
- [3] J. Neale and A.K. Mohsen, "Impact of CF-DAMA on TCP via satellite performance," *Proc. of Global Telecommunications Conference, GLOBECOM'01*, Vol. 4, pp. 2687-2691, Nov 2001.
- [4] ETSI, "Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM) Services and Architectures: QoS Functional Architecture," *ETSI TS 102 462*, Dec 2005.
- [5] M.A. Vázquez Castro, M. Ruggiano, L.S. Ronga and M. Werner, "Uplink Capacity Limits for DVB-RCS Systems with Dynamic Framing and Adaptive Coding," *Proc. 23rd In. Com. Sat. Sys. Conf., Rome*, Sep 2005.
- [6] A. Girard, C. Rosenberg and M. Khemiri, "Fairness and Aggregation: A Primal Decomposition Study," *Networking 2000, Lecture Notes in Computer Science 1815, Springer-Verlag*, pp. 667-678, May 2000.
- [7] R. Mazumdar, L.G. Mason and C. Douligeris, "Fairness in network optimal flow control: optimality of product forms," *IEEE Trans. on Comm.*, Vol. 39, No. 5, pp. 775-782, May 1991.
- [8] L. Boyd and S. Vandenberghe, *Convex optimization*, Cambridge University Press, 2003.
- [9] R. Jain, D. Chiu and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems," *Tech. Rep. DEC TR-301, Digital Equipment Corp.*, Sep 1984.