A bandwidth allocation strategy for multimedia traffic in a satellite network

Raffaele Bolla°, Franco Davoli°, Mario Marchese*
°Department of Communications, Computer and Systems Science (DIST)
University of Genoa, Italy
*CNIT - Italian Consortium for Telecommunications
DIST - University of Genoa Research Unit
Via Opera Pia 13, 16145 Genova (Italy)
Ph. +39-010-3532985
Fax. +39-010-3532154
° lelus, franco@ dist.unige.it; * mario.marchese@cnit.it

Abstract - **The paper proposes a bandwidth allocation scheme adapted for satellite networks. The traffic considered is divided into two categories: guaranteed traffic and non-guaranteed best-effort traffic. The control scheme is aimed at keeping the call blocking probability of the guaranteed traffic below a given threshold and at minimising the packet discarding probability of the best-effort portion in case of degradation of the satellite channel. In that way a minimum grade of service should be guaranteed for both traffics. The allocation scheme is designed as a two-level hierarchical mechanism: the upper level allocates a portion of the overall bandwidth to each station, the lower level shares the received portion between guaranteed and non-guaranteed traffic. The results show the behaviour of the algorithm and its performance.**

I. INTRODUCTION

The increasing interest to deliver multimedia services via satellite, in particular over less utilised bandwidths as Ka-band (20-30 GHz), is justified by the several advantages of the satellite environment with respect to cable networks. Ka-band has been object of test since 1991. The first tests were performed over the Italian satellite ITALSAT, which is also the reference of this work. Many national and international programs and projects in Europe, Japan and USA concern satellite networks and applications over Ka-band. NASA ACTS (summarised in [1] and [2]), and CNIT-ASI [3], which partially supports the present work, deserve a particular attention, among many others.

In order to provide multimedia services over satellite it is important that the used systems manage efficiently the various satellite and bandwidth resources [4]. The topic has been widely investigated in the literature concerning cabled networks: many congestion control and resource allocation techniques have been applied to obtain a fair allocation of voice, video and data flow. The need of guaranteeing a certain QoS has allowed to develop dynamic bandwidth allocation techniques, which take into account the current status of the channel. These works may represent a reference to design control schemes for satellite channels; nevertheless, the satellite environments are characterised by several peculiarities, which imply the introduction of suited control strategies. Differently from cabled and also wireless networks for personal communications, satellite channels vary their characteristics depending on the weather and the effect of fading heavily affects the performance of the whole system [5], in particular for systems operating in the Ka-band [6, 7].

This paper introduces a bandwidth allocation scheme suited for a satellite environment by considering as a starting point a bandwidth allocation scheme for cable networks [8] and by taking care of the fading effect in the allocation process.

The paper is structured as follows. Section II contains the description of the general framework: the characteristics of the network topology, of the bandwidth allocation scheme and of the traffic models used. Section III describes the control system: the cost functions and the optimisation problem. The results are contained in Section IV. Section V reports the conclusions.

II. GENERAL FRAMEWORK

A. Network topology

N land stations compose the network considered. The land stations are of two types: the Master Control Station (MCS), which contains the Centralised Network Control Center (CNCC) and has the role of checking and monitor the status and the available resources of the overall network; the remote stations, which manage the portion of the bandwidth assigned by the CNCC. The traffic considered may be synchronous or asynchronous. Synchronous flows require a certain level of Quality of Service (QoS) and have to be completely guaranteed; asynchronous traffic has no strict performance requirements and the network does its best to provide a minimum level of quality ("best effort" traffic).

B. The bandwidth allocation scheme

The designed control scheme manages the up-link available bandwidth and, for each land station, it is aimed at keeping the call blocking probability of the guaranteed traffic below a given threshold and minimising the packet discarding probability of the best effort portion.

The control architecture is organised in two hierarchical levels, along similar lines as in [8]. The upper level, called Centralised Bandwidth Allocator (CBA), periodically allocates the available bandwidth C [bit/s] by assigning a portion $C^{(i)}$ [bit/s] of the total bandwidth to each earth station i. The CBA acts with slow timing

This work has been partially supported by the Italian Space Agency (ASI) under the CNIT contract ASI-ARS-99-72 and by the National Research Council (CNR) under the CNIT contract CNR 99.0333.

0-7803-6451-1/00/$10.00 © 2000 IEEE

and is located in the CNCC of the MCS. The lower level, called Local Controller (LC), acts with a faster timing than the CBA and is located in each remote earth station i. It shares the bandwidth $C^{(i)}$, allocated to station i, between guaranteed ($C_g^{(i)}$ [bit/s]) and non-guaranteed ($C_{ng}^{(i)}$ [bit/s]) traffic. It performs Call Admission Control (CAC) of the incoming guaranteed calls and measures the statistics necessary for successive allocations. Let $N_{max}^{(i)}$ be the maximum number of guaranteed traffic calls acceptable at station i to provide a certain grade of service. Time t has been dropped to simplify the notation; actually all the quantities used may be thought as time variant ($C^{(i)}(t), C_g^{(i)}(t), C_{ng}^{(i)}(t), N_{max}^{(i)}(t)$). $Q^{(i)}$ is the buffer dimension dedicated to the non-guaranteed traffic in the LC multiplexer of the station i.

C. Traffic models

1) Synchronous Guaranteed Traffic

Each synchronous connection is defined as Constant Bit Rate (CBR) flows at B kbit/s. Each station is considered independent of the others. Let $\lambda^{(i)}\left[s^{-1}\right]$ be the arrival rate of the connection requests and $\frac{1}{\mu^{(i)}}$ [s] the average duration of each connection, for the station i. Due to the fact that only one type of synchronous traffic is considered, no index is introduced to distinguish the possible flows. Exponential distributions are used both from the inter-arrival time and the service time. If, as defined in the previous sub-section, $N_{max}^{(i)}$ is the maximum number of calls acceptable at station i, the call blocking probability at station i is

$$P_B^{(i)}(N_{max}^{(i)}) = \frac{\dfrac{\left(\lambda^{(i)}/\mu^{(i)}\right)^{N_{max}^{(i)}}}{N_{max}^{(i)}!}}{\sum_{n=0}^{N_{max}^{(i)}} \dfrac{\left(\lambda^{(i)}/\mu^{(i)}\right)^n}{n!}} \qquad (1)$$

2) Asynchronous Non-guaranteed Traffic.

A self-similar traffic model with a Pareto distribution [9], is considered. All the mathematical details, contained in [9] and [10], have been dropped. It is important to say that the quantity of interest is the cell loss probability of the non-guaranteed traffic in the queue of station i, averaged over the number of guaranteed connections.

The bandwidth allocated to the non-guaranteed traffic at station i ($C_{ng}^{(i)}(t)$) at time t is variable and it depends on the number of guaranteed connections at the same instant (let them be $n^{(i)}(t)$).

$$C_{ng}^{(i)}(t) = C^{(i)}(t) - B \cdot n^{(i)}(t) \qquad (2)$$

The quantities in (2) have been defined in section II, except for the time dependence. Time has been inserted to underline the fact that the non-guaranteed traffic, at each instant of time, takes the residual bandwidth not used, at that instant, by the guaranteed traffic. Index time will be dropped again where not explicitly necessary.

The quantity $n^{(i)}(t)$ can assume only discrete values from 0 to $N_{max}^{i}(t)$; as a consequence $C_{ng}^{(i)}(t)$ will assume only discrete values with a certain probability, depending on the probability of having $n^{(i)}(t)$ connections at time t at the station i.

The indication of the packet loss rate at station i averaged over the number of guaranteed connections is identified as $\overline{P}_{loss}^{(i)}(t, C_{ng}^{(i)}(t), N_{max}^{(i)}(t))$. It is important to note the dependence of the loss probability on the threshold $N_{max}^{(i)}$ and on the overall bandwidth $C^{(i)}$ allocated to station i.

III. THE CONTROL SCHEME

A. The lower level optimisation problem

The aim of this sub-section is defining the optimisation problem for the LC layer at the generic station i. LC will share the bandwidth $C^{(i)}$ allocated to the station i between guaranteed and non-guaranteed traffic. The lower level optimisation problem will evaluate the threshold $N_{max}^{(i)}$ for a specific value of $C^{(i)}$.

The problem has been formalised as follows: given the bandwidth $C^{(i)}$, allocated to station i, evaluate the minimum number of acceptable connections ($N_{max}^{(i)}$, i.e. the maximum number of calls, which the station i can accept), such that the call blocking probability (1), be, if possible, lower than $\gamma^{(i)}$. $N_{max}^{(i)}$ may assume values in the range $\left[0,\ldots,\left\lfloor \frac{C^{(i)}}{B} \right\rfloor\right]$.

Let $C_{min}^{(i)}$ be minimum bandwidth needed to obtain a call blocking probability lower than $\gamma^{(i)}$.

$$C_{min}^{(i)} = \arg\min_{X^{(i)}} \left\{ X^{(i)} \in \Re : P_B^{(i)}\left(\left\lfloor \frac{X^{(i)}}{B} \right\rfloor\right) \le \gamma^{(i)} \right\} \qquad (3)$$

The value of $C_{min}^{(i)}$ may be computed off-line.

The minimum threshold will be

$$N_{max}^{(i)} = \begin{cases} \left\lfloor \dfrac{C_{min}^{(i)}}{B} \right\rfloor & \text{if } C_{min}^{(i)} < C^{(i)} \\ \left\lfloor \dfrac{C^{(i)}}{B} \right\rfloor & \text{if } C_{min}^{(i)} \ge C^{(i)} \end{cases} \qquad (4)$$

If the bandwidth allocated by the upper layer is higher than the minimum threshold, the available bandwidth is shared between the guaranteed and non-guaranteed flow, assuring a portion sufficient to guarantee the required QoS (represented by $\gamma^{(i)}$) for synchronous traffic. If the value of $C^{(i)}$ is lower than the minimum bandwidth required, all the bandwidth is given to synchronous traffic.

Both from the computation and the technological point of view, it is simpler to assume $C^{(i)} \in \Re : C^{(i)} = k \cdot mpb, \forall k \in N, C^{(i)} \leq C$, where mpb is the minimum portion of bandwidth that can be allocated. If mpb is very small, the algorithm is very flexible but the computational load increases.

B. Channel modelling

The fading effect on the channel 'seen' by station i is modelled as a reduction of the capacity $C^{(i)}$ allocated. The real capacity utilised by station i may be written as $C^{(i)}_{real} = \beta^{(i)}_{level}(t) \cdot C^{(i)}$, where $\beta^{(i)}_{level}(t)$ is a coefficient to weight the channel degradation. The index 'level' identifies the level of the degradation. The time identification has been re-introduced just to focus on the time dependence of the quantity. It will be dropped again in the following. A certain probability $p^{(i)}_{level}$ is associated to each level of degradation. So, $\beta^{(i)}_{level}(t)$ represents the channel degradation level, $p^{(i)}_{level}$ its probability.

C. The upper level optimisation problem.

The upper level optimisation problem is aimed at defining the values $C^{(i)}$ for each station i. Let $Z^{(i)}$ be a generic variable.

$$J^{(i)}(Z^{(i)}) = \begin{cases} \overline{P}^{(i)}_{loss}(Z^{(i)}, N^{(i)}_{max}(Z^{(i)}) & \text{if } Z^{(i)} \geq C^{(i)}_{min} \\ H & \text{if } Z^{(i)} < C^{(i)}_{min} \end{cases} \quad (5)$$

is the cost function for station i. H is a very large number. $N^{(i)}_{max}(Z^{(i)})$ is the function defined in (6).

$$N^{(i)}_{max}(Z^{(i)}) = \begin{cases} \left\lfloor \dfrac{C^{(i)}_{min}}{B} \right\rfloor & \text{if } C^{(i)}_{min} < Z^{(i)} \\ \left\lfloor \dfrac{Z^{(i)}}{B} \right\rfloor & \text{if } C^{(i)}_{min} \geq Z^{(i)} \end{cases} \quad (6)$$

It can be noted that the formula is the same as in (4), but, if in (4), a specific number has to be computed after receiving the value $C^{(i)}$ from the upper level controller, (6) is just a function and $C^{(i)}$ is the value to be computed.

Being N the total number of stations, the overall cost function is:

$$J(Z^{(1)}, Z^{(2)}, ..., Z^{(N)}) = \sum_{i=1}^{N} \sum_{level} p^{(i)}_{level} \cdot J^{(i)}(\beta^{(i)}_{level} \cdot Z^{(i)}) \quad (7)$$

Even in this case $Z^{(i)} \in \Re : Z^{(i)} = k \cdot mpb, \forall k \in N, Z^{(i)} \leq C$. The aim is to find the particular values of $Z^{(i)} = C^{(i)}$ that minimise the function in (7). The minimisation problem is described in (8),

$$C^{(i)}, i = 1,...,N : J(C^{(1)},...,C^{(N)}) \leq J(Z^{(1)},...,Z^{(N)}), \forall (Z^{(1)},...,Z^{(N)}) \neq (C^{(1)},...,C^{(N)}) \quad (8)$$

with constraints,

$$\begin{cases} \sum_{i=1}^{N} Z^{(i)} = C \\ Z^{(i)} \geq 0, \forall i \in [1,...,N] \end{cases} \quad (9)$$

The problem deriving from (8) and (9) admits solution if $\sum_{i=1}^{N} C^{(i)}_{min} < C$; otherwise the allocation is performed as follows:

$$\begin{aligned} &\sum_{i=1}^{N} C^{(i)}_{min} = C \Rightarrow C^{(i)} = C^{(i)}_{min} \\ &\sum_{i=1}^{N} C^{(i)}_{min} > C \Rightarrow C^{(i)} = \frac{C}{\sum_{i=1}^{N} C^{(i)}_{min}} \cdot C^{(i)}_{min} \end{aligned} \quad (10)$$

A simpler choice for the global cost function might be to ignore the channel degradation and to use the formula in (11) below. This solution is used as a comparison in the results.

$$J(Z^{(1)}, Z^{(2)}, ..., Z^{(N)}) = \sum_{i=1}^{N} J^{(i)}(Z^{(i)}) \quad (11)$$

Note that cost (7) is implicitly supposed to span an infinite time horizon (stationary distribution are used).

IV. RESULTS

Four stations are involved. Three levels have been used to model the degradation: level={optimal (opt), degradation (degr), serious degradation (ser_degr)}. $\beta^{(i)}_{opt} = 1, \forall i \in [1,N]$, $\beta^{(i)}_{degr} = 0.5, \forall i \in [1,N]$, $\beta^{(i)}_{ser\_degr} = 0, \forall i \in [1,N]$. The minimum portion of bandwidth (mpb) that can be allocated has been fixed to 128 kbits/s. C=8 Mbits/s; the following parameters have been used:

$$\gamma^{(i)} = 0.05, \frac{1}{\mu^{(i)}} = 1200s, B = 128\,\text{kbits/s}, Q^{(i)} = 8000\,\text{cells}, \quad \forall i \in [1,N]$$

Length of a cell = 424 bits, $B^{(i)}_{ng} = 324$ kbits/s (Peak generation rate).

The tests presented are aimed at showing the behaviour of the strategy proposed, called Adaptive Bandwidth Allocation in Satellite Channels (ABASC) algorithm and at highlighting the advantages with respect to the strategy that does not take into account the channel degradation (formula (11)), identified as 'no degradation considered' algorithm.

The channels of three stations are not subject to degradation at all, while the degradation is increasing with the number of the test, concerning the fourth station. The traffic flows imposed are reported in Table I.

TABLE I
TRAFFIC FLOWS

Traffic	Station 1	Station 2	Station 3	Station 4
λ [conn/s]	0.006	0.006	0.003	0.003
λ_{burst} [burst/s]	1200	1200	600	600

The channel behaviour of the first three stations is fixed and reported in Table II. Table III contains the channel degradation 'seen' by station 4.

TABLE II
CHANNEL PARAMETERS

	$p^{(i)}_{opt}$	$p^{(i)}_{degr}$	$p^{(i)}_{ser\_degr}$
Station 1, 2, 3 (i=1, 2, 3)	1	0	0

TABLE III
CHANNEL DEGRADATION, STATION 4

	$p^{(4)}_{opt}$	$p^{(4)}_{degr}$	$p^{(4)}_{ser\_degr}$
Test 1	1	0	0
Test 2	0.99	0.01	0
Test 3	0.98	0.02	0
Test 4	0.95	0.05	0
Test 5	0.9	0.1	0
Test 6	0.8	0.2	0
Test 7	0.7	0.3	0
Test 8	0.6	0.4	0
Test 9	0.5	0.5	0
Test 10	0.4	0.6	0
Test 11	0.2	0.8	0
Test 12	0	1	0

The allocations performed have been reported in Table IV, both for the ABASC and the 'no degradation considered' scheme (ndc). The minimum bandwidth required $C^{(i)}_{min}$ is also considered.

TABLE IV
BANDWIDTH ALLOCATIONS (MEASURED IN mpb)

	Station 1	Station 2	Station 3	Station 4
$C^{(i)}_{min}$	12	12	7	7
ndc	20	20	12	12
ABASC				
Test1	20	20	12	12
Test2	19	20	11	14
Test3	19	20	11	14
Test4	19	20	11	14
Test5	19	20	11	14
Test6	19	20	11	14
Test7	19	20	11	14
Test8	19	20	11	14
Test9	19	19	10	16
Test10	19	19	10	16
Test11	18	18	10	18
Test12	18	18	10	18

Clearly, if there is no degradation (Test1) the two schemes provide the same results; even if a minimum degradation in station 4 is introduced (e.g. Test2), the ABASC allocations vary: more bandwidth is allocated to the station more penalised by the degradation. The bandwidth is taken from the stations where there is no degradation.

Table V, Table VI and Fig. 1 show the percentage of blocked calls (i.e. the expression in percentage of the call blocking probability) versus the channel degradation (test), for the two allocation methods. Fig. 2, Fig. 3 and Fig 4 show the percentage of dropped packets (i.e. the packet dropping probability) of asynchronous traffic for the same tests of the previous three figures. Stations 1 and 2 show the same behaviour, both for the call blocking and the packet dropping probability (Table 1 and Fig. 2) because the same amount of bandwidth is provided to both stations, which have also the same traffic load and degradation level. If there is no degradation (stations 1, 2 and 3), the performance of the guaranteed traffic are the same for both the ABASC and the 'no degradation considered' scheme. In the two cases, the percentage of blocked calls is constant for all the tests because the bandwidth allocation is always higher than the minimum threshold. The improvement if ABASC is used can be seen in Fig. 1. The algorithm proposed allows maintaining a constant call blocking probability even for serious degradations. This is not true for the other scheme. The behaviour may be explained as follows: even if the real bandwidth is strongly reduced because of the satellite channel degradation (e.g. it is halved, with different probabilities, in the case considered), the ABASC algorithm allows to allocate a portion of bandwidth sufficient to guarantee the minimum capacity required (7 mpbs) to station 4 in any case. The 'no degradation considered' scheme allocates always 12 mpbs to station 4, which, if reduced for the degradation, are not sufficient to guarantee the QoS required and the performance drastically reduces. The graphs in Fig. 2 and Fig. 3 show that the advantage in the call blocking probability is 'paid' by an increase in the percentage of dropped packets for the stations where there is no channel degradation if ABASC is used. This algorithm, in fact, transfers the residual bandwidth, utilised by the non-guaranteed traffic, to the degraded station. On the other hand, the advantage of using ABASC is outstanding from Fig. 4 for station 4. So, the ABASC strategy allows to maintain the required QoS for the guaranteed traffic in all the stations and to improve the performance concerning non-guaranteed traffic in the stations where there is disturbance. On the contrary, there is a decrease concerning the non-guaranteed traffic performance in the stations where there is no degradation but the drawbacks are really slight compared to the advantages obtained.

TABLE V
PERCENTAGE OF BLOCKED CALLS, STATIONS 1 AND 2

	Percentage of blocked calls
Test 1 to 12	3.12 %

TABLE VI
PERCENTAGE OF BLOCKED CALLS, STATION 3

	Percentage of blocked calls
Test 1 to 12	4.38 %

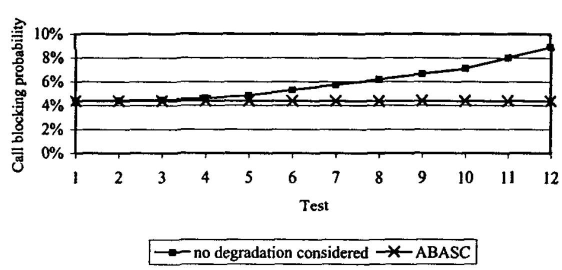

Fig. 1. Percentage of blocked calls, station 4.

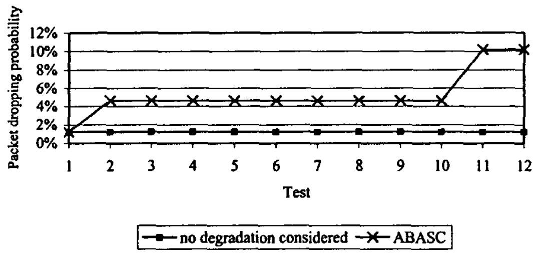

Fig. 2. Percentage of dropped packets, stations 1 and 2.

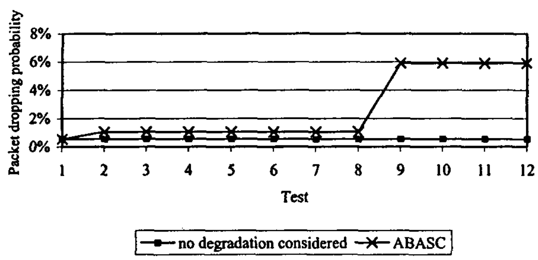

Fig. 3. Percentage of dropped packets, station 3.

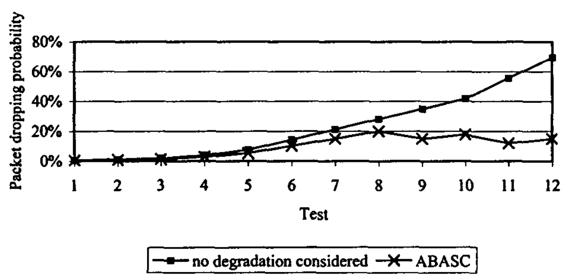

Fig. 4. Percentage of dropped packets, station 4.

V. CONCLUSIONS

The paper presents a bandwidth allocation algorithm to support multimedia traffic in a satellite environment (ABASC). A synchronous traffic, which needs precise Quality of Service (QoS) requirements, and an asynchronous "best effort" traffic have been considered. N land stations connected through a geostationary satellite, where a station has the role of master, which manages satellite network resources, compose the network.

The scheme proposed is aimed at keeping a given constraint on the call blocking probability of the guaranteed traffic and at minimising the packet discarding probability of the best effort portion, while guaranteeing a minimum grade of service for both traffics involved even in case of degradation of the satellite channel.

The results reported have shown that the ABASC strategy allows to maintain the required QoS for the guaranteed traffic in all stations and to improve the performance concerning non-guaranteed traffic in the stations where there is channel degradation. On the contrary, there is a slight decrease concerning the non-guaranteed traffic performance in the stations where there is no degradation.

REFERENCES

[1] D. E. Brooks, C. Buffinton, D. R. Beering, A. W. William D. Ivancic, M. Zernic, D. J. Hoder, "ACTS 118x Final Report, High Speed TCP Interoperability Testing", *NASA/TM - 1999-209272*, July 1999.

[2] R. T. Gedney, R. Schertler, F. Gargione, "*The Advanced Communications Technology Satellite*", SciTech Publishing, Mendham, NJ, 2000.

[3] D. Adami, A. Conti, D. Dardari M. Marchese, L. S. Ronga, "CNIT-ASI Project 'Integration of Multimedia Services on Heterogeneous Satellite Networks': An Overview", *Proc. Fifth Ka Band Utilization Conference*, October 1999, Taormina, Italy, pp. 219-226.

[4] V. Birmani, "Resource allocation for Ka-band broadband satellite systems", *CSHCN MS 99-6, Master of Science Thesis*, 1999, also in http://www.isr.umd.edu/TechReports/CSHCN/1999/ (last access August 29th, 2000).

[5] Louis J. Ippolito, "Propagation Consideration for Low Margin Ka-Band Systems". *Proc. Ka Band Utilization Conference*, Italy, September 1997, pp. 161-167.

[6] Richard T. Gedney e Thom A. Coney, "Effective Use of Rain Fade Compensation for Ka-Band Satellites". *Proc. of Ka Band Utilization Conference*, Italy, September 1997, pp. 153-160.

[7] Roberto J. Acosta, "Rain Fade Compensation Alternatives for Ka-Band Communication Satellites". *Proc. Ka Band Utilization Conference*, Italy, September 1997, pp. 145-152.

[8] R.Bolla, F.Davoli, S.Ricciardi, "A hierarchical control structure for multimedia access networks", *Proc. IEEE Internat. Conf. Commun. (ICC'99),* Vancouver, Canada, June 1999.

[9] B. Tsybakov, N.D. Georganas, "On self-similar traffic in ATM queues: Definition, overflow probability bound, and cell delay distribution". *IEEE/ACM Transaction on Networking*, 5(3), 1997, pp. 397-409.

[10] B. Tsybakov, N.D. Georganas, "Overflow probability in ATM queue with self-similar input traffic". *Proc. IEEE International Conference on Communications (ICC'97),* Montreal, Canada, 1997.