# Joint Congestion Control and Scheduling Algorithm in a Multi-Access and Multi-Service Satellite System

Francesco Delli Priscoli, Gianfranco Santoro, Carlo Colafranceschi David M. Krapez and Emiliano Guainella

***Abstract*— The work proposes a solution to the bandwidth allocation problem in a multi-user satellite telecommunications system, aiming to an efficient radio resource exploitation. This task is performed sharing dynamically, each signalling time, the resources available among the requiring users services, avoiding congestion states and guaranteeing an agreed Quality of Service. The proposed algorithm is fully compliant with the DVB-RCS standard, but due to its independence from the network technology it can be easily adopted in CDMA systems.**

*Index Terms*— **Bandwidth on Demand, Congestion Control, Scheduling, Satellite System, DVB-RCS, QOS.**

## I. INTRODUCTION

The design of a Medium Access Control (MAC) algorithm for a satellite network aims, on the one hand, to propose a management mechanism for an efficient and flexible partitioning of the available uplink capacity and, on the other hand, to match the QoS requirements guaranteed to each connection in progress. Nevertheless, high propagation delays that affect the involved links and the variability and the heterogeneity of the incoming traffic introduce further difficulties.

Several MAC algorithms for DVB-RCS [1] networks have been analysed and compared [2][3][4][5], and, basing on the simulation results, the authors decided to develop the algorithm discussed in [5], which, even though holds some drawback, allows to manage the bandwidth allocation fully dinamically.

In particular, this algorithm, once implemented into a DVB-RCS standard compliant system, doesn't permits to strictly control, during the all session, the respect of the QoS requirements agreed in the set-up phase.

To achieve these performances, has been necessary to introduce heavy modifications to the reference algorithm in order to respect the QoS requirements defined in [6].

For the sake of simplicity, we will refer to the algorithm described in [5] as **Predictive Bandwidth Allocation** algorithm **(PBA)** and to the proposed one as **Dynamic Bandwidth Allocation with QoS** algorithm **(DBAQ)**.

Fig. 1 shows the scenario addressed in this work, it consist of: a DVB geostationary orbit satellite, a Hub Station (HS), several Satellite Terminals (STs) and a Network Control Centre (NCC), which performs the traffic control, and some User Terminal (UT), which generates the traffic offered to the satellite access networks and forwarded to the core network.

Manuscript received January 21, 2005. This work was supported in part by the IST SAILOR Project.

F. Delli Priscoli, G. Santoro, C. Colafranceschi, D. M. Krapez, E. Guainella are with the Department of Systems and Computer Science, University of Rome "La Sapienza", 00184 Italy (phone: +39-06-44585444; fax: +39-06-44585367; e-mail: dellipriscoli@dis.uniroma1.it, gsantoro@dis.uniroma1.it).

## II. SYSTEM ARCHITECTURE

Connections, relevant to every UT, may belong to different service classes ($k \in [1, N]$) and are grouped in Real Time (RT) and Non Real Time (NRT). The former are characterized by the maximum tolerated bit loss percentage ($L^{\%}_{\max-k}$), where $k$ indicates the class of service which the connection belongs, and the maximum tolerated delay ($D_k^{\max}$) that packets can experience inside the ST; the latter by the minimum rate guaranted to the connection ($R^{QoS}_{\min-k}$.)

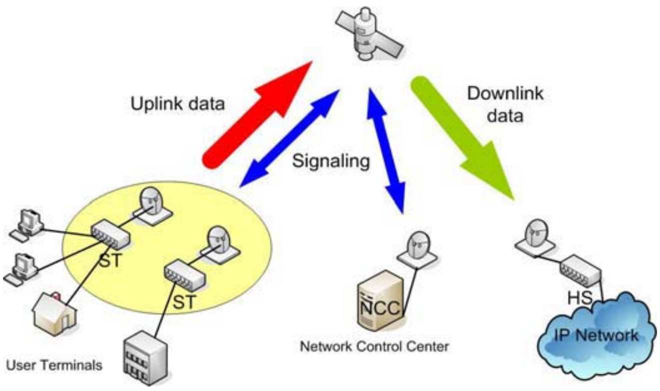


Fig. 1: Scenario overview

Each ST is provided with a set of N FIFO Buffers, one for each class of service, storing the IP datagrams incoming from the higher level, before being transmitted to the lower layers and to the uplink air interface. A Classifier, fed with the traffic arriving from the ST applications, is in charge of sorting the packets towards the N FIFO Buffers.

In the following, for the sake of brevity, the FIFO Buffer storing the packets relevant to a set of connections (i,j) and belonging to the *k-th* class of service, will be simply referred as queue (i,k), where *i* is the *i -th* ST.

The offered traffic to the *i-th* ST, once classified, will be handled with a per-flow policy, so that the bandwidth requests-assignments will be expressed in terms of flow/class of interest.

Let $\delta_{ass-ik}(h)$ denote the fraction of $R_{u-tot}$ (the overall uplink capacity) granted to the flow (i,k) at time *h*. Then, the parameter $\delta_{ass-ik}(h)$ is always included in the range [0,1] and the following fundamental uplink capacity constraint must be respected:

$$\sum_{i=1}^{S}\sum_{k=1}^{N}\delta_{ass-ik}(h) \leq 1 \qquad \textbf{(1)}$$

In the proposed scenario the NCC is in charge of computing and assigning to each *k-th* flow of each *i-th* ST the Dynamic Bit Rate, i.e. is in charge of selecting, at any time *h*, the parameters $\delta_{ass-ik}(h)$ (i = 1,...,S, k = 1,...,N).

The packets stored in the queue (i,k) can be either transmitted over the uplink air interface or, only for RT classes, discarded because they have waited more than the maximum tolerated delay ($D_k^{\max}$).

## III. CAPACITY ALLOCATION PROCEDURE

Before describing the assignment procedure, let us introduce the following definitions:

- Let $L$ denote the round-trip delay. Such a delay $L$, expressed in terms of $T_{short}$ (the shortest discrete interval considered) is equal to $2 \cdot \left\lceil D_{prop} / T_{short} \right\rceil$, where $D_{prop}$ is the maximum propagation delay in the run from any ST to the NCC, or in the opposite run; in the round-trip delay computation we neglect the NCC/ST computing times;
- Let $T_{\inf} = T_{sframe} / T_{short}$ denote the period occurring between two consecutive bandwidth requests, expressed in number of time intervals; in the simulator framework this value is set to 305*ms*;
- Let $R^*_{in-ik}(h; h+L+T_{\inf})$ represents the forecast, estimated by the *i-th* ST at time *h*, of the average traffic which will income into the queue (i,k) in the time interval $(h; h+L+T_{\inf})$. At a generic request time *h*, the STs compute, the average traffic incoming in the last $T_{window}$ seconds (the length of this time interval depends on incoming traffic period) and assume that in the following $T_{\inf} + L$ seconds the traffic holds this behaviour. Further forecasting methods could be also considered.

In the proposed demand-assignment mechanism the STs do not directly calculate the bandwidth they require. Conversely, they just send to the NCC some key parameters which are used by the NCC itself to perform appropriate bandwidth assignments.
So, each $T_{inf}$ seconds, the *i-th* ST sends to the NCC the following information:

1. the N predictions of the lengths of the queues (i,k) (k = 1,…,N) at time $h+L-T_{\inf}$ ;at time h the i-th ST computes these predictions, indicated as $q^*_{ik}(h+L-T_{\inf})$, according to the following expression:
$$q^*_{ik}(h+L-T_{\inf}) = q_{ik}(h) + R^*_{in-ik}(h; h+L-T_{\inf}) \cdot (L-T_{\inf}) + \\ -\delta_{ass-ik}(h+L-2\cdot T_{\inf}) \cdot (L-T_{\inf}) \cdot R_{u-tot}$$
2. the N predictions of the average bit rates of the traffic which will enter the queue (i,k) during the time interval $(h+L-T_{\inf}; h+L+T_{\inf})$, indicated as: $R^*_{in-ik}(h+L-T_{\inf}; h+L+T_{\inf})$;
3. the average delay ($D^{av}_{ik}$) encountered by packets of the queue (*i,k*) sent (or eventually discarded) in the last frame, over the time interval $(h-T_{frame}; h)$, before the current signalling transmission. Note that a superframe is divided into six frames of duration of 50ms

Basing on the information received from the STs, at time $h+(L/2)$, the NCC has to decide the capacity assignments $\delta_{ass-ik}(h+L)$ for any (i,k) pair (see Fig. 2).
These assignments must be performed aiming to fulfill the QoS requirements and to maximizing the bandwidth exploitation. Since these requirements are very different between RT and NRT services, we consider a two different policies of bandwidth assignments.
The NCC computes, during each time interval, the minimum bandwidth assignment in order to respect the QoS guarantees.
The exceeding bandwidth (if available) is shared among flows with a weight-based policy.
The request-assignment mechanism is shown in Fig.2.

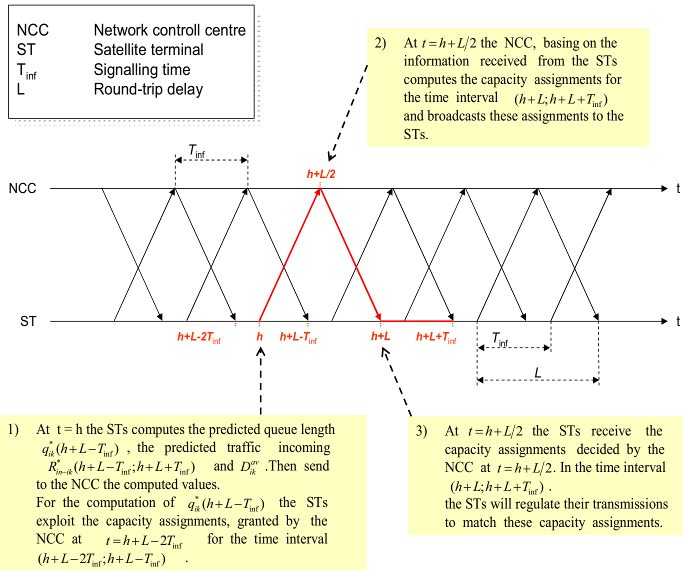


Fig. 2: Request assignment mechanism.

### A. *RT Assignments*

The lower boundaries for the RT classes assignments are computed by controlling the maximum queue length at time $h+L+T_{\inf}$.
Let $L^{\%}_{ik}$ denotes the bit loss percentage concerning this time instant. We define this parameter as the following ratio:

$$L^{\%}_{ik} = \frac{B^{loss}_{ik}}{B^{loss}_{ik} + B^{out}_{ik}} = \frac{B^{res}_{ik} - B^{\max}_k}{(B^{res}_{ik} - B^{\max}_k) + B^{out}_{ik}} \quad \textbf{(2)}$$

Where $B^{res}_{ik}$ denotes the residual number of bits that remain in the generic queue (*i,k*) at the instant $h+L+T_{\inf}$ after the bandwidth assignment received at time $h+L$.
$B^{\max}_k$ denotes the maximum residual length tolerated for the *k-th* queue, at the time instant $h+L+T_{\inf}$, to avoid losses. If $B^{res}_{ik} < B^{\max}_k$ means that $L^{\%}_{ik} = 0$.
$B^{out}_{ik}$ denotes the number of bits sent to the lower level during the time interval $(h+L-T_{\inf}; h+L+T_{\inf})$ and it's expressed by the following equation:

$$B^{out}_{ik} = \delta^{RT}_{ass-ik}(h+L-T_{\inf}) \cdot R_{u-tot} \cdot T_{\inf} + \delta^{RT}_{ass-ik}(h+L) \cdot R_{u-tot} \cdot T_{\inf} \quad \textbf{(3)}$$

$B^{\max}_k$, defined above, can be expressed as:

$$B^{\max}_k = D^{\max}_k \cdot R^{out}_{ik} \quad \textbf{(4)}$$

That is the maximum tolerated length for the *k-th* queue at the time instant $h+L+T_{\inf}$, so that with the assignment $\delta^{RT}_{ass-ik}(h+L+T_{\inf})$ the residue bits will be sent before a $D_k^{\max}$ time (this is a QoS requirement for RT services). At time $h+L/2$ the NCC is not aware of $\delta^{RT}_{ass-ik}(h+L+T_{\inf})$ so that we can approximate this assignment with the previous $\delta^{RT}_{ass-ik}(h+L)$ [1].

Therefore, $R_{ik}^{out}$ is the approximated output rate of the *k-th* queue after the assignment received at the instant $h+L+T_{\inf}$:

$$R_{ik}^{out} = \delta^{RT}_{ass-ik}(h+L)\cdot R_{u-tot} \tag{5}$$

Since $L_{ik}^{\%} \le L_{\max-k}^{\%}$ (see QoS requirements for RT services), from Eq. 3 follows:

$$B_{ik}^{res} \le B_{ik}^{\max} + \frac{L_{\max-k}^{\%}}{1-L_{\max-k}^{\%}} B_{ik}^{out} \tag{6}$$

By this constraint is possible to correlate the queues length with QoS requirements, setting the maximum tolerated sizes. Therefore, if $0 < B_{ik}^{res} \le B_{ik}^{\max} + \frac{L_{\max-k}^{\%}}{1-L_{\max-k}^{\%}} B_{ik}^{out}$, bit losses are lower than the QoS limit value $L_{\max-k}^{\%}$ and packets will be sent within a $D_k^{\max}$ time; whereas if occurs that $B_{ik}^{res} > B_{ik}^{\max} + \frac{L_{\max-k}^{\%}}{1-L_{\max-k}^{\%}} B_{ik}^{out}$, means that bit losses are over the upper threshold $L_{\max-k}^{\%}$.

From these considerations, replacing Eq. 3, Eq. 4, Eq. 5 into Eq. 6 and considering the lower boundary, the NCC performs the following assignment:

$$\delta^{RT}_{ass-ik}(h+L) = -\frac{\delta^{RT}_{ass-ik}(h+L-T_{\inf})\cdot T_{\inf}}{T_{\inf}+(1-L_{\max-k}^{\%})\cdot D_k^{\max}} + \\ +\frac{(1-L_{\max-k}^{\%})\cdot[R^*_{in-ik}(h+L-T_{\inf};h+L+T_{\inf})\cdot 2T_{\inf}+q^*_{ik}(h+L-T_{\inf})]}{R_{u-tot}\cdot[T_{\inf}+(1-L_{\max-k}^{\%})D_k^{\max}]} \tag{7}$$

*B. NRT Assignments*

NRT services must fulfill, instead, the constraint on the minimum output rate ($R^{QoS}_{\min-k}$) that must be guaranteed to the *k-th* class.

Since the predicted rate can be lower than the minimum rate, to respect the QoS constraint the NCC calculates the bandwidth assignment according to the following expression:

$$\delta^{NRT}_{ass-ik}(h+L) = \frac{\min\left\{R^{QoS}_{\min-k}; R^*_{in-ik}(h+L-T_{\inf};h+L+T_{\inf})\right\}}{R_{u-tot}} \tag{8}$$

*C. General Capacity Assignments*

Formulas expressed in Eq. 8 and Eq. 9, may be merged into a single expression as follows:

$$\delta^{\min}_{ass-ik}(h+L) = \\ \frac{(1-L_{\max-k}^{\%})\cdot[R^*_{in-ik}(h+L-T_{\inf};h+L+T_{\inf})\cdot 2T_{\inf}+q^*_{ik}(h+L-T_{\inf})]}{R_{u-tot}\cdot[T_{\inf}+(1-L_{\max-k}^{\%})D_k^{\max}]} + \\ -\frac{\delta_{ass-ik}(h+L-T_{\inf})\cdot T_{\inf}}{T_{\inf}+(1-L_{\max-k}^{\%})\cdot D_k^{\max}} + \frac{\min\left\{R^{QoS}_{\min-k}; R^*_{in-ik}(h+L-T_{\inf};h+L+T_{\inf})\right\}}{R_{u-tot}} \tag{9}$$

This formula, for RT services, can be brought back to Eq. 7, since minimum rates are not guaranteed ($R^{QoS}_{\min-RT}=0$) and, for NRT services, to Eq. 8, since $D^{\max}_{NRT} \longrightarrow \infty$.

The sum of the previous calculated bandwidth assignments couldn't respect the constraint expressed in Eq. 1, in case of $\sum_{i=1}^{S}\sum_{k=1}^{N}\delta^{\min}_{ass-ik} > 1$ is applied a normalization, so that the new assignments become:

$$\delta'_{ass-ik}(h+L) = \frac{w_{ik}(h-T_{frame};h)\cdot\delta^{\min}_{ass-ik}(h+L)}{\sum_{i=1}^{S}\sum_{k=1}^{N} w_{ik}(h-T_{frame};h)\cdot\delta^{\min}_{ass-ik}(h+L)} \tag{10}$$

where

$$w_{ik}(h-T_{frame};h) = \frac{D_{ik}^{av}(h-T_{frame};h)/D_k^{\max}}{\sum_{i=1}^{S}\sum_{k=1}^{N} D_{ik}^{av}(h-T_{frame};h)/D_k^{\max}} \tag{11}$$

are weights used to attribute different priorities to the queues according to delays experimented by the queued packets.

In case of $\sum_{i=1}^{S}\sum_{k=1}^{N}\delta^{\min}_{ass-ik} < 1$, the constraint in Eq. 1 is fulfilled, but aiming to exploit the overall bandwidth, the residual capacity, indicated by $RC = 1-\sum_{i=1}^{S}\sum_{k=1}^{N}\delta^{\min}_{ass-ik}(h+L)$ is shared basing on the weights defined in Eq. 11 and added to $\delta^{\min}_{ass-ik}$, as follows:

$$\delta'_{ass-ik}(h+L) = \delta^{\min}_{ass-ik}(h+L) + w_{ik}(h-T_{frame};h)\cdot RC$$

This equation permits the fully exploitation of the uplink capacity and the respect of Eq. 1.

As a further remark, note that the *i-th* ST can rearrange the received capacity assignments relevant to the flows (i,k) (k = 1,..., N) to take into account its actual situation. So, at a time $h+L$, the *i-th* ST computes the overall uplink capacity $\alpha_i(h)$ assigned to it, as:

$$\alpha_i(h+L) = \sum_{k=1}^{N}\delta'_{ass-ik}(h+L) \tag{12}$$

Then, the *i-th* ST allots its overall uplink capacity $\alpha_i(h)$ to the N flows according to criteria which can take into account updated information concerning effective queues length (information which was not available when the NCC computes the parameters $\delta'_{ass-ik}(h+L)$). So, these criteria can cause capacity assignments to the flows, hereafter indicated as $\delta_{ik}(h+L)$, which can differ from the ones granted by the NCC (i.e. $\delta'_{ass-ik}(h+L)$). Obviously, whatever criterion is adopted, the following constraint must be respected:

$$\sum_{k=1}^{N}\delta_{ik}(h+L) \le \alpha_i(h+L) \tag{13}$$

[1] This assumption can be done because the Dual Leaky Buckets, used in the implemented architecture, regularize the incoming traffic.

## IV. SIMULATION RESULTS

The simulation tool OPNET Modeler 9.0 has been adopted for modelling and testing performances of the proposed algorithm.

The MF-TDMA access is DVB-RCS compliant and the timing structure is divided into superframes ($T_{\inf}$) which duration is 305$ms$ and each superframe is divided into six frames ($T_{frame}$) which durations are 50 $ms$ (see [2]). Moreover, signalling is periodically sent every $T_{\inf}$ time.

The statistical characteristics of the considered applications are reported in Table 1. In particular, we considered four types of applications: two Real Time applications, namely Voice Over IP and Video Conference, and two Non Real Time applications, namely FTP and Web Browsing.

TAB. I
TRAFFIC SOURCES PARAMETERS

| | Video Conf. | VoIP | FTP | WEB |
|---|---|---|---|---|
| **Packet size (bits)** | uniform int (320,11680) | constant(580) | uniform int. (320,39680) | uniform int. (320,23680) |
| **Packet interarrival time (secs)** | constant (0.05) | constant (0.02) | uniform (0.05;0.15) | uniform (0.05;0.55) |
| **Average Rate (kbit/s)** | 120 | 29 | 200 | 40 |

The simulated scenario includes a spot beam with 5 STs (similar to the one described in Fig. 1), each connected to an Ethernet LAN (UT). Each of such UTs may be involved in a generic number of connections relevant to the applications described before. We considered, in this simulations set, three different conditions of average uplink load, at 75%, 90% and 105% of the overall uplink capacity. This capacity, for a single frequency bearer, is equal to 2,895,738 bit/s.

A 2-layer switching GEO satellite (round trip propagation delay $L = 500ms$), a Network Control Centre and a Hub Station in charge of connecting the UTs with servers.

The quality of service requirements that must be assured to each connection are shown in Table 2.

The first set of simulations, compares three different reassignment policies performed by STs over the time interval $[h+L; h+L+T_{\inf}]$, when the overall capacity is dynamically partitioned.

The three considered possibilities are:

1. *no bandwidth reassignment*, where STs use in such time interval, for every time frame, the same assignment received at the instant $h+L$ from the NCC.
2. *reassignment per superframe*, where STs use in such time interval, for every time frame, the same bandwidth reassignment calculated at the instant $h+L$ by the STs themselves.
3. *reassignment per frame,* where STs use, frame by frame, updated bandwidth reassignments themselves performed at the time instants $h+L+kT_{frame}$ ($k$=0,…,5).

In Fig. 4 are shown the performances in terms of bit loss percentages for Real Time classes, varying the bandwidth reassignment mode.

We observe that decreasing the average load, bit losses decrease also. As concerns the Video Conferencing class, we note that the QoS requirements are satisfied for each bandwidth reassignment policy and each average load. In every case, bit loss are lower than 0.25%, less than the maximum tolerated for this class of service (5%), so that this service class shouldn't need a local reassignment to be compliant to the QoS requirements. However, ranging from the absence of reassignment to the reassignment per frame, performances improve considerably. An important result is that when a reassignment per frame is considered, no bit loss are found.

TAB. II
QoS REQUIREMENTS

| Service Class | | Maximum queuing delay | Tolerated loss (IP datagrams) | Minimum rate[2] (IP datagrams) |
|---|---|---|---|---|
| **RT** | **Video Conference** | 500 ms | 2% | Not Applicable |
| | **VoIP** | 150 ms | 1% | Not Applicable |
| **NRT** | **FTP** | Not Applicable | Not Applicable | 92 Kbit/s |
| | **WEB Browsing** | Not Applicable | Not Applicable | 16 Kbit/s |

For as concerns the VoIP class, since the maximum tolerated bit loss is 2%, the reassignment per frame policy is the only that allows to respect the QoS contract. With this last policy, bit losses are very limited, but are not null, as for the Video case (the VoIP packets have stricter QoS constraints than the Video ones).

A second simulation set has been realized with the scope to demonstrate the obtained improvements implementing the **DBAQ algorithm** respect to the literature algorithm considered (**PBA algorithm**). The policy of reassignment used for both algorithms is per frame and performances are compared varying the offered traffic load.

Fig. 5 highlights that the DBAQ algorithm, differently to the PBA algorithm, is QoS respectful in terms of bit loss percentages, exploiting more efficiently the uplink interface (see Fig. 3). These improvements are due to a flexible bandwidth management, since NRT services have a decreasing assigned capacity, when the average load increases. Such decreasing of capacity entails an increasing of the assigned capacity to RT classes (Fig. 7). However, as regards NRT classes, QoS constraints are always fulfilled.

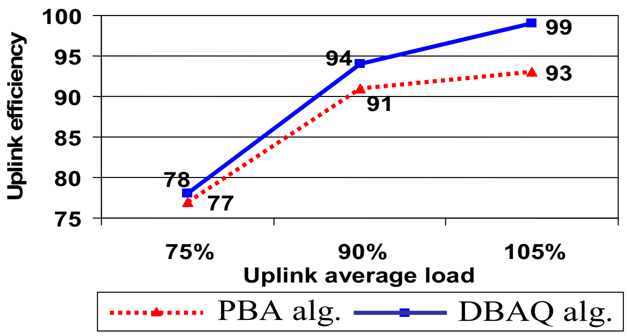


Fig. 3: Uplink efficiency.

[2] Guaranteed rates are considered as the 40% of the NRT average offered traffic rates.

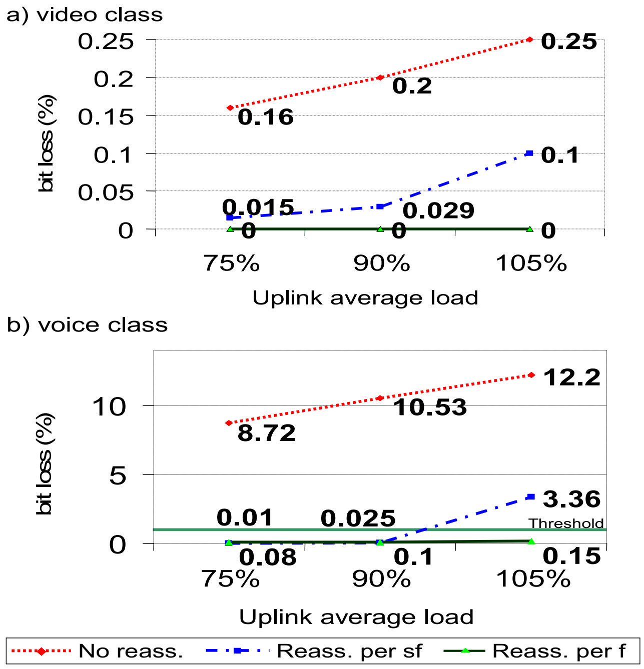


Fig. 4: Bit loss percentage, varying the reassignment policy: a) Video conferencing; b) VoIP service.

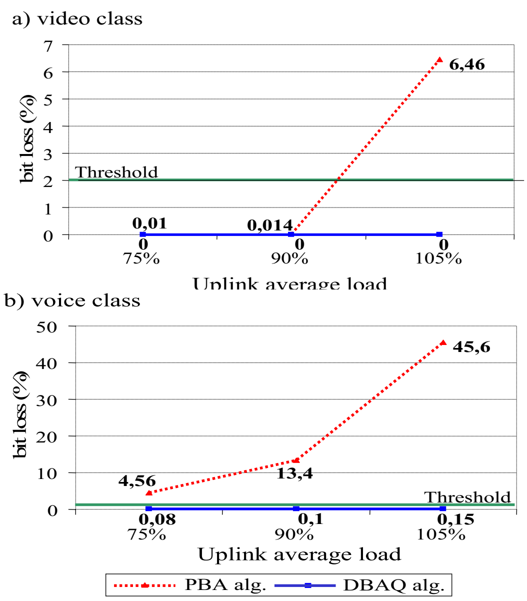


Fig. 5: Bit loss percentage. a) Video conferencing b) VoIP service.

## V. CONCLUSIONS

The simulation results show that the authors have developed an algorithm that efficiently performs capacity allocations respecting the agreed QoS requirements. The algorithm copes with high propagation delay typical of satellite networks and a fully dynamic capacity management necessary for an efficient link exploitation.

Further improvements to the proposed method can be investigated, introducing a more reliable predictive criterion, such as analysing the self-similarity of the Internet traffic sources, in order to reduce the errors made during queues size computations.

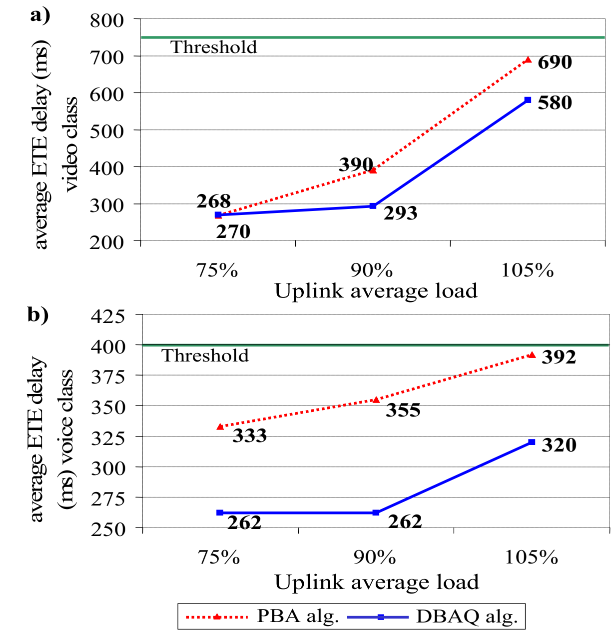


Fig. 6: Average End to End delays. a) Video conferencing b) VoIP service.

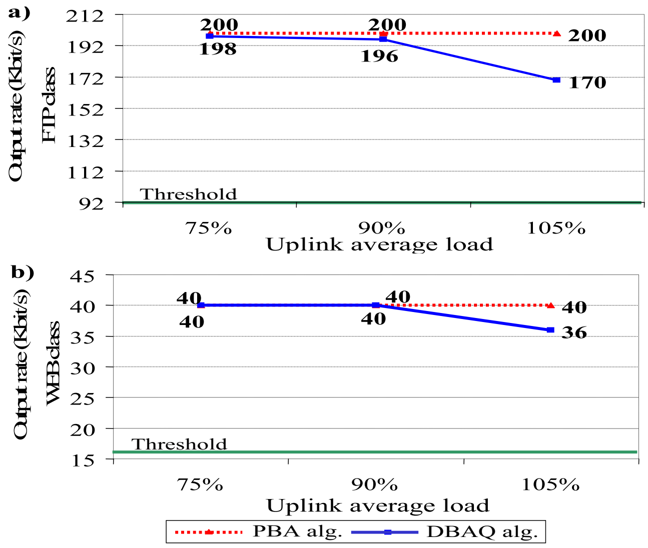


Fig. 7: Output rates. a) FTP service b) WEB Browsing.

## REFERENCES


[1] "*Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems*", ETSI EN 301 790 v1.2.2, December 2000.

[2] E. Impemba, T. Inzerilli, "*QoS Provisioning Support in DVB/RCS based Satellite Access Terminals*", 4th Intern. Network Conference, July 2004.

[3] E. Impemba, T. Inzerilli, "*QoS Support for Interactive Communication with DVB/RCS Satellites*", The Ninth IEEE Symposium on Computers and Communications, 29 June-1 July 2004.

[4] T. Inzerilli, E. Paone, "*Access Control with QoS Provision in the SATIP6 Project*", 13th IST Mobile and Wireless Communic., June 2004.

[5] F. Delli Priscoli, D. Pompili, G. Santoro, "*A QoS-Aware Bandwidth on Demand Assignment mechanism in a GEO Satellite System*", IST Mobile Summit, Lyon, July 2004.

[6] "*TIPHON: Definition of QoS classes*", ETSI TS 101 329-2, July 2000.

[7] Li Hwang, "*The convergence of traffic measurement and queuing analysis*", 1997.

[8] Blake S. et al. "*An Architecture for Differentiated Services*", RFC 2475, December 1998.

[9] V. Firoiu, J. Kurose and D. Towsley, "*Efficient Admission Control of Piecewise Linear Traffic Envelopers at EDF Schedulers*", January 1998.