This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

Adaptive Call Admission and Bandwidth Control in DVB-RCS Systems

Mario Marchese, Maurizio Mongelli
DIST – Department of Communications, Computer and Systems Science
CNIT – Italian National Consortium for Telecommunications
University of Genoa, Via Opera Pia 13, 16145, Genoa, Italy
{ mario.marchese, maurizio.mongelli }@ unige.it

Abstract—**The paper studies a protocol and a control architecture aimed at implementing bandwidth and call admission optimization over a DVB Return Channel Satellite Terminal (RCST) under Quality of Service (QoS) contraints. The approach can be applied in all cases where traffic flows, coming from a terrestrial portion of the network, are merged together within a single DVB flow, which is then forwarded over the satellite channel. The exact bandwidth to be provided to the flows entering the DVB layer is unknown because equivalent bandwidth techniques guarantee only approximate solutions in heterogeneous conditions. The paper firstly outlines merging operations of the flows over the DVB system and then it specifies the entities to be used at layer 3 and layer 2 of the RCST to coordinate the necessary actions to guarantee the QoS with minimum resource allocation.**

Keywords-DVB-RCS, call admission, measurement-based equivalent bandwidth, QoS mapping

I. INTRODUCTION

This paper deals with a satellite network, based on the DVB-RCS standard [1, 2], composed of a GEO stationary bent-pipe satellite, Return Channel Satellite Terminals (RCSTs), and a Network Control Center (NCC) that is connected to the Internet (Fig. 1). Local Area Networks (LANs) may be connected to RCSTs. RCSTs are fixed and use the Return Channel via Satellite (RCS) to transmit data and signaling. The NCC provides control and monitoring functions and manages network resources allocation to RCSTs. DVB-S is used for the forward link (from NCC to RCSTs) and DVB-RCS is employed for the return link (from RCSTs to NCC). DVB-RCS envisages 5 DVB classes, each implemented through a dedicated queue at DVB layer, and 5 corresponding resource allocation types whose requirements affect the overall available bandwidth that can be dynamically allocated.

The problem of bandwidth allocation among the RCSTs to satisfy QoS levels naturally arises in DVB environments, especially for the return channel where bandwidth is a scarce resource. This, in turn, leads to the problem of QoS mapping of IP flows over the DVB classes (see [3], [4] and references therein). This topic received the attention of the satellite community in the last years for what concerns research projects (e.g., [5], [6]) and scientific literature (an excellent overview can be found in [7]). The nature of the problem recalls the principles of "adaptive feedback control", as DVB classes dynamically ask and release bandwidth resources on the basis of measures of the traffic flows. The overall aim of dynamic bandwidth control is to avoid wasting resources. Several works address the dynamic bandwidth control problem in DVB networks (see [8] and references therein) by exploiting optimal control methodologies, e.g., [9], and TCP adaptations to the satellite, e.g., [10]. A peculiarity of the DVB technology, not addressed by the mentioned literature, is that bandwidth allocation must be implemented when IP flows with different statistics and QoS constraints are merged into a smaller set of DVB streams at layer 2 [4, 7]. This action is called vertical QoS mapping.

The underlying idea of this paper is to design control blocks and actions of a combined Call Admission Control (CAC) – Bandwidth Allocation scheme, implemented within the RCST Layer 2 Resource Manager (L2RM). The idea of driving the CAC with measurement-based equivalent bandwidth is derived from [11, 12]; the computation is extended to heterogeneous conditions in [13] and to the presence of noisy wireless channels [4].

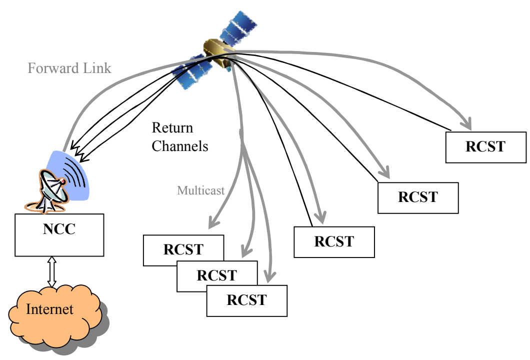

Figure 1. DVB-RCS-S2 communication system.

II. ARCHITECTURE OF RCST LAYER 2 RESOURCE MANAGER (L2RM)

Fig. 2 shows L2 data and control planes of the RCST. Given N_{L2} traffic flows at layer 2, the data plane is modeled as a set of traffic buffers, one for each flow, $\left(f_1^{L2},...,f_{N_{L2}}^{L2}\right)$, served with rate $\left(\theta_1,...,\theta_{N_{L2}}\right)$, respectively. The layer 2 receives the traffic flows from the layer 3, which is not the object of this paper.

A proposal concerning formal communication between layer 3 and layer 2 is reported in [3, 14]: the IP protocol stack is divided into lower layers (layer 2 and 1), called Satellite Dependent (SD) layers, and upper layers (layer 3-IP and above layers), called Satellite Independent (SI) layers. The interface between SI and SD layers (in practice the interface between IP and layer 2) is defined through an interface called SI-SAP (Satellite Independent – Service Access Point), which provides a set of communication primitives for IP - layer 2 communication and assures the separation of SI, independent of satellite technology, and SD layers, strictly dependent on the used satellite technology.

The control plane contains the L2RM, which is composed of N_{L2} Decision Makers (DMs), whose role, specified in detail in the following, is to compute an estimation $\left(\theta_1^{\min},...,\theta_{N_{L2}}^{\min}\right)$ of the minimum bandwidth necessary at traffic buffers to provide a given quality of service $\left(QoS_1,...,QoS_{N_{L2}}\right)$. After computing the minimum bandwidths, the vector $\left(\theta_1^{\min},...,\theta_{N_{L2}}^{\min}\right)$ is forwarded to the Bandwidth Allocator block that decides and communicates to traffic buffers real bandwidth allocations $\left(\theta_1,...,\theta_{N_{L2}}\right)$. Real bandwidth allocations should depend on the values of the minimum bandwidth. Mathematically: $\theta_1 = f\left(\theta_1^{\min}\right),...,\theta_{N_{L2}} = f\left(\theta_{N_{L2}}^{\min}\right)$. A proposal concerning function $f(\cdot)$ is reported in the following.

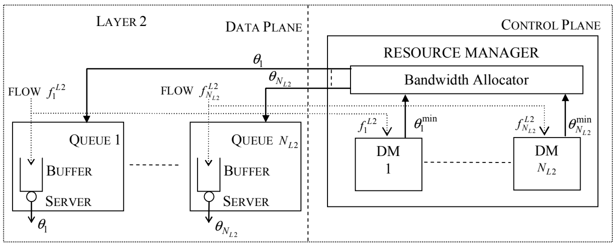

Figure 2. RCST Layer 2.

III. DECISION MAKERS: VIRTUAL QUEUES AND MINIMUM BANDWIDTH COMPUTATION

The architecture of generic $i-th$ DM is shown in Fig. 3. Each traffic flow $\left(f_1^{L2},...,f_{N_{L2}}^{L2}\right)$ is divided into two parts. One is directed towards the corresponding traffic buffer to be forwarded to the physical layer as shown in Fig. 2. The other part is sent to another buffer, the "virtual queue", which is an exact mirror of the traffic one but it does not interfere with routinary forwarding operations. Traffic buffers and virtual queues work in parallel but virtual queues are served with the estimated minimum rates $\left(\theta_1^{\min},...,\theta_{N_{L2}}^{\min}\right)$, computer by the Estimated Minimum Bandwidth Computation block in each DM.

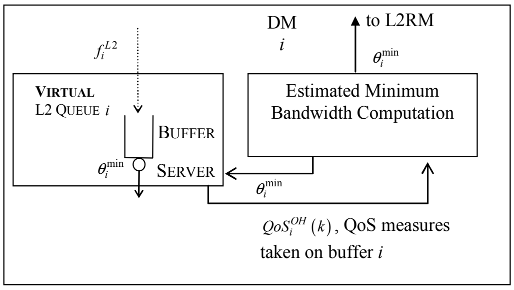

Figure 3. DMi architecture.

DM i computes $\theta_i^{\min}$ by following a set of steps. A sequence k=1,2,… of observation horizons for DMi is defined ($OH_i(k)$), during which the virtual queue is monitored in appropriate instants of time. These instants and monitoring information compose an information vector $I_i(k)$ for each $OH_i(k)$. $I_i(k)$ drives the service rate computation of queue i at time k+1 together with the previous allocations up until a time depth d , thus generating $\theta_i^{\min}(k+1)$.

$$\theta_i^{\min}(k+1)=F\left(\theta_i^{\min}(k),\theta_i^{\min}(k-1),...,\theta_i^{\min}(k-d),I_i(k),...,I_i(k-d)\right) \quad (1)$$

An example of information vector may be represented by the vector of differences $\left(e_1,...,e_{N_{L2}}\right)$ between the QoS levels measured over the virtual queues during the observation horizons $OH_i(k)$ and denoted by $QoS_i^{OH}(k)$, and the QoS threshold values QoS_i : $e_i(\cdot,k)=\left(QoS_i - QoS_i^{OH}(k)\right)^2$. A possible simplification of (1) is contained in (2) where $\theta_i^{\min}(k+1)$ is computed by using information at instant k .

$$\theta_i^{\min}(k+1) = F\left(\theta_i^{\min}(k),e_i(\cdot,k)\right) \quad (2)$$

Fig. 4 shows the steps to get $\theta_i^{\min}(k+1)$ taking (2) as a reference. Examples of algorithms to compute $\theta_i^{\min}$ are reported later. After getting $\theta_i^{\min}$ each DMi implements a scheme to evaluate the stabilization of $\theta_i^{\min}$ computation. The steady state of $\theta_i^{\min}, \forall i \in [1,...,N_{L2}]$ is captured under the following condition: $\left|\theta_i^{\min}(k+1)-\theta_i^{\min}(k)\right| \le \varepsilon_i$, where ε_i is the stabilization threshold for queue i. A reasonable practical value of ε_i might be $\varepsilon_i = \frac{1}{10}\cdot\theta_i^{\min}(k)$.

IV. BANDWIDTH ALLOCATION AT L2RM AND CAC

A. L2RM Bandwidth Allocator

The real bandwidth θ_i allocated to queue i is changed over time by following indications coming from DMi and depending on the value of $\theta_i^{\min}$. This paper proposes a possible action but other solutions may be applied. L2RM, on the basis of the value of $\theta_i^{\min}$, assumes three possible states for queue i: a) "no action required"; b) "imminent congestion"; 3) "bandwidth release". Ideally the 3 states can be defined by using only the value of $\theta_i^{\min}$:

$$\boxed{\begin{array}{l} \text{if } \theta_i < \theta_i^{\min} \Rightarrow \text{not enough bandwidth - imminent congestion} \\ \text{if } \theta_i = \theta_i^{\min} \Rightarrow \text{minimum bandwidth - no action required} \\ \text{if } \theta_i > \theta_i^{\min} \Rightarrow \text{too much bandwidth - bandwidth release} \end{array}} \quad (3)$$

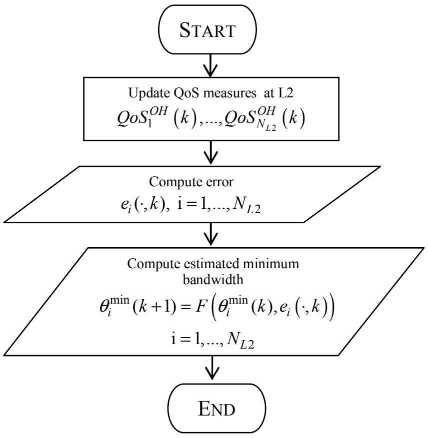

Figure 4. Estimated bandwidth computation by DM*i*.

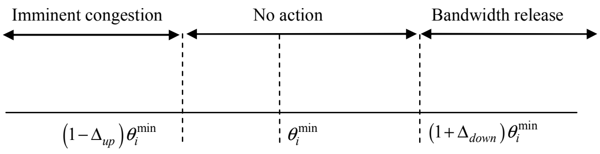

Figure 5. Queue i possible states.

Operatively the proposal in (3) is too hard and can generated unexpected bandwidth oscillations with consequent impact on real traffic. A smoother solution is preferred in practice: the mentioned states are defined by 2 thresholds $(1-\Delta_{up})$ and $(1+\Delta_{down})$ as follows:

$$\boxed{\begin{array}{l} \text{if } \theta_i < (1-\Delta_{up})\cdot\theta_i^{\min} \text{ not enough bandwidth - congestion} \\ \text{if } (1-\Delta_{up})\cdot\theta_i^{\min} \le \theta_i \le (1+\Delta_{down})\cdot\theta_i^{\min} \text{ minimum bandwidth - OK} \\ \text{if } \theta_i > (1+\Delta_{down})\cdot\theta_i^{\min} \text{ too much bandwidth - bandwidth release} \end{array}} \quad (4)$$

Fig. 5 graphically shows the concepts introduced in (4): if the allocated bandwidth θ_i is "close enough" to $\theta_i^{\min}$, no action is required; if θ_i value is too much below $\theta_i^{\min}$, bandwidth is underprovisioned and congestion may happen in the next future; if θ_i value is too much above $\theta_i^{\min}$, bandwidth is overprovisioned and may be released.

B. *Call Admission Control (CAC)*

The bandwidth update provided by the L2RM Bandwidth Allocator described above is limited by the maximum available bandwidth C at a specific RCST. This value is provided by the NCC, shown in Fig. 1. In this case, referring to the specific RCST in Fig. 2, which implements N_{L2} queues, there is the following constraint: $\sum_{i=1}^{N_{L2}} \theta_i \le C$. By referring again to the specific RCST in Fig. 2, a maximum bandwidth C_i might be forecast for queue i. The constraint is $\sum_{i=1}^{N_{L2}} C_i = C$. CAC is implemented within L2RM and must consider both the currently used bandwidth θ_i and the maximum available bandwidth, either C or C_i, as explained above. An incoming connection with peak bandwidth p is accepted at queue i if $\sum_{i=1}^{N_{L2}} \theta_i + p \le C$ or $\sum_{i=1}^{N_{L2}} \theta_i + p \le C_i$.

V. ALGORITHMS TO COMPUTE THE MINIMUM BANDWIDTH $\theta_i^{\min}$

Different forms of the control law $F(\cdot)$, appearing in formula (1) and (2), may be reasonably applied in the context of the paper to get $\theta_i^{\min}$.

A. *Reference Chaser Bandwidth Controller (RCBC)*

If the QoS of interest is the Packet Loss Probability (PLP) or the delay one could use the Infinitesimal Perturbation Analysis framework to derive a gradient-based formulation of the control law $F(\cdot)$ as follows:

$$\theta_i^{\min}(k+1) = \theta_i^{\min}(k) + \eta_k \left.\frac{\partial e_i(\cdot,k)}{\theta_i}\right|_{\theta_i = \theta_i^{\min}(k)} \quad (5)$$

where η_k is the gradient step size; more specifically, for the PLP case:

$$\left.\frac{\partial e_i(\cdot,k)}{\theta_i}\right|_{\theta_i = \theta_i^{\min}(k)} = 2\cdot\left.\frac{\partial \hat{l}_i(\theta_i)}{\partial\theta_i}\right|_{\theta_i = \theta_i^{\min}(k)} \cdot\left[\hat{l}_i^*(\cdot,k) - \hat{l}_i(\cdot,k)\right] \quad (6)$$

where:

$$\left.\frac{\partial \hat{l}_i(\theta_i)}{\partial\theta_i}\right|_{\theta_i = \theta_i^{\min}(k)} = -\frac{1}{T_k}\sum_{bp=1}^{N_{T_k}}\left[at_{T_k}^{bp}(\theta_i^{\min}(k)) - ll_{T_k}^{bp}(\theta_i^{\min}(k))\right] \quad (7)$$

where $\hat{l}(\cdot,k)$ is the measured loss rate of queue i over the $OH_i(k)$, $\hat{l}^*(\cdot,k)$ is the target loss rate coming from the required PLP value for queue i (PLP_i^*): $\hat{l}^*(\cdot,k) = \int_{OH_i(k)} PLP_i^* \cdot a_i(t)\,dt$ ($a_i(t)$ is the measured input rate of traffic class i over the $OH_i(k)$), T_k is the size of $OH_i(k)$. A busy period (*bp*), in (7), is a period of time in which the buffer is not empty. The quantity in brackets in (7) is the difference between the last loss during the busy period *bp* and the starting time of *bp*.

B. PID control

Other, more traditional, approaches for the control law $F(\cdot)$ are possible. For example, Proportional Integrative Derivative (PID) control laws may be applied as follows:

$$\begin{aligned} \theta_i^{\min}(k+1) &= \theta_i^{\min}(k) + w_{k+1}(k+1)\cdot e_i(\cdot,k+1) + \\ & w_k(k)\cdot e_i(\cdot,k) + w_{k-1}(k)\cdot e_i(\cdot,k-1) \end{aligned} \tag{8}$$

where the weights are the tuning parameters used to optimize the PID temporal behavior in dependence of the specific application of interest.

C. Equivalent bandwidth

Other control laws, not directly dependent on error $e_i(\cdot,k)$, are applicable. An example is:

$$\theta_i^{\min}(k+1) = m_i(k) + d\cdot\sigma_i(k),\ d = \sqrt{-2\ln(PLP_i^*) - \ln(2\pi)} \tag{9}$$

The equation reports an Equivalent Bandwidth (EqB) technique applicable for the PLP case [11, 12]. $m_i(k)$ and $\sigma_i(k)$ are the average and standard deviation, respectively, of the input rate process of queue i over $OH_i(k)$ and PLP_i^* is the PLP requirement at queue i.

VI. PERFORMANCE EVALUATION AND DISCUSSION

The bandwidth control algorithms under investigation are: RCBC, PID and EqB. PLP metric is considered here, thus disregarding the delay metric. VoIP traffic is taken as reference (later, heterogenesou flows are applied). According to ITU-T P.59, each source is an on-off process with exponentially distributed on and off time durations (mean 1.008 s and 1.587 s, respectively) and peak bandwidth of 16 kbps. VoIP traffic enters an IP buffer whose length and service rate (set by the traffic peak bandwidth) guarantee no packet loss rate. IP traffic is encapsulated in DVB, thus generating the process f_i^{L2}; f_i^{L2} enters the DVB buffer, where the VoIP loss rate in VoIP packets (of 80 bytes each) is measured every OH. PLP_{VoIP}^* is set to $1\cdot 10^{-2}$.

The PID weights are set via brute force analysis in order to guarantee the best PID performance. The application of CAC is considered together with heterogeneous aggregation and channel fading. The maximum available rate of the L2 queue is 2.0 Mbps; its buffer size amounts of 150 DVB cells. Together with the VoIP traffic outlined above, video sources are used; each video source is modelled as an exponentially modulated on-off process, with mean rate 0.25 Mbps and peak rate 0.7 Mbps and an average burst of 10.0 s. The active time of VoIP and video sources are log-normally distributed with an average of 200 s. the activation processes are Poisson arrival processes. Following [11], varying load conditions ranging from 0.5 to 5 are considered, where the value of 1.0 corresponds to a mean source activation rate equal to 240 sources/hour (4 calls per minute); an incoming call is randomly set to be VoIP or video with identical probabilities. Following [15], the performance metrics under investigation are the achieved PLP and link utilization (conceptually, the opposite of call blocking probability) under a given load. The delay metric is disregarded for the sake of simplicity; the maximum achievable delay with a full L2 buffer (of 150 DVB cells with rate of 2.0 Mbps) is however 112 ms, which is a rare event with a PLP lower than 1%. The target PLPs are 1% for VoIP and 0.1% for video.

The employed fading process has been taken from [15], where real attenuation samples are extracted from an experimental data set. Six different fading classes are defined, corresponding to combinations of channel bit and coding rate that give rise to redundancy factors. With the data adopted in [15], whose trace is depicted in Fig. 6, the fading lies in:

$$\phi(t) \in \{0.0,\ 0.15625,\ 0.3125,\ 0.625,\ 0.8333,\ 1.0\} \tag{10}$$

The bandwidth reduction at the L2 queue, denoted by $\theta_i(t)$, can be computed as $\theta_i(t) = \phi(t)\cdot\theta_i^{Max}(t)$. As only the rate $\theta_i(t)$ is available for data traffic, the $\theta_i^{Max}(t)$ has to tuned over time in order to maintain the required QoS. As RCBC and PID set automatically the bandwidth under the measured losses, they do not need to know the exact value of $\phi(t)$. On the other hand, EqB follows the bandwidth computation rule:

$$\begin{aligned} \theta_i^j(k+1) &= \left(2-\phi(k+1)\right)\cdot m_i(k) + \varsigma\cdot\sigma_i(k), \\ \varsigma &= \sqrt{-2\ln(PLP_{EqB}^*) - \ln(2\pi)} \end{aligned} \tag{11}$$

The $\left(2-\phi(k+1)\right)$ quantity corrects the bandwidth computation in dependence of the current information rate dedicated to protection codes (e.g., when $\phi(k+1) = 0.2$, the multiplying factor of $m_i(k)$ is 1.8). As PLP_{EqB}^* is the most stringent PLP requirement, EqB is not able to tune the bandwidth in proportion to the ongoing VoIP or video calls.

Figs 7 and 8 show the achieved PLP for each service under different loads; Fig. 9 deals with the inherent overall utilization of the available rate. In overall, RCBC reveals to be more precise in tuning the bandwidth; the PLP curves are almost flat and safely stay a little below the corresponding targets. This has clearly an impact on CAC, thus obtaining the highest levels of utilization over all the loads. The PLP performance of PID and EqB is acceptable as well, but fails to stay below the targets with the highest loads.

The maximum achievable utilization is 80% (surprisingly, obtained by all the techniques) under a load of 5. It is finally worth noting that the EqB has been applied without any optimization of the OH size. Even if the concept of "dominant time scale" could be applied to EqB [11], together with other remarkable tuning schemes (such as the concept of "reference source", as in [11]), the inherent overprovisioning under the PLP_{EqB}^* and the heuristic tuning with $\phi(\cdot)$ in (11) make the EqB hardly applicable in this context.

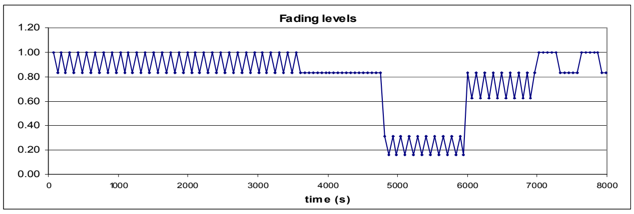

Figure 6. Fading levels [15].

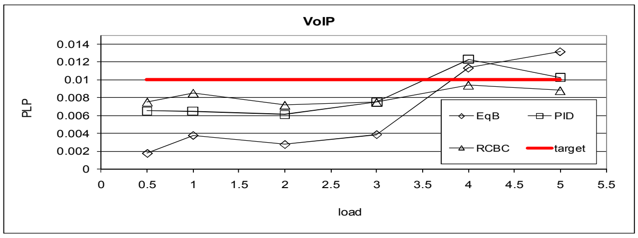

Figure 7. VoIP PLP.

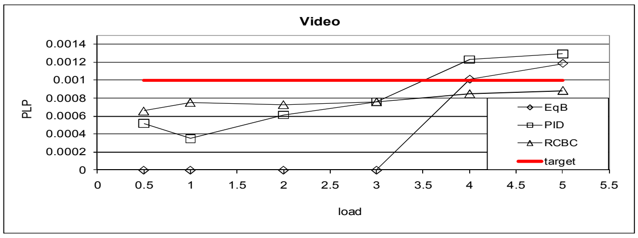

Figure 8. Video PLP.

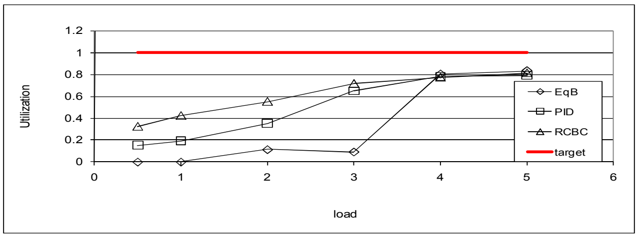

Figure 9. Utilization.

VII. Conclusion and Future Work

The paper proposes a new approach for bandwidth allocation and Call Admission Control over the DVB Return Channel of a satellite system. It is suitable for the optimization of the bandwidth under heterogeneous aggregation and Quality of Service constraints. The performance evaluation shows promising performance, thus opening the door to future evolutions as regards: delay and jitter control and the implementation of the control scheme inside a real architecture.

References

[1] ETSI TR 101 790. Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790. Technical Report. V1.2.1, (2003).

[2] ETSI EN 301 790. Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems. V1.5.1, (2009).

[3] ETSI TS 102 462. Satellite Earth Stations and Systems (SES), Broadband Satellite Multimedia (BSM); Broadband Satellite Multimedia Services and Architectures: QoS Functional Architecture. Technical Specification, V0.4.2, (2006).

[4] M. Marchese, M. Mongelli, “On-line Bandwidth Control for Quality of Service Mapping over Satellite Independent Service Access Points,” Computer Networks, vol. 50, no. 12, Aug. 2006, pp. 2089-2111.

[5] Applications layer QoS in DVB-RCS systems, Research Project, European Space Agency, http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=27802 , Oct. 2007.

[6] European Satellite Communications Network of Excellence (SatNEx) I&II, http://www.satnex.org/.

[7] H. Skinnemoen, H., A. Vermesan, A. Luoras, G. Adams, X. Lobao, “VoIP over DVB-RCS with QoS and Bandwidth on Demand,” IEEE Wireless Commun. Mag., vol. 12, Oct. 2005, pp. 46-53.

[8] Adaptive Resource Management and Optimization in Satellite Networks, G. Giambene, Ed., Springer Verlag, Berlin, Dec. 2006.

[9] S. Bracciali, R. Fantacci, T. Pecorella, L. Chisci, M. A. Vazquez Castro, “Proactive vs. Reactive DVB-RCS Terminal Using ACM Techniques,” Proc. IEEE International Conference on Communications 2008 (ICC 2008), Beijing, China, May 2008, pp. 1881-1885.

[10] M. Luglio, C. Roseti, F. Zampognaro, “Improving Performance of TCP-Based Applications over DVB-RCS Links,” Proc. IEEE International Conference on Communications 2009 (ICC 2009), Dresden, 14-18 June 2009, pp. 1-6.

[11] S. Georgoulas, P. Trimintzios, G. Pavlou, K. Ho, “Measurement-based Admission Control for Real-time Traffic in IP Differentiated Services Networks”, Proceedings of IEEE International Conference on Telecommunications (ICT 2005), Capetown, South Africa, May 2005, pp. 1-5.

[12] S. Georgoulas, P. Trimintzios, G. Pavlou, K. Ho, “An Integrated Bandwidth Allocation and Admission Control Framework for the Support of Heterogeneous Real-time Traffic in Class-based IP Networks,” Computer Communications, vol. 31, no. 1, Elsevier, Jan. 2008, pp. 129-152.

[13] M. Marchese, M. Mongelli, “Measurement-based Computation of Generalized Equivalent Bandwidth for Loss Constraints,” IEEE Commun. Letters, vol. 11, no. 12, Dec. 2007, pp. 1007-1009.

[14] ETSI TR 101 985. Satellite Earth Stations and Systems (SES), Broadband Satellite Multimedia (BSM); IP over Satellite. Technical Report. V0.2.0, (2002).

[15] N. Celandroni, F. Davoli, E. Ferro, “Static and Dynamic Resource Allocation in a Multiservice Satellite Network with Fading,” Internat. J. of Satellite Commun. and Networking, vol. 21, no. 4-5, July-Oct. 2003, pp. 469-487.