Space Communications 19 (2004) 171–182
IOS Press

E–CAP–ABASC versus CAP–ABASC: Comparison of two resource allocation strategies in satellite environment

Igor Bisio and Mario Marchese *
CNIT – Italian National Consortium for Telecommunications, University of Genoa Research Unit, Via Opera Pia 13, 16145, Genova, Italy
Tel.: +39 010 3532806; Fax: +39 010 3532154; E-mail: igor.bisio@cnit.it, mario.marchese@cnit.it

Abstract. Resource allocation schemes dedicated to Satellite environment often consider Internet traffic as the superposition of traffic sources without distinguishing between TCP and UDP flows, even if TCP and UDP have different traffic characteristics. In this paper we introduce a system control architecture with three types of flows entering the network, i.e., Constant Bit Rate (CBR), UDP and TCP, and a cost function including an analytical measure of the packet loss for TCP flow. We propose an extension of our previous bandwidth allocation control algorithm Constrained Average Probability–Adaptive Bandwidth Allocation in Satellite Channel (CAP–ABASC) to Extended CAP–ABASC. We provide performance evaluation of the two allocation strategies.

Keywords: Resource Allocation, TCP, congestion control, performance evaluation

1. Introduction

Considering the applications that require TCP/IP with the advantages offered by satellites, it is natural to think of TCP/IP-based applications over satellite networks, but the general characteristics (e.g., the presence of fading) of satellite links heavily affect the performance of the communication. Resource allocation is an issue of particular importance in this environment.

Within this framework, the paper considers the adaptive bandwidth allocation system, Constrained Average Probability–Adaptive Bandwidth Allocation in Satellite Channels, CAP–ABASC, proposed in [3]. The satellite network is composed of earth stations connected through a geostationary satellite. An earth station (or the satellite itself, if switching on board is allowed) represents the master, which manages the resources and provides the other stations with a portion of the overall bandwidth; each station shares the assigned portion between its traffic flows. Three types of traffic are considered: a *QoS* guaranteed traffic, operating at a fixed speed (measured in kbps) and two non-guaranteed best-effort traffic: UDP, modeled by a Pareto distribution [8,10], and TCP, whose packet loss model is introduced in [2] and briefly summarized in this paper. The fading is considered by assigning a probability of channel degradation to each link, along with a weighting coefficient to "measure" the degradation itself. The TCP packet loss formulation, together with the CBR and UDP models, is used to develop a new cost function and a bandwidth allocation scheme, termed Extended–CAP–ABASC, which forms the main contribution of this paper. The performance of E–CAP–ABASC is compared with CAP–ABASC.

The paper is structured as follows: Section 2 contains the description of the network topology and the channel model. Section 3 states the overall system architecture and Section 4 summarizes the source models for the considered traffic. Section 5 reports the E–CAP–ABASC proposal. Section 6 shows the performance evaluation and the conclusions are given in Section 7.

2. Network topology and channel model

The satellite network is composed of I earth stations. One station is the "master" and controls the al-

*Corresponding author.

location of the satellite resources. Each station gathers traffic from users, connected either directly or through local area networks. Even though the study is not linked to a particular satellite frequency band, Ka-band (20–30 GHz) may be considered the reference case, since fading due to rain is a serious concern. The real availability of channel bandwidth is strictly dependent on rain fading compensation. It is very important to study the fading effect, which results in bandwidth reduction [3]. Mathematically, it means that the nominal bandwidth C_{tot} (assigned to a station) is reduced by a factor β, a stochastic parameter distributed in the real numbers interval [0, 1].

$$C_{real} = \beta \cdot C_{tot}. \tag{1}$$

A specific value β corresponds to a fixed fading level. A technical interpretation of the factor β may be the bandwidth reduction due to the presence of a Forward Error Correction (FEC) scheme. Each fading level, happening with an associated probability p_f, a particular FEC. In satellite networks, link layer corruption due to noise is typical and packet loss occurs. Nevertheless, if FEC schemes are used, link layer corruption may be seen as congestion problem. In other words, FEC strategies make channel errors negligible, but reduce the available service capacity. In this paper we explicitly consider the capacity reduction through the factor β and assume that all the packet losses happening during communication may be considered due to congestion. With the use of FEC, packet loss due to link layer corruption tends to zero, and the simple channel model of Eq. (1) seems to be a reasonable approximation of satellite channel behavior.

3. Control system architecture

The traffic considered is divided into three types: CBR, UDP and TCP. CBR flow is privileged since a *QoS* threshold, in terms of call blocking probability, is assured through a call admission control (CAC) scheme. TCP and UDP, are considered as two separate components. The traffic is conveyed to each earth station and, virtually, enters the system bandwidth allocation. The overall system architecture is shown in Fig. 1.

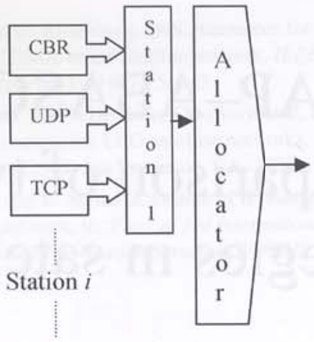

Fig. 1. Control system architecture.

4. Source models for CBR, UDP and TCP

The models reported in the following are taken from the literature. The models for CBR and UDP have been already used in CAP–ABASC [3] and are quickly summarized here. An analytical expression for TCP packet loss probability over geostationary satellite channels has been proposed by the authors, together with the performance evaluation [2]. The expression obtained is a function of the bandwidth available and is suited to control mechanisms.

4.1. CBR

The relevant performance metric, used in the allocation strategy, is the call blocking probability modeled by the Erlang B formula (M/M/m/m queuing system), given in Eq. (2):

$$P_B(k_{\max}) = B\left(\frac{\lambda}{\mu}; k_{\max}\right), \tag{2}$$

$k_{\max}$ is the maximum number of servers; λ and μ are the call arrival rate and the service rate, respectively, and both of them are exponentially distributed. Given $k(t) \leqslant k_{\max}$ active CBR sources at the instant t (i.e., having $k(t)$ busy servers), emitting data at R_{CBR} (kbps), the bandwidth used by real-time CBR traffic is, at time t:

$$C_{CBR} = k(t) \cdot R_{CBR}. \tag{3}$$

4.2. UDP

The performance metric used in the allocation scheme is the packet loss probability. The model used is Y/D/Cs/Q according to [10]. It has been used in the CAP–ABASC strategy to model all the IP-based traffic with no distinction between UDP and TCP. When the model was proposed, there was not any reference to congestion control algorithm and to the transport layer used. Nevertheless, the particular type of statistical ON-OFF fractal traffic nature suggests its use for UDP traffic, where no acknowledgement-based mechanism is provided for congestion control. The analytical approximation is given by Eq. (4):

$$P_{loss}^{UDP}(C_s) = \min\left\{ \frac{c\lambda_s(\alpha(\alpha-1))^{-1}}{(C_s-\lambda_s\bar{\tau})} Q_{UDP}^{-(\alpha+1)}, 1 \right\}, \quad (4)$$

C_s is the number of servers busy in the Y/D/Cs/Q system and Q_{UDP} is the length (measured in packets of 1500 bytes) of the IP buffer dedicated to UDP sources, which is supposed to tend to infinite. α is the Pareto parameter ($1 < \alpha < 2$), c is a normalization constant and λ_s is the arrival rate of UDP sources equal to $\lambda_{asy} \cdot T$, where T is the source packetization time and λ_{asy} the burst of packets generated by the UDP sources. The approximation reported in Eq. (4) is valid if $C_s > \lambda_s\bar{\tau}$, otherwise the packet loss probability is 1. It is important to specify the relation between C_s and the transmission bandwidth available for this kind of source. If the peak bandwidth for each UDP source is B_p, the average value of IP packet length is L. T is defined as L/B_p [10].

$$C_s = \frac{C_{UDP}}{L}T = \frac{C_{UDP}}{L}\cdot\frac{L}{B_p} = \frac{C_{UDP}}{B_p}, \quad (5)$$

C_{UDP} is the transmission bandwidth expressed in kbps dedicated to UDP traffic.

4.3. TCP

The scenario considered is shown in Fig. 2. T_n is the round trip time at the TCP layer for the n-th connection. It is supposed to be constant for each packet of the n-th connection.

W^{pipe} is the maximum volume of information that can be transmitted to the system with a channel server of capacity C_{TCP} (bits/second) and IP buffer of size Q_{TCP} (bits). C_n and Q_n (constant over time), are defined as the maximum portion of the capacity C_{TCP} and of the buffer Q_{TCP}, "seen" by the n-th connection. W_n^{pipe} is the maximum volume of information that can be transmitted to the system by the n-th connection, and is given by Eq. (6).

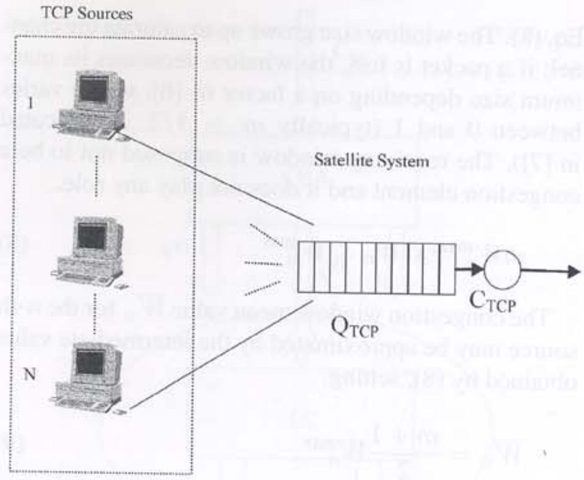

Fig. 2. TCP model.

$$W^{pipe} = \sum_{j=1}^{N} W_j^{pipe} = \sum_{j=1}^{N}(C_j \cdot T_j + Q_j). \quad (6)$$

The following "long term behavior hypotheses", are necessary for this model:

1. Each source always has data to send [9].
2. The number of sources saturates the channel [9]; this hypothesis implies 1.
3. The sources are synchronized [1,7,9].
4. Only the congestion avoidance phase is considered [9].
5. The evolution of the current congestion window for a generic i-th connection is described by a Markov regenerative process with rewards [9].
6. Losses are due only to congestion [7,9].

The flow synchronization hypothesis (3, in the previous list), extended to an aggregate of TCP sources, implies Eq. (7), where W_j, W_n, T_j, T_n are, respectively, the current congestion window (which, during the congestion avoidance phase, is the average number of packets in flight) and the round trip time for two generic sources j and n that belong to the set of integers $[1, N]$:

$$W_jT_j = W_nT_n, \quad \forall j, n \in [1, N]. \quad (7)$$

Being in the congestion avoidance phase (assuming one packet loss a time, which is hypothesis 4) and considering TCP Reno as reference, the dimension of the congestion window W_n of a generic source n varies between a minimum and maximum value as introduced in [7]. The TCP-Reno simplified model is shown in

Eq. (8). The window size grows up to saturate the channel; if a packet is lost, the window decreases its maximum size depending on a factor m [6], which varies between 0 and 1 (typically $m = 1/2$, as indicated in [7]). The receiving window is supposed not to be a congestion element and it does not play any role.

$$mW_n^{\max} \leqslant W_n \leqslant W_n^{\max}. \tag{8}$$

The congestion window mean value $\overline{W}_n$ for the n-th source may be approximated by the intermediate value obtained by (8), setting:

$$\overline{W}_n = \frac{m+1}{2} W_n^{\max}, \tag{9}$$

$W_n^{\max}$ is the maximum amount of data that can be sent by the n-th source within the TCP algorithm. It is physically limited by the quantity W_n^{pipe}, defined in Eq. (6).

$$W_n^{\max} = W_n^{pipe}. \tag{10}$$

According to reference [9] the average congestion window of the n-th TCP connection by using the hypotheses 1, 4 and 5 and assuming TCP Reno is given by:

$$\overline{W}_n(p_n) = \sqrt{\frac{8}{3b_n p_n}} + o\left(\frac{1}{\sqrt{p_n}}\right), \tag{11}$$

where p_n is the packet loss probability and b_n the number of packets covered by one acknowledgement for the connection n. $o(\cdot)$ is a function of p_n, which is negligible for very small p_n values.

From Eq. (7), it is also true that, for any j and n

$$\frac{W_j}{W_n} = \frac{T_n}{T_j}, \quad \forall j, n \in [1, N], \tag{12}$$

$$W_j = \frac{T_n}{T_j} W_n, \quad \forall j, n \in [1, N]. \tag{13}$$

Defining $R_n^{\max}$ as maximum value for the n-th TCP flow entering the system:

$$\sum_{j=1}^{N} R_j^{\max} = \sum_{j=1}^{N} \frac{W_j^{pipe}}{T_j} = \sum_{j=1}^{N} \frac{C_j \cdot T_j + Q_j}{T_j} = \sum_{j=1}^{N} C_j + \sum_{j=1}^{N} \frac{Q_j}{T_j} = C + \sum_{j=1}^{N} \frac{Q_j}{T_j}. \tag{14}$$

From Eq. (10) and Eq. (14),

$$\sum_{j=1}^{N} \frac{W_j^{\max}}{T_j} = C + \sum_{j=1}^{N} \frac{Q_j}{T_j}. \tag{15}$$

From equation (13), the maximum value is,

$$\sum_{j=1}^{N} \frac{T_n \cdot W_n^{\max}}{T_j^2} = T_n \cdot W_n^{\max} \sum_{j=1}^{N} \frac{1}{T_j^2} = C + \sum_{j=1}^{N} \frac{Q_j}{T_j}, \tag{16}$$

and

$$W_n^{\max} = \frac{C + \sum_{j=1}^{N} \frac{Q_j}{T_j}}{T_n \sum_{j=1}^{N} \frac{1}{T_j^2}}. \tag{17}$$

Applying Eq. (9),

$$\overline{W}_n = \frac{m+1}{2} \cdot \frac{C + \sum_{j=1}^{N} \frac{Q_j}{T_j}}{T_n \cdot \sum_{j=1}^{N} \frac{1}{T_j^2}}. \tag{18}$$

From Eq. (11) and Eq. (18), for small values of p_n:

$$\sqrt{\frac{8}{3b_n p_n}} = \frac{m+1}{2} \cdot \frac{C + \sum_{j=1}^{N} \frac{Q_j}{T_j}}{T_n \sum_{j=1}^{N} \frac{1}{T_j^2}}. \tag{19}$$

Extracting p_n from Eq. (19) and assuming $m = 1/2$ in the reminder of the paper,

$$p_n = \frac{32}{3b_n \cdot (m+1)^2} \cdot \frac{\left(T_n \sum_{j=1}^{N} \frac{1}{T_j^2}\right)^2}{\left(C + \sum_{j=1}^{N} \frac{Q_j}{T_j}\right)^2}. \tag{20}$$

Considering a GEO satellite, the round trip time may be supposed fixed and equal for all the sources, as in Eq. (21). This equality, together with the hypothesis 3 (synchronization), gives origin to "fairness", which is the condition when all the connections have equal share of bandwidth [5]. If the connections do not get exactly equal allocation, "fairness" may be quantified by an index that measures the "distance" from the ideal condition [5].

$$T_j = T_n = RTT, \quad \forall j, n \in [1, N]. \tag{21}$$

Substituting the expression (21) in Eq. (20); remembering that $\sum_{j=1}^{N} Q_j = Q_{TCP}$, being Q_{TCP} the overall buffer size, and that $\sum_{j=1}^{N}(1/T_j^2) = N \cdot (1/RTT^2)$, the equation (22) can be written as,

$$p_n = \frac{32N^2}{3 \cdot b_n \cdot (m+1)^2 \cdot (C_{TCP} \cdot RTT + Q_{TCP})^2} = c \cdot \frac{N^2}{3 \cdot b_n \cdot (C_{TCP} \cdot RTT + Q_{TCP})^2}. \tag{22}$$

c is a numerical constant equal to $128/27$. If $b_n = b$, for all n, the packet loss probability in Eq. (22) is independent of the index n (i.e. it is the same for each single source, keeping the other parameters fixed) and corresponds to the entire TCP aggregate. This assumption is maintained in the reminder of the paper.

5. E–CAP–ABASC

The control mechanism called E–CAP–ABASC (Extended–Constrained Average Probability–Adaptive Bandwidth Allocation in Satellite Channels), proposed in the paper, uses a global packet loss probability for best-effort traffic but separates the buffer dedicated to UDP and TCP and exploits Eq. (22), which explicitly takes in account the effect of the acknowledgement-based congestion control algorithms. Best-effort packet loss probability is defined as the probability that either one UDP packet or one TCP packet is lost or both of them are lost simultaneously. Formally, it may be written as a function of the bandwidth available and of the number of active TCP and UDP sources:

$$P_{loss}(C^{be}, N, M) = P_{loss}^{TCP} + P_{loss}^{UDP} - P_{loss}^{TCP} P_{loss}^{UDP} \tag{23}$$

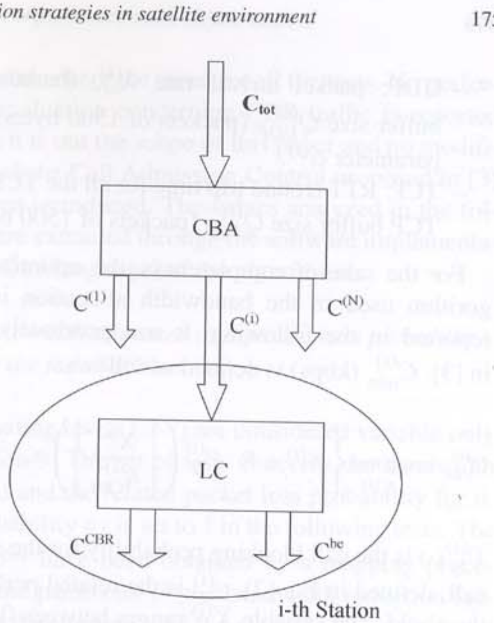

Fig. 3. Resource allocation control scheme.

P_{loss}^{TCP} is defined in Eq. (22). P_{loss}^{UDP} is modeled in Eq. (4), which, in CAP–ABASC, was used to describe generic best-effort traffic. C^{be} is the overall bandwidth for best-effort traffic (C_{TCP} plus C_{UDP}). The bandwidth assigned to UDP and TCP flows is proportional to the number of active sources, as in Eq. (24) and Eq. (25). M and N are the number of UDP and TCP sources, respectively.

$$C_{UDP} = \frac{M}{N+M} C^{be}, \tag{24}$$

$$C_{TCP} = \frac{N}{N+M} C^{be}. \tag{25}$$

The control scheme, as shown in Fig. 3, is composed of a higher and a lower layer. The higher layer is called Centralized Bandwidth Allocator (CBA) and distributes the bandwidth capacity among the earth stations. The lower layer is called Local Controller (LC) and splits the capacity allocated to each station into two contributions: C_{CBR} for CBR real time traffic and C^{be} for the Internet traffic including UDP and TCP. Each i-th earth station solves a local optimization problem to share the capacity assigned, so finding a threshold $k_{\max}^{(i)}$, which is the maximum number of acceptable CBR calls that allows guaranteeing a specific QoS level in terms of call blocking probability. The following parameters are defined for the stations:

- CBR: call arrival rate $\lambda^{(i)}$ (calls/s), call duration mean value $1/\mu^{(i)}$ (s), bit rate of the i-th call $R_{CBR}^{(i)}$ (kbps).

– UDP: packet arrival rate $\lambda_{asy}^{(i)}$ (burst/s), UDP buffer size $Q_{UDP}^{(i)}$ (packets of 1500 bytes), Pareto parameter $\alpha^{(i)}$.
– TCP: RTT (round trip time for all the TCP flows), TCP buffer size $Q_{TCP}^{(i)}$ (packets of 1500 bytes).

For the sake of completeness, the optimization algorithm used in the bandwidth allocation is briefly reported in the following. It was previously defined in [3]. $C_{\min}^{(i)}$ (kbps) is defined as follows:

$$C_{\min}^{(i)} = \arg\min_{X^{(i)}} \left\{ X^{(i)} \in \Re : P_B^{(i)}\left(\left\lfloor \frac{X^{(i)}}{R_{CBR}^{(i)}} \right\rfloor \right) \leqslant \gamma^{(i)} \right\} \quad (26)$$

$P_B^{(i)}(\cdot)$ is the call blocking probability for the i-th CBR call, defined in Eq. (2). $\gamma^{(i)}$ is the related performance threshold. The variable $X^{(i)}$ ranges between 0 and C_{tot}, which is the overall available channel bandwidth. $K_{\max}^{(i)}$ is the maximum number of calls guaranteed by $C_{\min}^{(i)}$. Eq. (27) establishes the relation among the two quantities.

$$C_{\min}^{(i)} = \left\lceil K_{\max}^{(i)} \cdot R_{CBR}^{(i)} \right\rceil. \quad (27)$$

For a given bandwidth $X^{(i)}$ assigned to the i-th station, Eq. (28) fixes the maximum number $\hat{K}_{\max}^{(i)}(X^{(i)})$ of real time (CBR) calls acceptable in the system.

$$\hat{K}_{\max}^{(i)}(X^{(i)}) = \begin{cases} \left\lfloor \dfrac{C_{\min}^{(i)}}{R_{CBR}^{(i)}} \right\rfloor & \text{if } X^{(i)} > C_{\min}^{(i)} \\ \left\lfloor \dfrac{X^{(i)}}{R_{CBR}^{(i)}} \right\rfloor & \text{if } X^{(i)} \leqslant C_{\min}^{(i)}, \end{cases} \quad (28)$$

$k^{(i)}(t) \leqslant \hat{K}_{\max}^{(i)}(X^{(i)})$ is the number of CBR calls in progress at time t, at station i. The bandwidth dedicated to CBR traffic is defined in Eq. (3) ($R_{CBR}^{(i)} \cdot k^{(i)}(t)$). The residual capacity ($C_{be}^{(i)} = X^{(i)} - R_{CBR}^{(i)} \cdot k^{(i)}(t)$) is available for the Internet traffic at the i-th station.

The basic idea is that a strong penalization is necessary if the average call blocking probability $\bar{P}_B^{(i)}(\cdot)$ is above the threshold $\gamma^{(i)}$. The constraint is explicitly indicated through the penalty function in Eq. (29).

$$F_{CAP}^{(i)}(X^{(i)}) = \begin{cases} 0 & \text{if } \bar{P}_B^{(i)}(\hat{K}_{\max}^{(i)}(X^{(i)})) \leqslant \gamma^{(i)} \\ H & \text{if } \bar{P}_B^{(i)}(\hat{K}_{\max}^{(i)}(X^{(i)})) > \gamma^{(i)}, \end{cases} \quad (29)$$

where H is a very large constant and

$$\bar{P}_B^{(i)}(\hat{K}_{\max}^{(i)}(X^{(i)})) = \sum_{f=1}^{F} p_f^{(i)} \cdot P_B^{(i)}(\hat{K}_{\max}^{(i)}(X^{(i)})), \quad (30)$$

F is the maximum number of fading levels and p_f their associated probability. As said above $P_B^{(i)}(\cdot)$ is defined in Eq. (2).

The Centralized Bandwidth Allocator (CBA) assigns the bandwidth portions to the earth stations by minimizing the cost function defined in Eq. (31):

$$J_{CAP}(X^{(1)}, \ldots, X^{(I)}) = \sum_{i=1}^{I} \sum_{f=1}^{F} p_f^{(i)} J_{CAP}^{(i)}(\beta_f^{(i)} X^{(i)}), \quad (31)$$

where

$$J_{CAP}^{(i)}(X^{(i)}) = P_{loss}^{(i)}(X^{(i)}, \hat{K}_{\max}^{(i)}(X^{(i)})) + F_{CAP}^{(i)}(X^{(i)}). \quad (32)$$

The solution is:

$$\{C^{(1)}, \ldots, C^{(I)}\} = \arg\min_{X^{(1)},\ldots,X^{(I)}} \{J_{CAP}(X^{(1)}, \ldots, X^{(I)})\} \quad (33)$$

$$\text{if } C_{tot} < \sum_{i=1}^{I} C_{\min}^{(i)} \Rightarrow C^{(i)} = \frac{C_{tot}}{\sum_{i=1}^{I} C_{\min}^{(i)}} \cdot C_{\min}^{(i)} \quad (34)$$

$P_{loss}^{(i)}$ is computed by using the new formula proposed in Eq. (23), applied for each station (i).

6. Performance comparison

The aim is to compare values of packet loss probability obtained using E–CAP–ABASC with results obtained with the packet loss probability using CAP–ABASC. Although QoS metrics such as throughput, end-to-end delay and delay jitter may be important, the allocation strategies are compared in terms of packet loss probability. This approach allows direct and fair comparison with CAP–ABASC, which studied only packet loss probability. E–CAP–ABASC (identified as E-CAP in the following) is compared with the CAP–ABASC strategy (where there is just one buffer for both kinds of best effort traffic), identified as CAP-1, as well as with an improved version of CAP–ABASC (CAP-2), which uses two separate buffers for TCP and UDP traffic but describes the packet loss of both through Eq. (4). It allows stating the additional value of the new allocation scheme regarding both buffer separation and use of a dedicated TCP packet loss model. The comparison is performed by varying the percentage of TCP and UDP traffic loading the system. The tests are carried out with 4 earth stations (numbered from 0 to 3) and using the following parameters for each i-th station.

CBR: $\lambda^{(i)} = 0.006$ calls/s, $1/\mu^{(i)} = 600$ s, $R_{CBR}^{(i)} = 128$ kbps, threshold $\gamma^{(i)} = 0.05$.

UDP: $\lambda_{asy}^{(i)} = 10$ burst/s, $Q_{UDP}^{(i)} = 8000$ packets, $\alpha^{(i)} = 1.5$, packet length $L = 1500$ bytes, peak rate $B_p = 64$ kbps and variable number M of UDP sources.

TCP: variable number N of TCP sources, a $RTT = 520$ ms and $Q_{TCP}^{(i)} = 8000$ packets, coherently with the UDP buffer.

To make a fair comparison among the algorithms, the single buffer dedicated to best-effort traffic (CAP-1) is set to 16 000 packets and $\lambda_{asy}^{(i)} = 20$ burst/s. These values are twice the corresponding values used in the CAP-2 and E-CAP cases. The parameter choices are directly imposed by the requirements of the model applied for UDP traffic, where the buffer dimension is supposed infinite.

The performance evaluation is structured into two parts: the first one shows the bandwidth allocation and the packet loss probability values only for the station in a fading condition; the second one shows the overall bandwidth allocation needed for all the stations to keep the packet loss probability under a threshold, and the gain of E-CAP with respect to CAP-2 for each station.

The overall number of best-effort active sources is 100 ($M + N = 100$), in all cases. The constraint over the call blocking probability for the CBR traffic is set to 0.05 and is kept the same for all the tests. No performance evaluation concerning CBR traffic is reported because it is out the scope of this paper and no modifications of the Call Admission Control proposed in [3] have been introduced. The values analyzed in the following are extracted through the software implementation (in C++ language) of the algorithms.

6.1. Bandwidth allocated and packet loss probability for the faded station

The fading levels ($\beta^{(i)}$) are considered variable only for Station 3. The set of tests concerns the bandwidth allocated and the related packet loss probability for it. The probability p_f is set to 1 in the following tests. The values $\beta^{(i)}$ have been obtained by a mapping procedure of the parameter β (the bandwidth reduction factor) with real measures of the signal to noise ratio and FEC (from [4]) and reported in Table 1. Stations 0, 1 and 2 are in clear sky conditions.

The overall bandwidth available C_{tot} is set to 8 Mbps. The tests are performed by unbalancing the TCP and UDP traffic of the Station 3, which is the earth station experiencing critical meteorological conditions. The results reported in Tables from 2 to 6, by increasing the value of β and expressing the bandwidth in kbps,

Table 1

Fading class and related β level for station 3

Fading class	C/N$_0$ (dB)	β
1	>77.13	1
2	74.63–77.13	0.8333
3	72.63–74.63	0.625
4	69.63–72.63	0.3125
5	66.63–69.63	0.15625
6	<66.63	–

Table 2

Bandwidth allocated (in kbps) to station 3 and packet loss probability comparison (β level for station 3 equal to 0.15625)

Best-Effort Traffic	Bandwidth allocated (kbps)			Packet loss probability		
TCP–UDP	CAP-1	CAP-2	E-CAP	CAP-1	CAP-2	E-CAP
10%–90%	3968	0	4352	0.82950	1	0.4
20%–80%	3968	0	4352	0.82950	1	0.4287
30%–70%	3968	0	4608	0.82950	1	0.4368
40%–60%	3968	0	4224	0.82950	1	0.6617
50%–50%	3968	0	3968	0.82950	1	0.8325
60%–40%	3968	0	0	0.82950	1	1
70%–30%	3968	0	0	0.82950	1	1
80%–20%	3968	0	0	0.82950	1	1
90%–10%	3968	0	0	0.82950	1	1

Table 3

Bandwidth allocated (in kbps) to station 3 and packet loss probability comparison (β level for station 3 equal to 0.3125)

Best-Effort Traffic	Bandwidth allocated (kbps)			Packet loss probability		
TCP–UDP	CAP-1	CAP-2	E-CAP	CAP-1	CAP-2	E-CAP
10%–90%	3968	2944	3968	0.05837	1	0.0035
20%–80%	3968	2944	3968	0.05837	1	0.0046
30%–70%	3968	3328	4096	0.05837	0.724	0.0053
40%–60%	3968	3968	4096	0.05837	0.152	0.026
50%–50%	3968	3968	3968	0.05837	0.088	0.064
60%–40%	3968	3968	3968	0.05837	0.152	0.152
70%–30%	3968	3328	3328	0.05837	0.724	0.722
80%–20%	3968	2944	2944	0.05837	1	1
90%–10%	3968	2944	2944	0.05837	1	1

Table 4

Bandwidth allocated (in kbps) to station 3 and packet loss probability comparison (β level for station 3 equal to 0.625)

Best-Effort Traffic	Bandwidth allocated (kbps)			Packet loss probability		
TCP–UDP	CAP-1	CAP-2	E-CAP	CAP-1	CAP-2	E-CAP
10%–90%	2688	1536	2688	0.00258	1	0.0015
20%–80%	2688	2816	2688	0.00258	0.3305	0.0018
30%–70%	2688	2688	2688	0.00258	0.0120	0.0021
40%–60%	2688	2688	2688	0.00258	0.0083	0.0027
50%–50%	2688	2688	2688	0.00258	0.0073	0.0036
60%–40%	2688	2688	2816	0.00258	0.0083	0.0048
70%–30%	2688	2688	2816	0.00258	0.0120	0.01
80%–20%	2688	2816	2688	0.00258	0.3305	0.328
90%–10%	2688	1536	1536	0.00258	1	1

are referred to Station 3. The packet loss probability (for CAP-1, CAP-2 and E-CAP) is computed through: Eq. (4) for CAP-1; Eq. (4) and Eq. (23) for CAP-2; Eq. (4), Eq. (22) and Eq. (23) for E-CAP. The bandwidth allocated is the solution of Eq. (33) but the three algorithms (CAP-1, CAP-2 and E-CAP) use the different packet loss equations stated just above. Table 2 sets $\beta = 0.15625$. CAP-1 strategy is independent of the Internet traffic balance, because it does not separate TCP and UDP traffic. The estimated packet loss probability is equal for all the tests performed. The CAP-2 allocation scheme always assigns no bandwidth to station 3, "considered" too much faded by the allocation algorithm of CAP-2.

E-CAP–ABASC, distinguishing between TCP and UDP flows by using different buffers and models, assigns portions of bandwidth to the faded station and allows a satisfying performance also for Station 3. The packet loss probability is always 1 if UDP and TCP traffic are undistinguished. It is worth noting that P_{loss}^{UDP} value is independent of the number of sources, which is considered only within the bandwidth assigned to the UDP aggregate.

Table 3 sets the fading level to 0.3125. Without any distinction between UDP and TCP, CAP-1 performance does not change when the best-effort traffic balance is varied. It means that, without any distinction between TCP and UDP traffic, the bandwidth allocation is too rough. CAP-2 allows a minimum tuning but only using the characteristics of TCP traffic. With a proper model, it is possible to get a correct allocation and, consequently, achieve improved performance.

Similar considerations may be done for Table 4, reporting the $\beta = 0.625$ case, and Table 5, containing the results for $\beta = 0.8333$. Table 6 reports the clear sky test. The advantage of E–CAP–ABASC with respect to CAP–ABASC (both CAP-1 and CAP-2 cases) is evident in many cases both in terms of packet loss probability and in terms of "bandwidth saving":

- the performance in terms of packet loss probability is outstanding for E–CAP–ABASC;
- the overall bandwidth is distributed among the stations so that the "bandwidth saving" of the E–CAP–ABASC is not obtained at cost of Station 3, which is faded, but at cost of Stations 0, 1 and

Table 5

Bandwidth allocated (in kbps) to station 3 and packet loss probability comparison (β level for station 3 equal to 0.8333)

Best-Effort Traffic	Bandwidth allocated (kbps)			Packet loss probability		
TCP–UDP	CAP-1	CAP-2	E-CAP	CAP-1	CAP-2	E-CAP
10%–90%	2304	1152	2304	0.00189	1	0.0012
20%–80%	2304	2304	2304	0.00189	0.1041	0.0014
30%–70%	2304	2304	2304	0.00189	0.0084	0.0016
40%–60%	2304	2304	2304	0.00189	0.0058	0.0020
50%–50%	2304	2304	2304	0.00189	0.0053	0.0026
60%–40%	2304	2304	2304	0.00189	0.0058	0.0038
70%–30%	2304	2304	2304	0.00189	0.0084	0.0067
80%–20%	2304	2304	2304	0.00189	0.1041	0.1028
90%–10%	2304	1152	1152	0.00189	1	1

Table 6

Bandwidth allocated (in kbps) to station 3 and packet loss probability comparison (β level for station 3 equal to 1)

Best-Effort Traffic	Bandwidth allocated (kbps)			Packet loss probability		
TCP–UDP	CAP-1	CAP-2	E-CAP	CAP-1	CAP-2	E-CAP
10%–90%	2048	2048	2048	0.00166	1	0.0010
20%–80%	2048	2048	2048	0.00166	0.0381	0.0012
30%–70%	2048	2048	2048	0.00166	0.0070	0.0014
40%–60%	2048	2048	2048	0.00166	0.0051	0.0018
50%–50%	2048	2048	2048	0.00166	0.0046	0.0023
60%–40%	2048	2048	2048	0.00166	0.0051	0.0032
70%–30%	2048	2048	2048	0.00166	0.0070	0.0055
80%–20%	2048	2048	2048	0.00166	0.0381	0.0369
90%–10%	2048	2048	2048	0.00166	1	1

2, working in clear sky, so enhancing the performance of the network as a whole. This is not possible if CAP–ABASC is used.

It is worth noting that the bandwidth assignment performed by CAP-2 is independent of the relative percentage of TCP and UDP traffic: the "10% TCP 90% UDP" case originates the same bandwidth partitions of the opposite case, "90% TCP 10% UDP"; actually, the performance is "symmetric" with respect to the "50% TCP 50% UDP" case, where TCP and UDP receive the same bandwidth portions. That is the effect of aggregating IP traffic in the network without differentiating between UDP and TCP. The congestion control of TCP would reduce its rate in case of packet loss. E-CAP–ABASC algorithm uses this TCP feature, improving the bandwidth utilization and efficiency of the overall system.

6.2. *Overall bandwidth allocation and gain*

Figures 4, 5 and 6 contain a comparison of the overall bandwidth needed for all the four stations (the sum of the bandwidth necessary to each earth station) to get a packet loss probability lower than 0.001 (0.1%) for all the stations. Station 3 is again the only faded station. The fading value, differently from the previous case, it is not strictly linked to an FEC scheme, i.e., $\beta^{(3)}$ values are not taken from Table 1 but the following set of numbers is used: $(0.1, 0.2, \ldots, 0.9, 1)$. Results are then interpolated. The bandwidth allocated by CAP-1 is totally independent of the traffic balance, Figures from 4 to 6 show the bandwidth values only for CAP-2 and E-CAP. Both allocations are computed by solving the problem stated in Eq. (33), but using different packet loss probability expressions, as stated for Tables 2–6.

Three unbalance cases are considered: "90% TCP 10% UDP", "50% TCP 50% UDP", and "10% TCP 90% UDP". Concerning the "90% TCP 10% UDP" case (Fig. 4), the needed bandwidth is similar for the two compared mechanisms because the advantage of E–CAP–ABASC differentiation is limited by the small amount of bandwidth assigned to UDP, which generates an increase of the overall packet loss probability. The 90% of TCP sources assures the slight bandwidth gain appearing in Fig. 4. Figure 5 reports the "50%

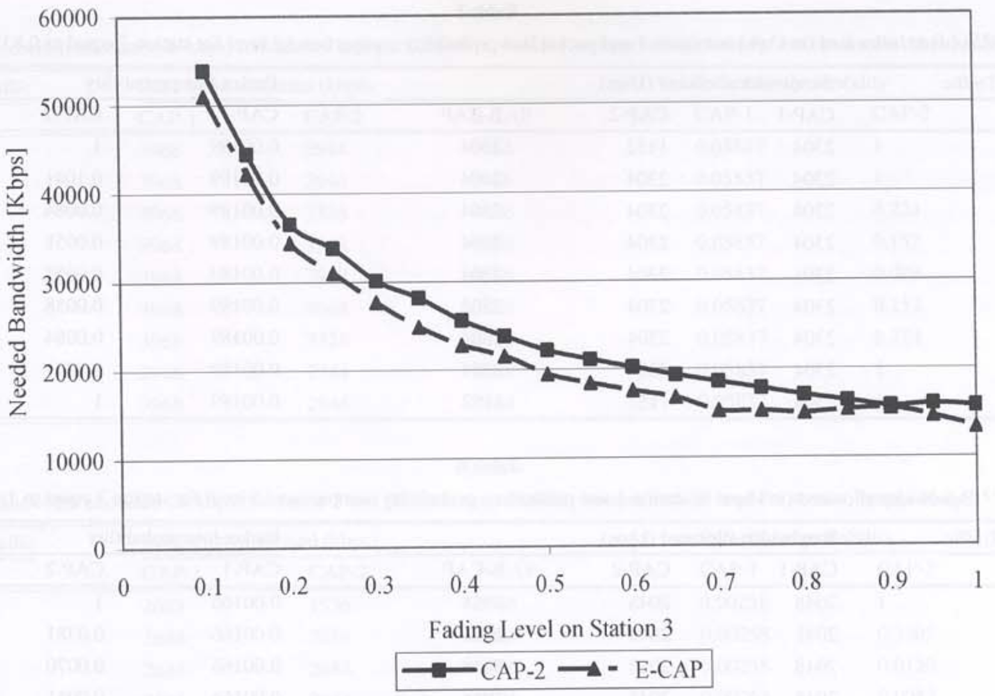

Fig. 4. Needed bandwidth (kbps) versus fading level of station 3 (90% TCP 10% UDP).

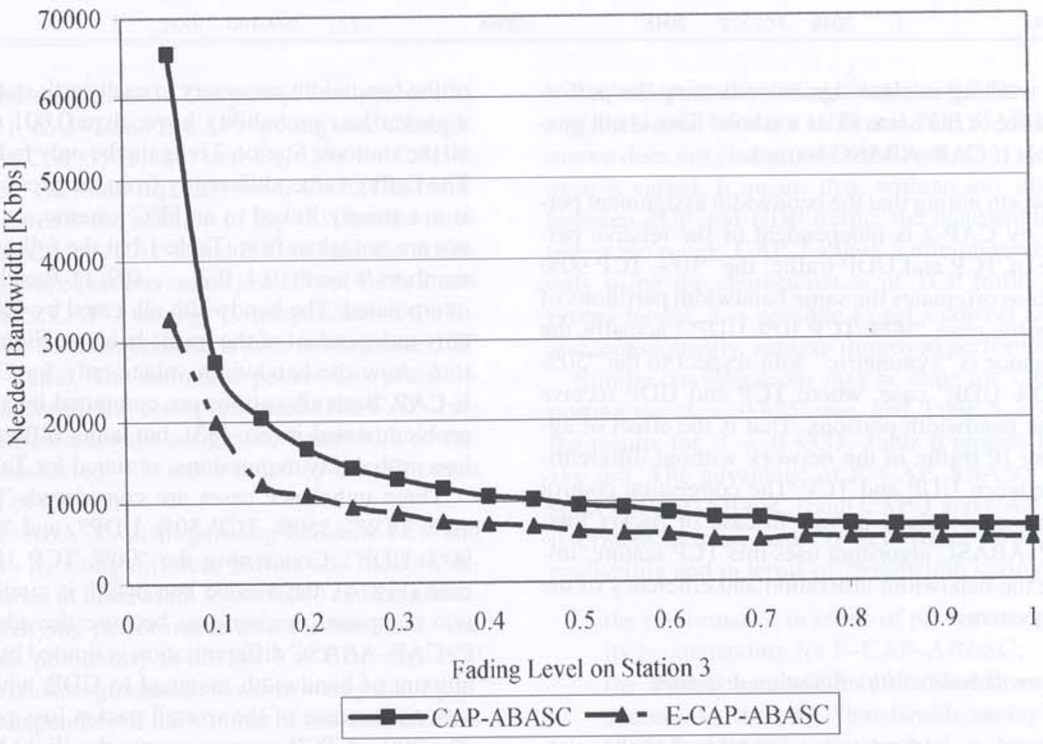

Fig. 5. Needed bandwidth (kbps) versus fading level of station 3 (50% TCP 50% UDP).

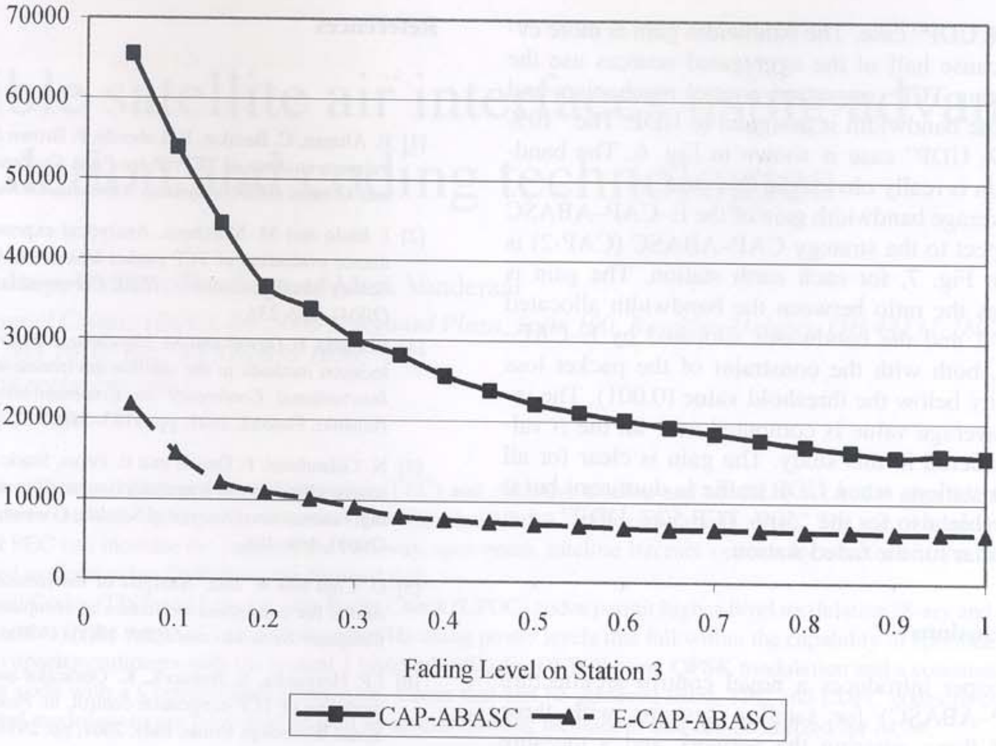

Fig. 6. Needed bandwidth (kbps) versus fading level of station 3 (10% TCP 90% UDP).

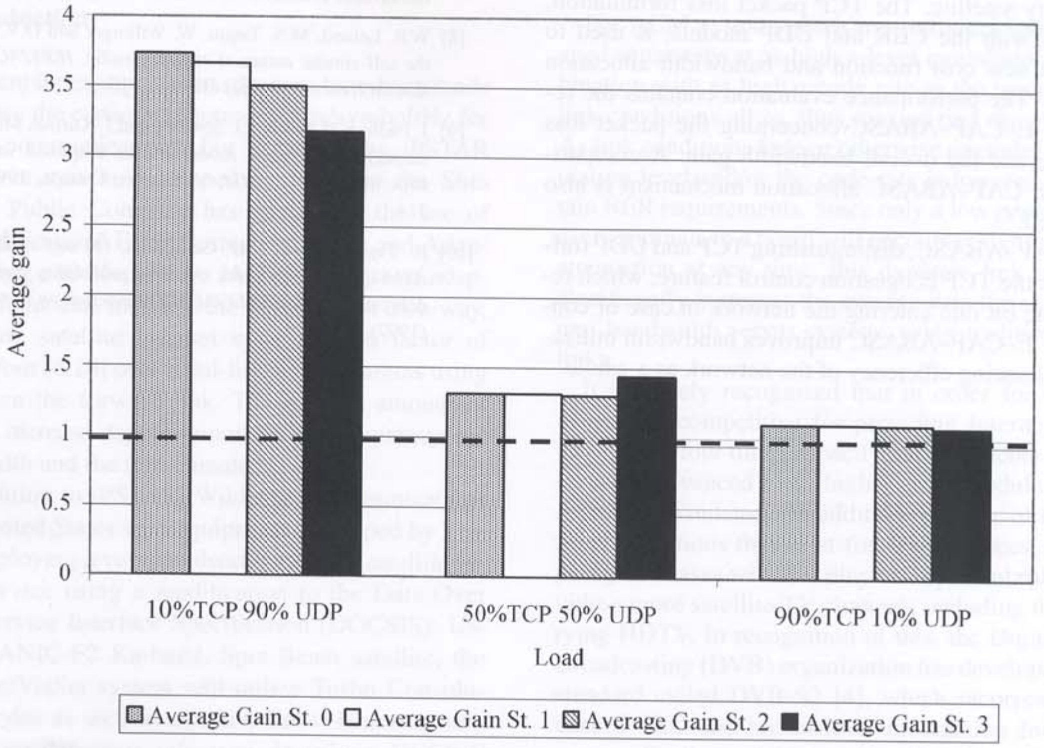

Fig. 7. Average bandwidth gain of E–CAP–ABASC with respect to CAP–ABASC.

TCP 50% UDP" case. The bandwidth gain is more evident because half of the aggregated sources use the rate-limiting TCP congestion control mechanism and 50% of the bandwidth is assigned to UDP. The "10% TCP 90% UDP" case is shown in Fig. 6. The bandwidth gain is really obvious in this case.

The average bandwidth gain of the E–CAP–ABASC with respect to the strategy CAP–ABASC (CAP-2) is shown in Fig. 7, for each earth station. The gain is defined as the ratio between the bandwidth allocated by CAP-2 and the bandwidth allocated by E–CAP–ABASC, both with the constraint of the packet loss probability below the threshold value (0.001). The indicated average value is computed over all the β values considered in this study. The gain is clear for all the earth stations when UDP traffic is dominant but it is noticeable also for the "50% TCP 50% UDP" case, in particular for the faded station.

7. Conclusions

The paper introduces a novel control architecture (E–CAP–ABASC) for satellite systems with three types of flows entering the network and a measure of the packet loss for TCP. The satellite network is composed of earth stations connected through a geostationary satellite. The TCP packet loss formulation, together with the CBR and UDP models, is used to derive a new cost function and bandwidth allocation scheme. The performance evaluation contains the results of E–CAP–ABASC concerning the packet loss measure and the overall bandwidth gain. A comparison with CAP–ABASC allocation mechanism is also presented.

E–CAP–ABASC, distinguishing TCP and UDP traffic, uses the TCP congestion control feature, which reduces the bit rate entering the network in case of congestion. E–CAP–ABASC improves bandwidth utilization, enhancing efficiency of the network as a whole.

References

[1] E. Altman, C. Barakat, E. Laborde, P. Brown and D. Collange, Fairness analysis of TCP/IP, in: *Proc. Conference on Decision and Control (ICDC)*, Sydney, Australia, 2000, pp. 231–242.

[2] I. Bisio and M. Marchese, Analytical expression and performance evaluation of TCP packet loss probability over geostationary satellite channels, *IEEE Communications Letter* **8**(4) (2004), 232–234.

[3] R. Bolla, F. Davoli and M. Marchese, Adaptive bandwidth allocation methods in the satellite environment, in: *Proc. IEEE International Conference on Communications (ICC 2001)*, Helsinki, Finland, 2001, pp. 3183–3190.

[4] N. Celandroni, F. Davoli and E. Ferro, Static and dynamic resource allocation in a multiservice satellite network with fading, *International Journal of Satellite Communications* **21**(4–5) (2003), 469–488.

[5] D. Chiu and R. Jain, Analysis of the increase/decrease algorithms for congestion avoidance in computer networks, *J. of Computer Networks and ISDN* **17**(1) (1989), 1–14.

[6] J.P. Hespanha, S. Bohacek, K. Obraczka and J. Lee, Hybrid modeling of TCP congestion control, in: *Proc. HSCC International Workshop*, Rome, Italy, 2001, pp. 291–304.

[7] T. Lakshman and U. Madhow, The performance of TCP/IP for networks with high bandwidth-delay products and random loss, *IEEE/ACM Transactions on Networking* **5**(3) (1997), 336–350.

[8] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, On the self-similar nature of ethernet traffic, *IEEE/ACM Transaction on Networking* **2**(1) (1994), 1–15.

[9] J. Padhye, V. Firoiu, D. Towsley and J. Kurose, Modeling TCP throughput: a simple model and its empirical validation, in: *Proc. ACM SIGCOMM*, Vancouver, Canada, 1998, pp. 303–314.

[10] B. Tsybakov and 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), 397–409.