

Providing Differentiated Service to TCP Flows Over Bandwidth on Demand Geostationary Satellite Networks

Merkourios Karaliopoulos, *Student Member, IEEE*, Rahim Tafazolli, *Member, IEEE*, and Barry G. Evans, *Senior Member, IEEE*

Abstract—The elasticity of transmission control protocol (TCP) traffic complicates attempts to provide performance guarantees to TCP flows. The existence of different types of networks and environments on the connections' paths only aggravates this problem. In this paper, simulation is the primary means for investigating the specific problem in the context of bandwidth on demand (BoD) geostationary satellite networks. Proposed transport-layer options and mechanisms for TCP performance enhancement, studied in the single connection case or without taking into account the media access control (MAC)-shared nature of the satellite link, are evaluated within a BoD-aware satellite simulation environment. Available capabilities at MAC layer, enabling the provision of differentiated service to TCP flows, are demonstrated and the conditions under which they perform efficiently are investigated. The BoD scheduling algorithm and the policy regarding spare capacity distribution are two MAC-layer mechanisms that appear to be complementary in this context; the former is effective at high levels of traffic load, whereas the latter drives the differentiation at low traffic load. When coupled with transport layer mechanisms they can form distinct bearer services over the satellite network that increase the differentiation robustness against the TCP bias against connections with long round-trip times. We also explore the use of analytical, fixed-point methods to predict the performance at transport level and link level. The applicability of the approach is mainly limited by the lack of analytical models accounting for prioritization mechanisms at the MAC layer and the nonuniform distribution of traffic load among satellite terminals.

Index Terms—Bandwidth on demand (BoD), fixed point, geosynchronous earth orbit (GEO) satellite networks, media access control (MAC), split connections, transmission control protocol (TCP).

I. INTRODUCTION

THE DOMINANCE of the Internet protocol (IP) within data networks seems today to be clearer than ever before. The expectations in the late 1980s/early 1990s that asynchronous transfer mode (ATM) would be the future network protocol have not materialized. Internet applications constitute the overwhelming majority of data applications running over wired data networks; at the same time the penetration of IP in wireless networks is increasing, embracing different access technologies.

Broadband satellite networks have been advertised as a significant part of the future global network infrastructure.

Manuscript received December 15, 2002; revised July 1, 2003 and September 20, 2003. This work was supported in part by the Hellenic State Scholarships Foundation (I.K.Y.).

The authors are with the Centre for Communication Systems Research, University of Surrey, Guildford, Surrey GU2 7XH, U.K. (e-mail: M.Karaliopoulos@surrey.ac.uk; R.Tafazolli@surrey.ac.uk; B.Evans@surrey.ac.uk).

Digital Object Identifier 10.1109/JSAC.2003.819983

With significant advantages related to their inherent broadcast capabilities, they embody a promising platform for the provision of medium and high bit rates to end-users. Three different transport technologies have been proposed in the last decade for these systems. ATM was central to the majority of initial system proposals, given the speculations for a future, all-ATM network at that time and the desire to achieve close integration with the core terrestrial networks. On the contrary, the second family of systems originates from the broadcasting community: digital broadcasting systems making use of the MPEG-2 transport services, with or without a satellite return channel, emerged as a promising alternative for the implementation of broadband satellite networks, one of their main advantages being the availability of standards (for example, see [1]). However, in the light of the established IP dominance in terrestrial networks, native-IP broadband satellite networks become more and more attractive, since they allow easier integration with the terrestrial network infrastructure. Such systems have been the subject of a number of European projects and the “raison d’être” of the ETSI Broadband Satellite Multimedia (BSM) working group [2].

This paper investigates one of the issues arising in the aforementioned systems: the interaction of the most popular Internet transport protocol, transmission control protocol (TCP), with bandwidth on demand (BoD) mechanisms envisaged for these networks. It is organized as follows. A brief review of TCP and BoD is given in Section II. The problem is then stated and possible analytical approaches that might be of use in addressing it are reviewed in Section III. Their main assumptions and weaknesses are highlighted before the formulation of the problem in fixed-point terms is presented. In Section IV, we describe the system under consideration and the simulation methodology adopted for its study. Different options available at transport and MAC/BoD level, capable of providing differentiated service to TCP flows, are demonstrated in Section V. One contribution of this work is the evaluation of proposed TCP modifications within a BoD-shared satellite environment. The actual performance gain obtained with these modifications is found to be less significant or even nonexistent, in comparison with results reported for the single TCP connection case or for multiple TCP connections, when BoD mechanisms are not taken into consideration. The system load range over which these options perform efficiently, as well as the “dynamic range” of the provided differentiation, are investigated. In parallel, the potential of the fixed-point approach to predict performance is discussed and its predictions—where the method is applicable—are compared against simulation results. Section VI concludes our study.

II. TCP AND BoD

A. Transmission Control Protocol (TCP)

TCP is the protocol supporting the most popular Internet applications, the ones responsible for the overwhelming majority of Internet traffic. TCP performs dynamic window-based flow control, assisted by the feedback carried in ACK packets. Under no congestion/loss, TCP can open its window more quickly, exponentially, during slow-start and less aggressively during congestion avoidance. The window value together with the round-trip time (RTT) experienced by the connection, i.e., the time elapsing between the transmission of a packet and its successful acknowledgment, dictate the actual connection rate. The maximum TCP send window value poses an upper limit to the amount of data that TCP can have “in-flight” with a pending acknowledgment. Its value is dictated by the receiver buffer availability (*receive window*) or may be set by the application running on top of TCP. The RTT is a control signal for TCP rate. Its decrease/increase is interpreted in respective increase/decrease of the TCP rate. When congestion occurs at the network, TCP may also experience packet loss. Duplicate acknowledgment or TCP retransmission timer timeouts are interpreted as congestion alarms. In these cases, TCP performs error recovery functions [3] and reduces its rate taking action that differs from implementation to implementation (for example, see [4] and [5]).

TCP experiences difficulties over paths featuring high propagation delays, particularly when these paths involve wireless, error-prone links. Given that this is almost always the case with satellite links (networks), considerable effort has been devoted in the last decade to the enhancement of TCP performance in such environments. The majority of the respective proposals are related to modifications of the end hosts’ protocol stacks: the sender and/or, to a lesser extent, the receiver implementation. Their main drawback is that the TCP sender cannot infer on its own whether the loss of a packet is due to congestion or corruption and act accordingly; it rather reduces its rate even when this is not necessary. Solutions involving network elements have, therefore, been proposed and can be broadly classified into two categories: In the first category, the network element—a router or a gateway at the border of wired and wireless network—may provide explicit feedback to the sender regarding the reason for a packet loss. In the second category, these intermediate elements may perform local error recovery functions or introduce other features, of varying visibility to the end hosts, which attempt to hide the impact of errors from the sender. For a comprehensive summary of these proposals the reader is referred to [6] and the references therein.

B. Satellite MAC Protocols and Bandwidth on Demand

Satellite MAC-layer protocols have been the subject of research for more than forty years. From the fixed assignment protocols of mid and late 1960s to the basic random access and combined reservation/random access protocols of the 1970s (see for example [7]–[9]) and the integrated voice and data protocols of the 1980s (for example, see [10]), MAC protocols have been continuously evolving adapting to the terrestrial network

technologies and the corresponding traffic pattern characteristics. The more recent emergence of multiservice networks led to multiclass protocols featuring a broader range of traffic handling capabilities in order to support most efficiently the increasing variety of these networks’ applications. In fact, the envisaged resource allocation mechanisms at MAC level are, to a large extent, independent of the overlaying network protocol: the standard capacity allocation mechanisms, formally described within the digital video broadcasting–return channel via satellite (DVB-RCS) standard [1] for example, may be used for the support of both the ATM-based and IP-based quality-of-service (QoS) frameworks ([11], [12]).

With broadband satellite systems also targeting residential users, BoD mechanisms are deemed necessary to accommodate the burstiness of TCP/IP traffic in a resource-friendly manner. Resources (e.g., time slots) may be allocated per-request, be reserved for variable time intervals or be accessed in random access mode. Spare, nonreserved capacity, may be distributed to terminals via free capacity allocation (FCA) mechanisms. The resource assignments are scheduled by a controller (scheduler) residing either on-board the satellite or at the ground. The scheduler may exercise different disciplines, consider priorities amongst traffic classes or/and even rely on traffic prediction for allocating resources. For a comparative review of satellite MAC protocols, the interested reader is referred to [13].

C. TCP and BoD Interaction

Little literature has been found on the interaction of TCP and BoD. The interaction of TCP with BoD mechanisms in the context of DVB-RCS networks is discussed in [14] and [15], using simplistic approximations for the resource allocation procedures described in the standard [1]. The authors in [14] consider a single TCP connection, while they model the volume-based dynamic capacity (VBDC) allocation and FCA modes of DVB-RCS are modeled as independent processes: return link slots become available to the terminal either deterministically (periodically) or following some authoritative distribution (exponential or uniform). In [15], it is the BoD-related delay (access delay) that is independently modeled: the access delays related to different DVB-RCS capacity allocation modes are one of the inputs to the simulation, i.e., chosen *a priori* irrespective of the link capacity and the number of connections that share the link.

A TCP-aware MAC implementation is proposed in [16]. The fundamental requirement in this approach is that the DAMA Controller, located either at the network control center (NCC) or on-board the satellite, can read protocol headers deeper in the MAC frame payload and predict capacity demand by different terminals following the per-connection TCP segment-ACK packet exchange. The main disadvantage of the scheme is that it increases the BoD controller complexity and raises additional concerns about its scalability, on top of the security related (i.e., compatibility with IPsec) objections.

It is anticipated that an actual BoD simulation environment is required for a more elaborate study of the TCP performance over BoD satellite networks; this is what we have attempted to provide in Sections IV and V. We demonstrate that the MAC

layer provides enough flexibility for the support of qualitative service differentiation for TCP flows over these networks via rather simple mechanisms that do not necessitate per-flow state at the MAC layer and/or the overhead of estimation/prediction related to [16]. But before proceeding with the simulation study, we briefly review analytical approaches to the study of TCP/IP networks and outline an analytical method for the study of TCP performance over satellite networks. We discuss its main limitations in addressing the service differentiation problem in the context of BoD networks, advocating, in the same time, the simulation approach taken in the rest of this paper.

III. ANALYTICAL TREATMENT

Analytical approaches are always favored in performance analysis particularly when they are accompanied by neat, simple-to-use formulas. Such formulas have been invoked extensively in problems related to telecommunication network analysis and, more generally, resource management. The dimensioning of telephone networks has relied heavily on the use of Erlang and Engset formulas; their extensions for multiuser, multiservice networks, as well as the broad use of the effective bandwidth concept in the study of ATM networks, advocate further the value of these approaches.

TCP traffic, however, does not lend itself to these analytical methods. The lack of an inherent rate due to the adaptive nature of TCP, renders techniques that have been heavily used in the context of telephone and ATM network analysis inappropriate.

In the *processor sharing (PS)-based models* introduced in [17]–[19], classical queueing theory was invoked in order to study the interaction of adaptive TCP/IP flows with the network. The common drawback of these *flow-level* models is that they cannot provide results for the packet-level dynamics of the traffic (expressed as buffer queues, losses etc.). The most attractive property of PS-based models is their *insensitivity property*: the model outcome is not dependent on the actual distribution of connection arrivals or transferred file sizes but only on their mean. This allows generalizing solutions derived under Markovian theory to the more realistic cases of power-tail distributions reported in literature for TCP traffic features (see for example [20] and references therein).

The processor sharing-based models are an abstraction of the TCP-controlled resource sharing and ignore the actual details of the protocol. On the contrary, *fixed-point methods* feature separate, explicit models for the TCP source and the resource (network links, buffers) that are coupled with each other: the TCP send rate is an explicit function of the loss and delay experienced by the connection in the network nodes and links [21]–[23], which are in turn a function of the TCP send rates. Classical queueing models such as the M/D/1/K [24], the M/M/1/K model [25] and some variation of it in [26], earlier proposed in [27], as well as active queue management (AQM) models [28], [29] have been proposed for modeling the resource.

These formulas for the TCP send rate already introduce some inaccuracy: all three models ignore aspects of TCP congestion control such as the fast retransmit algorithm and the fast recovery phase, while [21], [22] do not consider the slow-start phase at all. Likewise, the resource models in [24]–[27] fall back

on the Poisson assumption, this time at packet-level, in order to take advantage of well-known queueing models. However, the packet arrival process for TCP/IP traffic is largely different than a point Poisson process, even at high aggregation levels [30].

A. Fixed-Point Approximation

In the following, we outline a fixed-point method for predicting TCP performance in geostationary satellite networks. In doing so, we adopt the vector representation introduced by Roughan *et al.* in [24]. We generalize their formulation to cater for networks including asymmetric, MAC-shared links and consider its application to the case of BoD satellite networks.

1) *Generic Formulation*: Let L be the set of network links and R the set of routes (paths) of the TCP connections, corresponding to subsets r_i of L . If M and N are the cardinalities of the two sets, i.e., $M = |L|$ and $N = |R|$, then two $N \times M$ routing matrices $\bar{\mathbf{A}}_F$ and $\bar{\mathbf{A}}_R$ can be defined, one for the forward path of the connection and one for the reverse (ACK packets path). Their elements a_{ij}^F, a_{ij}^R equal unity when the link $j, j \in L$ is included in the forward/reverse connection path i , respectively, $i \in R$, and 0, otherwise

$$a_{ij}^{F,R} = \begin{cases} 1, & \text{if } j \in r_i \\ 0, & \text{if } j \notin r_i \end{cases} \quad (1)$$

Let also $\mathbf{c} = (c_j)$ denote the capacities of the network links and $\mathbf{d} = (d_j)$ the one-way propagation delays, related to them. Then, the vector of the RTTs of the N paths $\mathbf{t} = (t_i)$ can be written as

$$\mathbf{t} = (\bar{\mathbf{A}}_F + \bar{\mathbf{A}}_R) \cdot (\mathbf{d} + \mathbf{d}\mathbf{q}) = \mathbf{t}_p + (\bar{\mathbf{A}}_F + \bar{\mathbf{A}}_R) \cdot \mathbf{d}\mathbf{q} \quad (2)$$

$\mathbf{d}\mathbf{q} = (dq_j)$ are the queueing delays at the output buffers of the links and $\mathbf{t}_p = (t_i^p)$ is the component of the total path delay that is related to the link propagation delays.

If n_i is the number of connections sharing the same path i , w_i the average window sizes (at path level) and $m_i^{\text{TCP}}, m_i^{\text{ACK}}$ the TCP maximum segment size and the ACK packet size, respectively, in path i , the total send rate in the forward and the reverse path are given by

$$\mathbf{s}_F = \mathbf{n} \cdot \mathbf{SR}(\mathbf{W}^{\max}, \mathbf{m}_{\text{TCP}}, \mathbf{p}, \mathbf{t}) \quad (3)$$

$$\mathbf{s}_R = \mathbf{n} \cdot \mathbf{SR}(\mathbf{W}^{\max}, \mathbf{m}_{\text{TCP}}, \mathbf{p}, \mathbf{t}) \cdot (1 - \mathbf{p}) \cdot \mathbf{u} \cdot \frac{\mathbf{m}_{\text{ACK}}}{\mathbf{m}_{\text{TCP}}} \quad (4)$$

The notations $x \cdot y$ and $x./y$, where x and y are vectors, denote element-wise multiplication/division, respectively. The vector \mathbf{p} is defined in (9), \mathbf{W}^{\max} is the maximum TCP send window (see Section II-A), while \mathbf{u} accounts for the reverse path load reduction due to delayed ACKs, $0.5 \leq u_i \leq 1$, $i \in R$.

$SR_i(W_i^{\max}, m_i^{\text{TCP}}, p_i, t_i)$ yields the mean send-rate of the connections on path i in segments/s, according to any of the formulas in [21]–[23].

The offered load upon the network queues¹ is then written

$$\mathbf{b} = \mathbf{s}_F^T \cdot \bar{\mathbf{V}}_F + \mathbf{s}_R^T \cdot \bar{\mathbf{V}}_R \quad (5)$$

¹In each network node there is one-to-one correspondence between queues and output links.

where the elements of the $N \times M$ matrices $\mathbf{V}_F, \mathbf{V}_R$ are given by

$$v_{ij}^{F,R} = \begin{cases} \prod_{j' \in r_i: \text{pos}(j',i) < \text{pos}(j,i)} (1 - q_{j'}), & \text{if } j \in r_i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

In (6), $\text{pos}(j, i)$ yields the position of the queue (link) j within path i and q_j is the aggregate loss probability due to both congestion and link errors over the respective link.

The assumption is that

$$q_k = 0, \quad \text{pos}(k, i) = 1 \quad \text{for } k : \text{pos}(k, i) < \text{pos}(m, i) \\ \forall m \in r_i, \quad m \neq k$$

in other words, there is no TCP data loss at the sender parts of the connections at the end hosts, whose queues are the first ones on the connections' paths.

What is then needed is to develop efficient resource models, i.e., models for the actual packet loss or delay that are experienced at the buffers of the different links. In general, losses and delays may incur at any link on the connection's path

$$dq_j = f_d^j(b_j; c_j) \quad (7)$$

and

$$q_j = q_j^c + q_j^l = f_l^j(b_j; c_j) + q_j^l \quad (8)$$

where f_d^j and f_l^j are functions (formulae) yielding the delay/loss suffered at each link j of capacity c_j , when load b_j is offered to it. The term q_j^l is the loss probability related to link errors. In order to make the analysis feasible, the assumption of independent losses at each link is invoked, so that, if $\mathbf{1}$ is the vector of size N whose components are all equal to 1, the path losses experienced by the TCP senders on the connection paths may be written

$$\mathbf{p} = \mathbf{1} - \overline{\mathbf{A}_F} \cdot (\mathbf{1} - \mathbf{q}). \quad (9)$$

The TCP send rate is then dictated by the queueing losses (q_j^c) and the link errors (q_j^l) encountered in each link on the connection path and the limitations set by the buffer availability at the TCP receiver peer or the application \mathbf{W}^{\max} .

The medium access controlled links need particular attention in order to be accommodated in this framework: with regard to queueing delay/loss they can be regarded as the distributed equivalent of a single conventional multiplexer, i.e., as a single distributed switch. On the contrary, with regard to link errors a shared link corresponds to many point-to-point links due to the physical separation of the sender-receiver pairs.

One way to expand the described vector representation so that it can accommodate MAC controlled links is to slightly

modify (6), (8), and (9). For the MAC-shared link j^s , (8) may be rewritten as

$$q_{j^s} = q_{j^s}^c = f_l^{j^s}(b; c) \quad (10)$$

while under realistically small loss rates

$$\mathbf{p} = \mathbf{1} - \overline{\mathbf{A}_F} \cdot (\mathbf{1} - \mathbf{q}) + \mathbf{p}^l \quad (11)$$

and the factors accounting for the successive traffic thinning on its path are written as shown in the equation at the bottom of the page. In other words, the link error probabilities at the shared link are separated from the link congestion losses and are added at connection (i.e., path) level as separate terms. These adaptations are applicable when there is a single MAC-shared link on the connections' paths or, equivalently, when other MAC-shared links (e.g., Ethernet network at the TCP server/client side) do not introduce link errors.

2) *Application to Bottleneck BoD-Shared Satellite Networks:* We consider now our problem as a special case of this generic formulation. We assume that the bottleneck link is the shared satellite link, neglecting the delays and losses incurred at the other links; the control signal for TCP in our context is the delay experienced at the MAC level of the shared satellite link. In comparison with the generic problem formulation

$$dq_j = dq_k \cdot \delta_{jk} \quad (12)$$

$$\overline{\mathbf{A}_F}(:, k) = \overline{\mathbf{A}_R}(:, k) = \mathbf{1} \quad (13)$$

where k is the index of the shared satellite uplink and δ_{jk} is the Kronecker delta.

Replacing index k with sat, (7) yields

$$dq_{\text{sat}} = f_d^{\text{sat}} \cdot \left(\frac{\sum_{i=1}^N SR^i (W_i^{\max}, m_i^{\text{TCP}}, p_i, t_i) \cdot (1 + (1 - p_i) \cdot u_i \cdot m_i^{\text{ACK}})}{t_i^{\text{PROP}} + 2 \cdot dq_{\text{sat}}}, c_{\text{sat}} \right) \quad (14)$$

i.e., we derive an equation of the form $x = f(x)$ for dq_{sat} .

The missing component is the function $f_d^{\text{sat}}(b_{\text{sat}}, c_{\text{sat}})$ that relates the MAC access delay dq_{sat} , at the shared satellite link of bandwidth c_{sat} , to the offered load to the link b_{sat} . In this context, the most relevant analytical study we are aware of is the analysis provided in [31] for a member of the combined free/demand assignment multiple access (CFDAMA) protocol family, the one with preassigned request slots. The system considered therein combines the two resource allocation mechanisms that are considered subsequently in our simulations (Section IV), i.e., demand and free assignment. Nevertheless, certain assumptions of the analysis question its applicability.

$$v_{ij}^{F,R} = \begin{cases} \prod_{j' \in r_i: \text{pos}(j',i) < \text{pos}(j,i)} (1 - q_{j'}), & \text{if } j \in r_i \cap \text{pos}(j, i) \leq \text{pos}(j^s, i) \\ \prod_{j' \in r_i: \text{pos}(j',i) < \text{pos}(j,i)} (1 - q_{j'}) - p_i^l, & \text{if } j \in r_i \cap \text{pos}(j, i) > \text{pos}(j^s, i) \\ 0, & \text{otherwise} \end{cases}$$

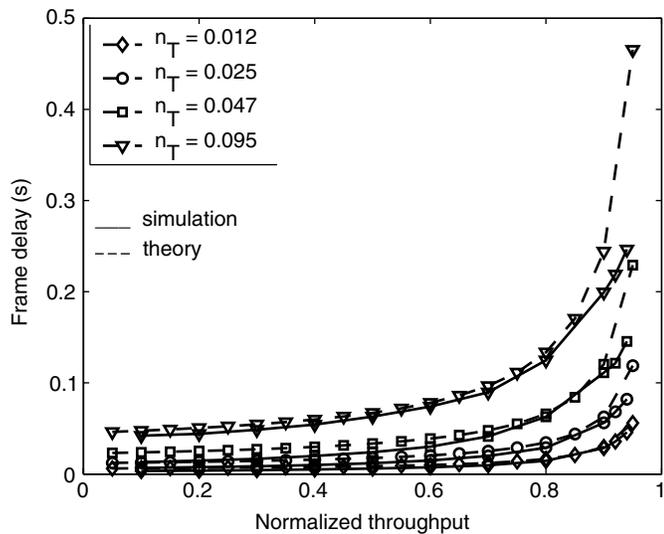


Fig. 1. Comparison of the analytical approximation in [31] against simulation results for the MAC frame delay under Poisson traffic and variable normalized terminal population.

First of all, the analysis is provided for Poisson traffic. Second, the request allocation period is higher than the one-hop delay (for the case of on-board BoD controller), setting a lower bound to the resource allocation period. Third, the analysis only addresses the case of load equally distributed among the terminals sharing the link. Fourth, it is a no-loss model, i.e., the MAC buffering space is assumed infinite.

We first examine whether this CFDAMA analysis can be applied to the system considered in the subsequent simulations (see Section IV). In other words, we apply Poisson traffic to our BoD system and compare the MAC delays with the ones provided by the respective analysis. The number of terminals considered is small enough, so that we can take advantage of the geometrical approximation provided in [31] for small number of terminals. Fig. 1 depicts the mean frame delays predicted by the analysis and the ones obtained by simulation, when the simulated BoD system is exposed to Poisson traffic. Results are shown for various values of the normalized terminal population n_T [31]

$$n_T = \frac{N_T \cdot \tau}{\text{RTD}} \quad (15)$$

where N_T is the number of terminals sharing the link, τ is the slot size (duration), and RTD the round-trip delay related to the BoD procedure (including propagation and processing delays). The agreement between the geometric approximation and the simulation results is tighter at moderate and high loads, although the former yields pessimistic predictions about the expected frame delay close to the saturation area.

In subsequent comparisons of simulation results against fixed-point based predictions in Section V-A, we will use the geometric approximation as the basis for the BoD induced delay. Given the aforementioned limitations of the method in addressing the service differentiation mechanisms targeted in this work, we subsequently resort to simulation as the primary means for our study.

IV. SIMULATION SETUP AND METHODOLOGY

For our simulations, we have chosen *ns2* [33], a discrete event simulator dedicated to the research of IP networks and protocols, widely used within the Internet community. We extended the basic satellite code of *ns2* with BoD features [34] whilst, at the same time, we were able to take advantage of the built-in packet-level simulator functionality.

A. Simulator Parameters and Assumptions

The BoD scheme under consideration within this paper draws heavily on the procedure described in [35] and some terminology is adopted from it. Periodically, and as long as new packets arrive at their queues, the *BoD entities*, physically located in satellite terminals and serving one or more data flows, submit requests to a *BoD controller* residing on-board the satellite. The controller takes requests into consideration, builds the burst time plan (BTP) and broadcasts it to the BoD entities. Both functions are repeated with a period equal to an integer number n_s of time-division multiple-access (TDMA) frames (*resource allocation period*). Requests may be submitted in preassigned slots, piggybacked on data slots, or compete in random access manner (i.e., slotted/unslotted Aloha). Within the context of this study, we consider only the first option. The BTP corresponding to requests submitted within a certain allocation period is activated by the entities after an integer number L of resource allocation periods (*system response time*), allowing for the propagation and processing delays at the terminals and the Controller. Requests that cannot be satisfied in the corresponding allocation period are stored and served with priority in the next resource allocation period. Spare capacity (*free slots*) is distributed prioritizing entities that have not submitted a request in the respective resource allocation period on a round-robin basis.

The algorithms for estimating the requests are borrowed from [35]. If $q(k)$ are the queued packets at the terminal queue at the beginning of the k th resource allocation period, then the request submitted to the BoD controller is for $r(k)$ slots shown in (16), at the bottom of the page. In (16), $\lceil x \rceil^+$ denotes the smallest positive integer, including zero, that is greater than x , $a(k)$ are the number of slots per frame allocated to the BoD entity for the next n_s frames, and $w(k)$ is the estimate of the *owed* slots by the controller to the entity, because in one or more of the previous resource allocation periods, the controller assigned fewer or even none of the slots requested. Apparently, $w(k) \geq 0$. In

$$r(k) = \left\lceil \frac{q(k) - n_s \cdot a(k) - n_s \cdot \sum_{j=1}^{L-1} r(k-L+j) - n_s \cdot w(k)}{n_s} \right\rceil^+ \quad (16)$$

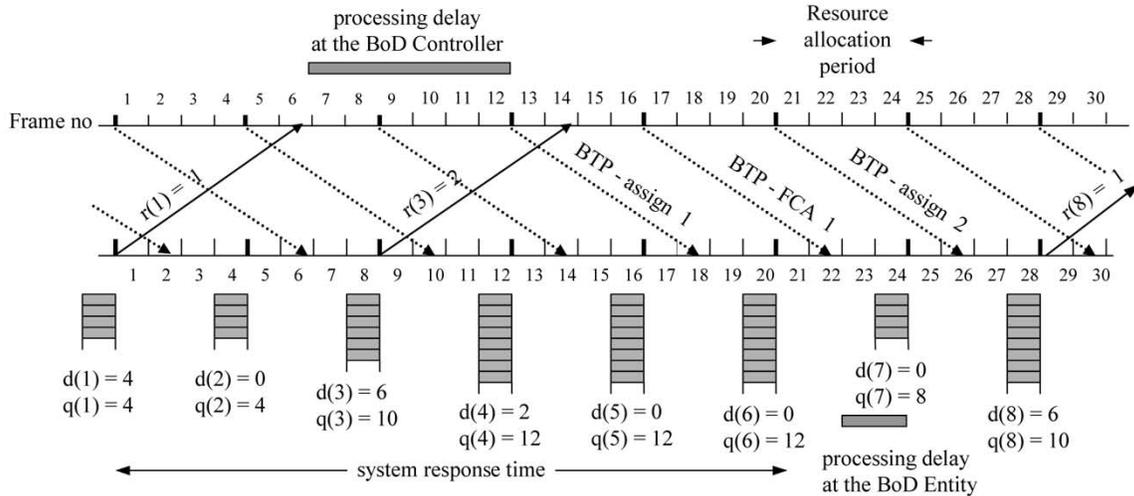


Fig. 2. Evolution of BoD procedure with time and example exchange of messages between BoD controller and BoD entity $-n_s = 4, L = 5$.

effect, the process resembles the VBDC allocation mode, as defined in [1].

An example of the BoD scheme operation is provided in Fig. 2. The snapshot shows the estimated requests, the assignments of the controller, and the queue size at the beginning of the resource allocation periods; $d(k)$ denote the number of new frames that arrive at the entity queue within the $(k - 1)$ th resource allocation period. In this example, the link is lightly loaded so that the BoD controller can satisfy the requests of the entity, namely $w(k) = 0 \quad k \in [1 \dots 8]$, and there is spare capacity that is assigned to the entity in FCA mode.

In our simulations, we mainly consider terminals with low traffic aggregation. A variable number of TCP connections compete for the link bandwidth. Connections are subject to BoD mechanisms in both the forward and reverse direction (ACK packets), i.e., we consider a mesh system architecture without notion of forward and return link. In the majority of simulations, there is one-to-one correspondence between TCP connections and BoD entities, namely each entity serves the TCP sender and TCP receiver parts of two unidirectional TCP connections. Further simulation results with more connections per BoD entity are provided in Section V-E. Simulations were run for satellite link capacities of 512/1024/2048 kb/s yielding 32/64/128 16 kb/s slots per frame, respectively, for frame duration equal to 24 ms. The resource allocation period was set to 96 ms (four frames).

TCP segments are fragmented implicitly [36] at the link-layer into 48-byte payload data frames. We select moderate values for the maximum TCP window size and let the link buffer be large enough to handle the incoming TCP bursts, in effect avoiding losses at the link buffers. In short, we are concerned with a lossless system. In such a system, TCP adapts to the delays occurring at the MAC queues within the satellite network.

We have chosen the TCP NewReno [5] variant for our simulation experiments. Since no losses occur, this choice does not affect obtained results; most of the available TCP implementations (e.g., Reno, NewReno, SACK) differ in the way they treat losses and suggest different action during the fast retransmit/fast recovery phase. The granularity of the TCP timer is 100 ms and the timestamps option is activated for all simulations. The default value for the maximum TCP send window (hereafter de-

noted as $rwnd$) is 8 kB. Further values, when used, are explicitly stated. The satellite links are considered error-free. This assumption and the impact of link errors in general are discussed further in Section V-E.

B. Approach-Performance Metrics

As a metric for end user satisfaction, we adopt the mean TCP connection throughput, in line with the metrics considered in [18] within the broader context of elastic traffic.² Whereas real-time services like packet voice, audio or video streams introduce packet-level requirements, provision of better than best-effort service to TCP flows is mostly related to the provision of some minimum rate to the flow. This is the case for the most popular applications running on top of TCP (Web browsing, the increasingly popular peer-to-peer applications, FTP).

Connections are initiated with a random phasing and are run for time intervals of 60–100 s. The connection throughput is measured and averaged after the first quarter of the connection time has elapsed. At the system level, we mainly plot the achieved utilization of the satellite link. Each point in the graphs is the same average of 10–20 simulation runs.

V. RESULTS AND DISCUSSION

A. Transport Layer Differentiation

The TCP includes a number of variables. By solely selecting appropriate values for these variables it is already possible to provide advantage to some connections over others. Likewise, differentiated performance can be obtained if TCP options that enhance the protocol performance, such as those described in [6] and—in a more generic context—in [38], are selectively activated for a subset of competing connections.

1) *TCP Receive Window*: An exhaustive enumeration of all possible variables and options that could provide transport layer service differentiation is certainly not possible within the space

²The focus of the study is subsequently on persistent connections. For non-persistent, short-lived TCP connections, the impact of the slow-start phase of the TCP connection becomes important. We discuss some of the slow-start implications in [37].

limitations of this paper. Instead, we focus our attention on one of the settings at TCP level that can have a significant impact on the protocol performance, the receive window (*rwnd*). There are two good reasons for doing so.

First, it is interesting to evaluate the performance gain achieved via increasing *rwnd* in BoD-shared satellite networks, since it is widely suggested as one of the necessary protocol modifications/adaptations for enhancing TCP performance over satellite links. Second, its impact can be captured by the available analytical TCP models [21]–[23] allowing us to test the predictions of the analytical method presented in Section III-A. On the contrary, these models cannot cater for the microscopic details of the wide variety of options that have been proposed for TCP.

2) *Split-TCP Connections*: The second transport-layer mechanism we investigate is the split-TCP mechanism. A proxy at the border between the terrestrial and the satellite network intercepts the connection set-up (SYN) and termination (FIN) packets and initiates another TCP connection over the satellite link, creating the respective state. When the other end of the satellite connection is another gateway, a third connection has to be setup for the last portion of the connection, from the second gateway to the end user. The split-TCP mechanism is only one of the potential features of TCP performance enhancing proxies (TCPPEPs). These solutions are viewed with skepticism within the IETF community, since their deployment within the Internet may contrast on a number of occasions with the end-to-end design principle dominating the latter [39]. Concerns regarding reliability and security aspects of TCPPEPs [40] have not prevented the satellite operators from deploying these mechanisms and explore countermeasures for their inefficiencies [41].

The advantages of split connections have been reported in numerous studies: split-TCP yields a performance gain that increases with errors over the satellite link and level of congestion at the terrestrial portion of the end-to-end connection, at the expense of increased buffering requirements at the gateway nodes, hosting the TCPPEP [42]. Nevertheless, the split-TCP mechanism has been mostly studied in the context of a single TCP connection. Few are the studies that considered more than one connection and even fewer those that evaluated the mechanism within a MAC-shared satellite environment.

3) *Simulation Results and Comparison With Analysis*: To allow the fixed-point method to account for the split-TCP mechanism, the elements a_{ij}^F , a_{ij}^R of the new routing matrices $\bar{\mathbf{A}}_F$, and $\bar{\mathbf{A}}_R$ can be defined as

$$\mathbf{A}_F(:, j) = \mathbf{A}_R(:, j) = \delta_{jk} \cdot \mathbf{1} \quad (17)$$

where k is the index of the satellite link and the elements w_i and W_i^{\max} are replaced by the respective values in the satellite portion of the connection. The assumption behind this modification is that the satellite component of the connection determines the actual connection throughput. In general, its validity is dependent on the relative values of the actual throughputs achieved by the individual connection components—hence, on all those factors affecting the throughput (RTT, delay/losses, enabled options in each of the connection components). It holds as far as

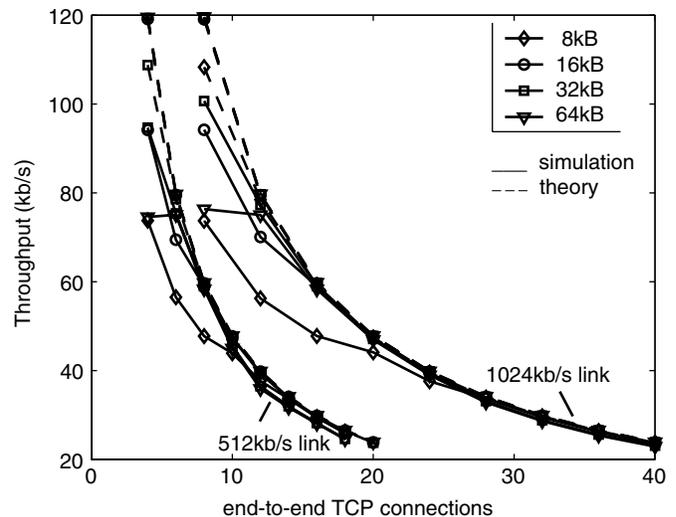


Fig. 3. Throughputs achieved by end-to-end connections for different *rwnd* values and comparison with the predictions of the fixed-point method.

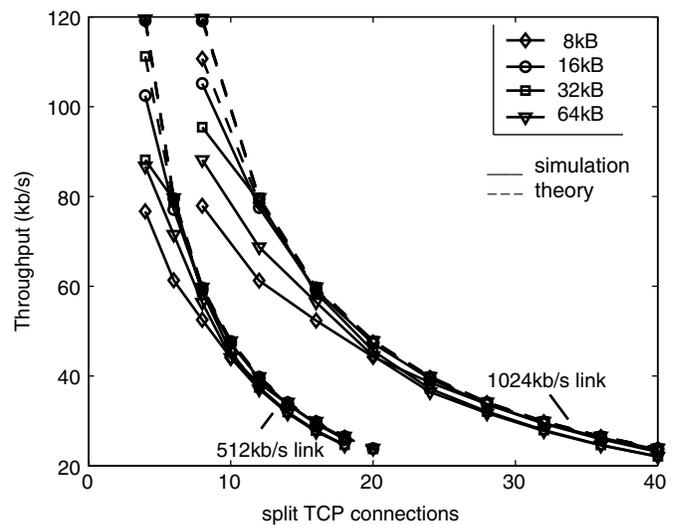


Fig. 4. Throughputs achieved by split TCP connections for different *rwnd* values of the satellite component of the connection and comparison with the predictions of the fixed-point method.

the throughput of the terrestrial component of the split connection is at least such that it does not let the satellite component starve [42].

The throughputs achieved by variable numbers of end-to-end and split connections are shown in Figs. 3 and 4, respectively. Split connections are more aggressive for the same *rwnd* values and saturate the link more quickly. Given our assumption for an error-free satellite link and the fact that the RTTs of the terrestrial portions of the connections are assumed to be small for these runs (~ 10 ms), the performance enhancement should be regarded as a lower bound on the gain that can be achieved over end-to-end connections. However, the impact of *rwnd* is similar in both cases: as *rwnd* increases the link utilization reaches its upper limit (saturation area) in the presence of fewer connections (Figs. 5 and 6) and then the throughput obtained by them is the same, irrespective of the *rwnd* value. In general, performance enhancement may be achieved as far as there is some

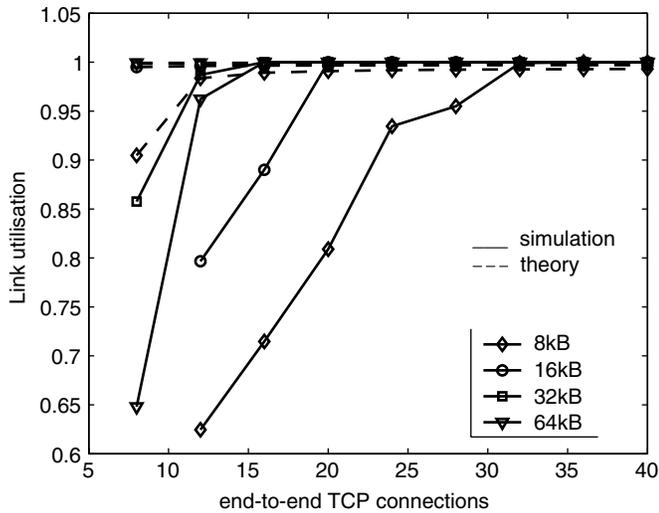


Fig. 5. Link utilization for different *rwnd* values of the end-to-end TCP connections and comparison with the fixed-point method predictions.

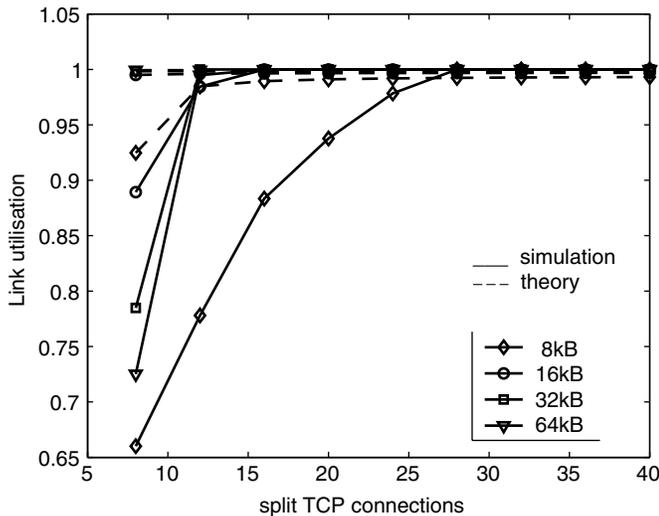


Fig. 6. Link utilization for different *rwnd* values of the satellite components of the split TCP the connections and comparison with the fixed-point method predictions.

spare capacity, not utilized by the connections because they are *rwnd-limited*. One may introduce the notion of the path effective bandwidth product, as the product of the path round-trip delay times the bandwidth available to a single connection **given** the presence of other connections sharing the link. As long as this is higher than the individual connection *rwnd*, performance improvement is possible via *rwnd* increase. Otherwise, the connections are *capacity-limited*; when this is the case, increase of the *rwnd* value only leads to accumulation of TCP segments into the link buffers of the satellite gateway, i.e., buffering is transferred from end hosts to the network.

Even worse, increase of the *rwnd* value does not deterministically increase the mean connection throughput, even when there is spare capacity to take advantage of. For higher values of *rwnd*, connections may experience timeouts with a severe impact on the individual connection throughput and the utilization (e.g., for *rwnd* equal to 32/64 kB in Figs. 3–6).

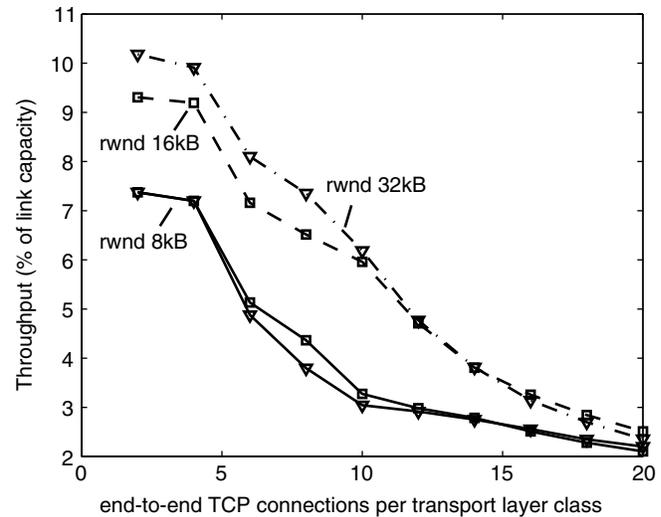


Fig. 7. Throughput of two classes of end-to-end TCP connections—differentiation option: *rwnd*, equal number of connections per class.

Figs. 3–6 also compare the predictions made by the fixed-point method against simulation results. The analysis clearly overestimates the achieved throughput and the subsequently achieved link utilization for low number of connections. In higher load values, the agreement is closer. The noticed deviation is mainly due to the burstiness of the TCP traffic. The delay values dictated by the solution of the fixed-point equations correspond to higher values of utilization under the more uniform Poisson traffic, for which the MAC resource model is derived. On the contrary, TCP segments arrive in bursts; the link remains underutilized over some intervals, while it is saturated over others, resulting in increase of the average access delay.

4) *Competing End-to-End and Split Connections*: In experimenting with the *rwnd* parameter, we separated the connections sharing the link into two transport-layer classes with regard to the *rwnd* setting. The *rwnd* value for the one class was set to 8 kB, whereas the *rwnd* for the second was set to 16 kB and 32 kB in two different experiment sets. Fig. 7 suggests that the higher *rwnd* provides a relative advantage that becomes maximum at medium-to-high loads, where the available resource is less than what the connections can attain within the limitations of their *rwnd*, but decreases rapidly when the link becomes saturated. Moreover, at these load levels doubling the *rwnd* from 16 to 32 kB does not benefit the split connections, since the additional in-flight data are accumulated at the MAC buffers. Note that increase of *rwnd* for end-to-end connections also implies higher congestion risk at the terrestrial links traversed by the connections.

With the split-TCP mechanism we followed a different approach. We considered two cases, each one featuring a fixed number of competing connections corresponding to different levels of system load: in the first, a 1.024 Mb/s link BoD-shared link is subject to the load generated by 16 connections, whereas in the second, the link was driven to saturation by a mix of 32 connections. Within each one of the two mixes we varied the proportion of each type of connections (end-to-end versus split) for different *rwnd* settings for the satellite component of the connection.

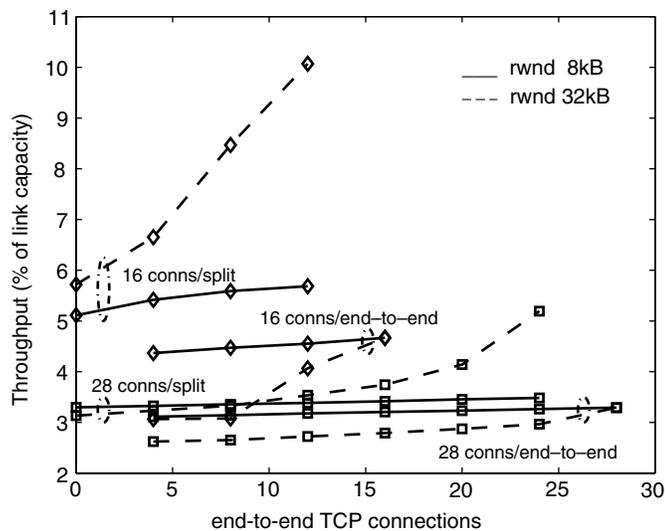


Fig. 8. Throughput of end-to-end and split connections versus number of end-to-end connections in the mix—the *rwnd* refers to the satellite component of the split connections.

In both cases, split connections achieve better performance and in both cases the throughput achieved—in absolute numbers—from all connections increases with decrease of the portion of split connections in the mix (Fig. 8). The lag in performance for end-to-end connections is larger at low load, since split connections make better use of free capacity as well, and increases with higher *rwnd* values of the satellite component of the connection. For higher load, when spare capacity is almost extinct and the connections rely only on their requests to obtain capacity, the performance lag is less but clear. In this case, the performance gain is only due to the inherent dynamics of split connections (smaller loop). Again, these results can be regarded as a lower bound on the achievable performance enhancement. The gain of the split approach is higher if link errors are met or the terrestrial portions of connections become longer.

B. MAC-Layer Differentiation

1) *Priority Scheduling*: Prioritization of traffic may also take place at the MAC/BoD level. The satellite link differs from a conventional time-multiplexing link in that packets are distributed over the terminal queues rather than being buffered at a single point. Multiplexing still takes place—over the air this time—and access to the medium is arbitrated by the scheduler (BoD controller), whose function resembles that of schedulers used in wired networks.

In our paper, we experimented with two types of schedulers: a first-come first-served (FCFS) scheduler, which treats all requests in the same, best effort (BE) manner and a priority-aware scheduler. The latter discriminates between two MAC classes and serves requests of the nonpriority class only when and if all requests of the priority class have been satisfied. Among connections of the same class, slots are distributed on a round-robin basis.

2) *FCA Policy*: A second way to allow for different treatment of connections at the access level is related to the policy imposed on the distribution of the spare capacity. An example

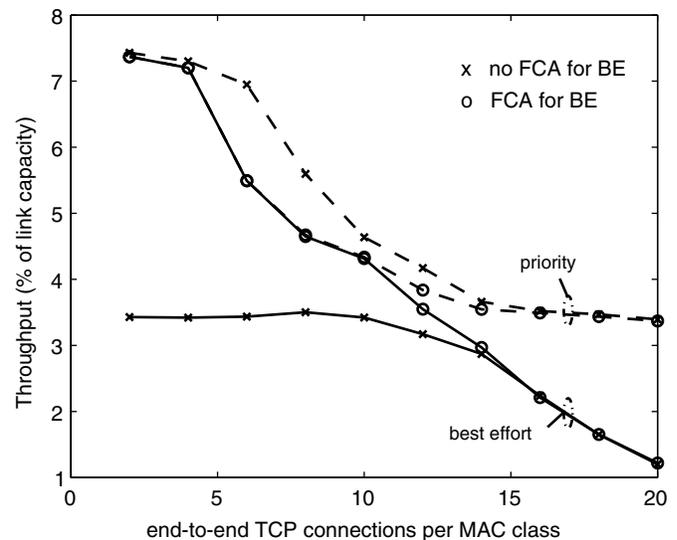


Fig. 9. Mean throughputs of MAC priority and MAC best effort connections under the two policies regarding the free capacity distribution—equal number of connections per MAC class.

of such a policy, which is investigated further in our study, is to make spare capacity available only to the MAC priority class.

Fig. 9 plots the average connection throughput when the link is shared among equal numbers of end-to-end BE and priority (prio) connections in two scenarios: in the first one, free capacity is made available to both BE and priority connections, whereas in the second only to the latter. It is illustrative regarding the differentiation space of the two MAC-level mechanisms, i.e., the priority scheduling and the FCA policy. The FCA policy dominates at low loads: when BE connections have access to spare capacity, the priorities of the scheduler actually have no effect and the two classes of connections obtain the same performance. On the contrary, restriction of spare capacity to priority flows only, has a dramatic effect. BE flows are hard-limited to the capacity corresponding to their actual requests, whereas priority connections, having the exclusive privilege to spare capacity, increase their advantage over the former. At higher loads, we face the opposite situation. It is the scheduling mechanism that provides service differentiation to the flows, whereas the impact of FCA policy fades out together with spare capacity as such.

The nonavailability of free capacity to BE flows penalizes the system utilization at low loads. On the contrary, the impact becomes negligible for higher loads, where the amount of spare capacity is reduced and the capacity allocation is performed almost exclusively on the basis of capacity requests.

3) *The Impact of Movable Boundary*: The introduction of a *boundary* can function as an additional regulator on the service differentiation offered to different traffic classes. The (movable) boundary concept was introduced in the context of the hybrid multiplexers [43], multiplexing circuit-switched voice with packet-switched data. It has been adopted in MAC protocols (see for example [10]) for integrated voice/data services with similar benefits: It leads to better service for data and higher resource utilization without affecting the service offered to voice.

Introducing a movable boundary for hard-limiting the capacity that is made available to the MAC-priority connections, one may leverage the performance gap between the two MAC

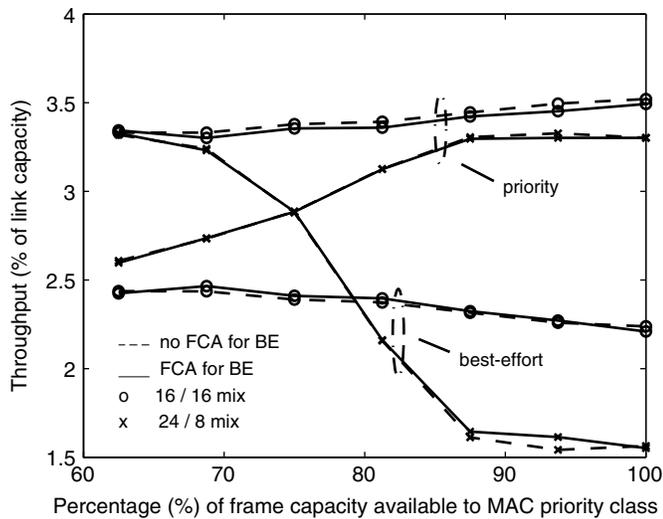


Fig. 10. Mean throughputs of MAC priority and MAC best effort connections versus the movable boundary position—32 TCP connections, two different mixes of connections per MAC class (16 and 24 split, respectively).

classes. Fig. 10 shows the impact of boundary under two different connection mixes that drive the link in saturation. This impact is much more dramatic for the 24/8 mix (24 priority versus 8 best effort connections) rather than the uniform 16/16 mix. In the first case, it is even possible to cancel the advantage of priority connections due to their privileged treatment by the BoD scheduler, if a low enough value is chosen for the boundary. In general, the impact of the boundary is not only a function of the number of TCP connections sharing the link but also of the proportion of each MAC class in the connection mix.

C. Combined Transport-MAC Layer Differentiation

It should be now clear that both the transport and MAC layers include knobs that can be tuned in order to favor some connections over others, in essence providing qualitatively better service to the former. Even more significantly, the available options provide enough flexibility, in that differentiation can be provided over the full range of system load, using different capabilities each time.

The satellite network can build a variety of radio bearers (transfer capabilities), equal in size to the Cartesian product of the available tuning knobs at the two layers. Coordination between the layers is dependent on the available interlayer procedures of the radio interface. There are “heavier” radio interfaces, like the standardized third-generation systems, featuring well-defined primitives between the radio resource allocation layer and the layers/sublayers of the radio interface [44]. In “lighter,” less connection-oriented interfaces, the necessary information may be carried on the packet headers in line with the proposed IP QoS frameworks. In both cases, appropriate mapping rules are required at the data and control plane.

Figs. 11 and 12 demonstrate possible bearers that may be generated on the basis of the available options at the MAC and the transport layer. In Fig. 11, we plot the throughputs obtained when split connections are mapped to the MAC priority class (*split-prio*) and end-to-end connections are mapped to the MAC best effort class (*e2e-BE*). The connections of the first MAC

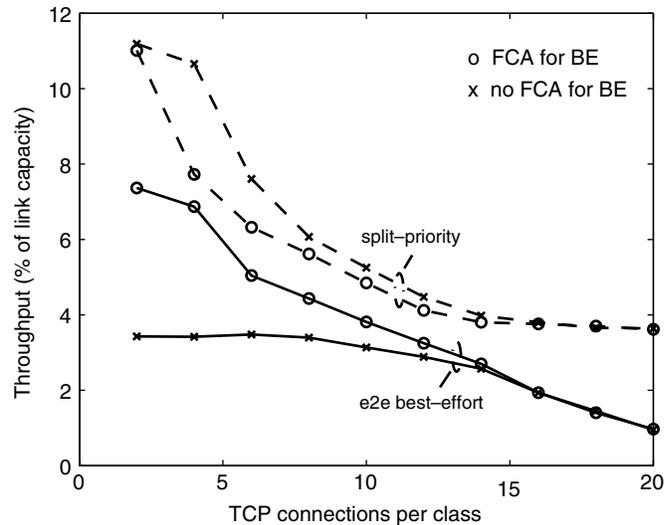


Fig. 11. Average throughput of two connection classes under the two policies regarding free capacity distribution—split (end-to-end) connections mapped to MAC priority (best effort) class.

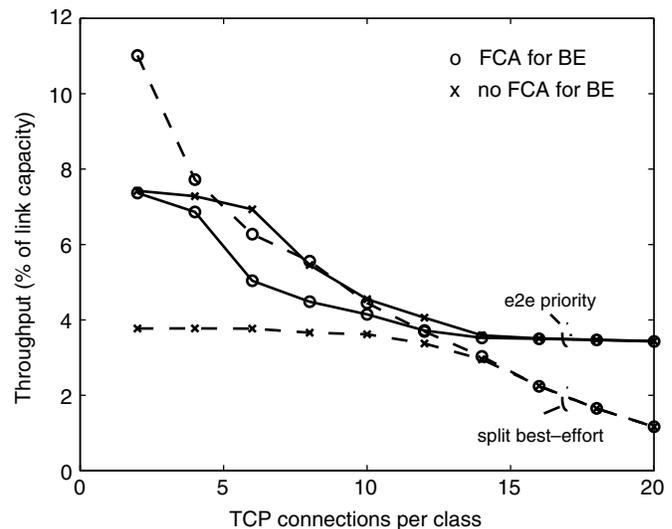


Fig. 12. Average throughput of two connection classes under the two policies regarding free capacity distribution—split (end-to-end) connections mapped to MAC best effort (priority) class.

class achieve consistently better performance thanks to the split mechanism at low loads and the MAC prioritization at higher loads. The performance gap increases when the MAC priority class monopolizes the spare capacity but only over the range of load where the latter is available.

In Fig. 12, the mapping of transport-upon MAC-classes is reversed, i.e., split connections are mapped to the MAC BE class (*split-BE*), whereas end-to-end connections are treated with priority from the BoD scheduler (*e2e-prio*). This scenario may be seen as a demonstration of what could be the relative performance achieved by the connections if the MAC layer adheres strictly to the layering principle and does not take into account the decisions at the transport layer. The curves demonstrate once again the load areas dominated by each differentiation mechanism. The split connections enjoy better performance at low loads, as far as spare capacity is available

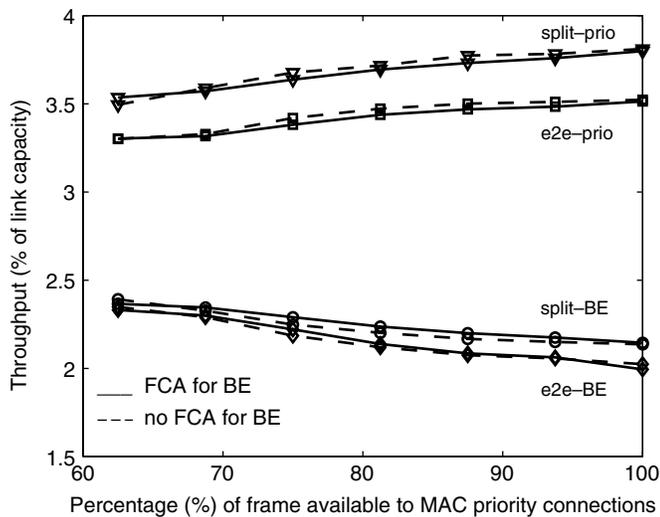


Fig. 13. Variation of the average throughput of the four combined transport/MAC-layer connection classes with the boundary position under the two policies regarding free capacity distribution—saturated link, no spare capacity.

to them. The combination of the split mechanism with spare capacity availability effectively cancels the prioritization decisions at the MAC layer. Only for higher load, where the scheduling discipline impact becomes important, can the e2e-prio connections achieve higher throughput. When free capacity is not available to split-BE, e2e-prio connections prevail over the former throughout the load range (number of connections). In this case, the throughput of the split-BE connections is upper-limited by the capacity made available to them via requests. The trend in their performance is similar to the one of e2e-BE connections in Fig. 11; however, the absolute throughput values are slightly higher.

The potential of these combinations to lead to separate, distinguishable service levels is investigated in Figs. 13 and 14. TCP connections that share the link are equally distributed among the four classes, each one corresponding to a certain transport/MAC class combination (i.e., split-prio, split-BE, e2e-prio and e2e-BE). Two scenarios are investigated: in the first scenario, the link is saturated (no spare capacity), whereas in the second scenario, the link is lightly loaded. Different values of the boundary and the two possible policies regarding the distribution of spare capacity are considered.

In the first case, the TCP connections—predictably—obtain the best performance when they are split and, on top of this, are prioritized by the BoD scheduler. There is a performance improvement that reaches 10%, while the available frame capacity, controlled by the boundary varies from 60% to 100% (Fig. 13). End-to-end priority connections follow, confirming that at high load the scheduling discipline is the decisive factor in service differentiation. On the other hand, both types of BE connections are restricted to lower portions of the capacity, their percentage being reduced while priority connections are allowed to access a bigger portion of the total MAC frame. Nevertheless, split connections maintain their advantage over end-to-end connections, allowing four distinct service levels to be provided.

In the second case, split-prio connections still achieve the best performance but now the relative weights of the FCA policy and

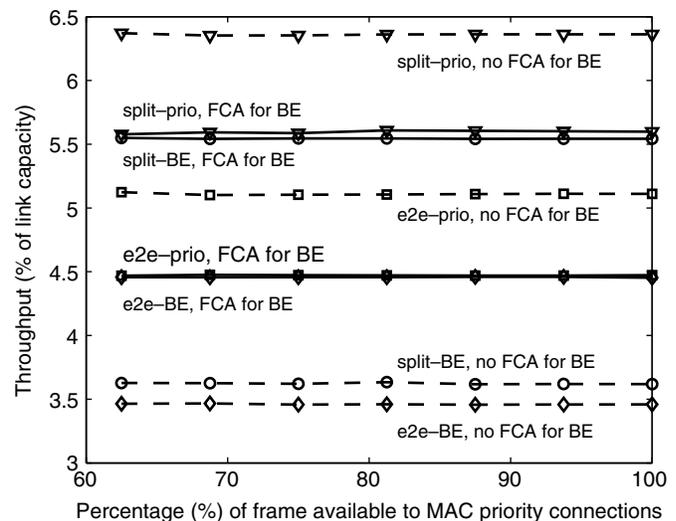


Fig. 14. Variation of the average throughput of the four combined transport/MAC connection classes with the boundary position under the two policies regarding free capacity distribution—lightly loaded link.

TABLE I
FACTORS AND LEVELS FOR THE STUDY OF SERVICE DIFFERENTIATION MECHANISMS AS A 2^3 TEN FACTORIAL DESIGN

Variable	Factor	Level 1	Level 2
<i>A</i>	<i>Transport-level mode</i>	<i>end-to-end</i>	<i>split</i>
<i>B</i>	<i>MAC scheduling discipline</i>	<i>FCFS</i>	<i>priority-aware</i>
<i>C</i>	<i>FCA policy</i>	<i>FCA to BE</i>	<i>no FCA to BE</i>

the split transport mechanism are different (Fig. 14). Split-BE connections outperform e2e-prio connections, as long as spare capacity is available to them; the latter exhibit no advantage over e2e-BE connections under the same condition (i.e., availability of spare capacity to BE connections). The four distinct service levels when MAC BE connections are restricted from spare capacity are reduced to two when this restriction is removed (split versus BE). The boundary impact is negligible since both the link load and the proportion of priority connections in the connection mix (50%) are low enough.

D. Allocation of Variation Analysis

The allocation of variation technique [45] provides an alternative, more formal demonstration of the weight of each mechanism on the TCP performance. In this context, we may consider an *experimental design* investigating the impact of three *categorical variables or factors* on the performance of the to-be-prioritized flows: the transport-level mechanism, the MAC scheduling discipline, and the policy regarding the spare capacity. There are two alternatives (*levels*) for each one of the three factors (Table I). Assuming an additive model for the experiment analysis and the measured TCP throughput as the *response variable* of the experiment, the *observed responses* y_{jkl} , $1 \leq j, k, l \leq 2$, corresponding to experiments with all eight possible combinations of factor levels, can be expressed as

$$y_{jkl} = \mu + a_j + \beta_k + \gamma_l + \delta_{ABjk} + \delta_{ACjl} + \delta_{BCKl} + \delta_{ABCjkl} + e_{jkl}$$

TABLE II
EFFECTS AND VARIATION EXPLAINED IN THE STUDY OF SERVICE
DIFFERENTIATION MECHANISMS (SATURATED 512 kb/s LINK)

Factor	Effect	Percentage of variation	Confidence interval
A	-125.86	1.55	(-133.1, -118.61)
B	-1000.02	97.99	(-1007.27, -992.78)
C	-8.24	0.006	(-15.48, -0.99)
AB	49.46	0.239	(42.21, 56.69)
AC	-0.17	0.000	(-7.42, 7.06)
BC	0.88	0.000	(-6.37, 8.12)
ABC	-3.22	0.001	(-10.47, 4.02)

where μ is the *grand mean* of all observed responses and $\alpha_j, \beta_k, \gamma_l$, are the *effects* of the three prioritization mechanisms (A, B, C—Table I) under one of the two alternatives available for each. The term δ_{XYij} accounts for the first-order interaction between factors X and Y at levels i and j , respectively, δ_{XYZijk} is the effect of the second-order interaction between factors X, Y , and Z at levels i, j , and k , respectively, and e_{jkl} is the error term computed for a single experiment on the basis of the experiment *replications* (i.e., simulation runs) performed for it. For each graph point (given number of TCP connections loading the link), we come up with a three-factor full-factorial analysis experiment with ten replications.

Table II lists the estimated percentage of variation explained by each considered experiment factor under link saturation conditions (high load), as well as the variation explained by their interactions. It also provides the confidence intervals for these effects. The weight of the MAC prioritization mechanism at these loads is evident by simple inspection of columns 2 and 3: it is responsible for 98% of the variation with regard to the *grand mean* μ [measured in this context in kilobytes per second (kB/s)], with a narrow enough confidence interval. The split mechanism has a clear but much less significant impact, whereas the FCA policy does not have any. Both its percentage of variation score and the zero-inclusive confidence intervals of the first-order interaction effects δ_{AC}, δ_{BC} confirm its minimum impact at high loads (as expected and discussed already in Section V-B).

E. Additional Considerations

1) *Aggregation*: Higher levels of TCP connection multiplexing at the BoD entities let the same number of connections obtain higher throughput (Fig. 15) and utilize the link more efficiently (Fig. 16). From the system point of view, the link is saturated with fewer connections under higher aggregation levels. The free capacity of the system is now shared among fewer terminals (more accurately: BoD entities) and that reduces the probability that spare capacity is assigned to a terminal that might not need it. The relative benefit from the connection multiplexing at the BoD entity level decreases as the link approaches the saturation area: there the capability of the entities to grab and use the spare capacity does not matter any more, since there is no spare capacity left.

2) *Link Errors*: All the results shown so far have been derived under the assumption of error-free links. In general, the validity of this assumption is strongly related to the satellite environment under consideration and the respective physical layer design. Regarding the latter, the trend has always been

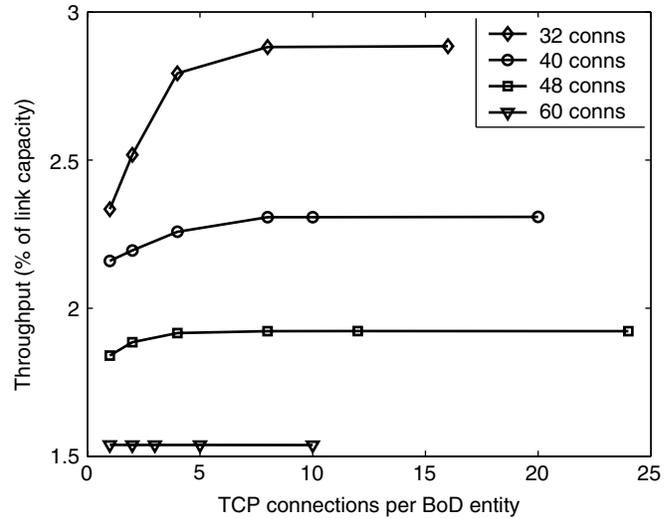


Fig. 15. Average connection throughput variation versus number of TCP connections per BoD entity for fixed total connection number.

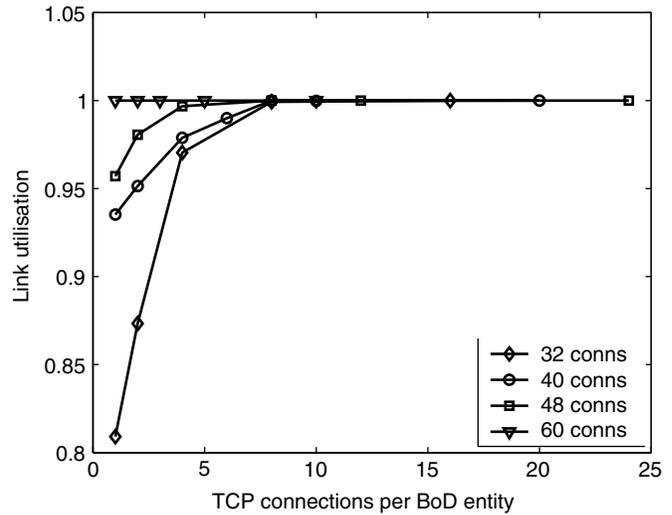


Fig. 16. Link utilization variation versus number of TCP connections per BoD entity for fixed total connection number.

to “overdimension” the link, i.e., to add all the required redundancy so that an error-free link is emulated. Physical layer implementations providing bit-error rates (BERs) as low as 10^{-10} after decoding are available for fixed broadband satellite networks. Such error rates do not have an impact on individual TCP connections, even when files of several megabyte (MB) are transferred.

Link errors could well cancel the aforementioned differentiation mechanisms. In the ideal—but also strictly theoretical—case of uniform link losses, differentiation is still feasible; split connections increase their relative advantage over end-to-end connections, at least at moderate error rates [42]. This is not the case though in the more realistic case of uncorrelated errors at the different transmitter-receiver paths over the satellite link, where connections might achieve performance that has little to do with the desired one, as reflected in its mapping and treatment at transport and/or MAC layer. Error control techniques become mandatory if the service differentiation

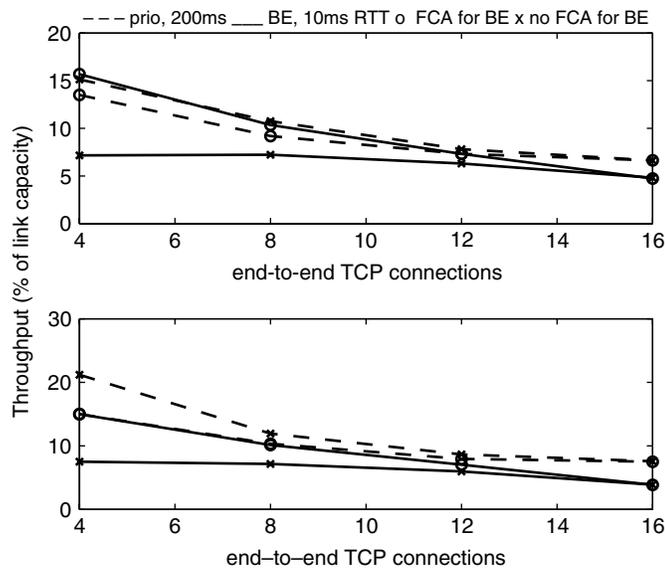


Fig. 17. Mean connection throughput for the two MAC-layer classes under nonuniform RTTs—the TCP bias against longer RTTs cancels the impact of the MAC scheduling mechanism (top); split connections preserve the intended differentiation (below).

is to be preserved. Adaptive FEC [46] or/and ARQ [36], may preserve the existing or introduce a further level of differentiation in the connection treatment, at the expense of additional data redundancy and signaling overheads. Combination of connectionless link-layer protocols and the split-connection mechanism into the same bearer can provide performance enhancement even over highly loaded BoD satellite links [42].

If the problem were approached only from the system point of view, then some form of pure channel-dependent scheduling would maximize the link utilization. Any attempt to respect the class hierarchy over the satellite link requires more complex scheduling algorithms than the simpler “give slots according to the channel quality” approach and introduces a tradeoff with respect to the resource utilization and the network capability to preserve the qualitative differentiation among service classes.

3) *Different Propagation Delays in the Terrestrial Network*: The inherent bias of TCP against connections with longer RTTs is another factor that may cancel the service differentiation mechanisms of the network. In Fig. 17, end-to-end connections that are treated as best effort at the MAC-level outperform at low load those treated with priority but exhibiting 200 ms higher RTTs over the terrestrial network, as long as they are allowed access to spare capacity. In [47], alternative window increase mechanisms that can alleviate this inherent bias are discussed; appropriate scheduling algorithms may also serve the same purpose. In the satellite network, the splitting mechanism can constitute yet another way to address this bias. By splitting the to-be-prioritized connections, the advantage of the best effort connections with regard to their small RTT is cancelled. The former obtain performance at least as good as the latter at low loads, whereas they clearly outperform them at higher loads. In fact, the split connections can decouple the overall TCP performance from the TCP parameters (e.g., *rwnd*) of the terrestrial portion of the connection as well as its RTT and the congestion level met at the rest of the network.

However, the differentiation objective can be served better at low loads by the FCA policy. If the shorter-RTT flows do not have access to spare link capacity, the desired differentiation can be preserved under light load, irrespective of the treatment of the connections at the transport layer.

VI. SUMMARY AND CONCLUSION

In the last decade, the study of IP networks has attracted tremendous interest and several new approaches, considering the flow control problem from zero basis, have been presented. The investigation of the flow control problem within an optimization framework has led to various proposals about the “optimal” way to perform flow control, often with reference to potential charging schemes (see, for example, [48], [49], and references therein).

Despite the impact of these studies and considering the current popularity of TCP, it should be taken for granted that TCP will remain the main flow control protocol within Internet, at least in the short-term future. From an IP point of view, broadband satellite networks will be yet another heterogeneous domain over which TCP/IP has to be efficiently supported, in agreement with the adopted IP QoS framework, however hard or soft this is.

The first contribution of this paper is the demonstration in a simulated BoD environment of possible mechanisms available at transport and MAC layers to provide differentiated service to TCP flows. The BoD scheduling algorithm and the policy regarding spare capacity distribution are two MAC-layer mechanisms that dominate at different traffic load areas: the former is effective at high levels of traffic load whereas the latter has an impact at low traffic load. Their combination with transport-level mechanisms and options provide the satellite operator with adequate flexibility in designing radio bearer services over the satellite network.

A second contribution of this paper is the evaluation of the actual gain achieved with proposed TCP options and modifications in a MAC-shared satellite environment. This gain is always significantly less than the one estimated in studies addressing a single TCP connection or multiple TCP connections, when the impact of the MAC layer is not considered. Additional measures for maintaining the service differentiation in the case of link errors were discussed and the potential of the split-TCP mechanism to alleviate the bias against TCP connections with long RTTs was outlined.

The problem was also addressed within a fixed-point framework. We have generalized formulation available in literature to cater for asymmetry and MAC-shared links, two common features of satellite networks. In the case of the infinite MAC buffer system, it was possible to derive a fixed-point solution for the delay experienced at MAC level. The applicability of the method to our study of differentiation mechanisms is mainly limited by the lack of analytical models that account for the MAC prioritization mechanisms and the nonuniform distribution of traffic load among satellite terminals.

In this paper, we have shown how the level of differentiated service depends on the system load and where the available

mechanisms succeed or fail in fulfilling this objective. However, the provision of absolute, *quantitative* QoS guarantees is ultimately dependent on whether the number of connections sharing the link can be controlled, implying some form of access control in the network. The ideas proposed in [50]–[52] might be the starting point for the introduction of additional functions into the entry points of the satellite network, if such QoS guarantees are to be provided to TCP connections.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions. They are also grateful to L. Wood and K. Narenthiran for their comments on earlier versions of this work.

REFERENCES

- [1] *Digital Video Broadcasting (DVB)*, Nov. 2002. Interaction channel for satellite distribution systems.
- [2] H. Skinnemoen and H. Tork, "Standardization activities within broadband satellite multimedia," in *Proc. IEEE ICC 2002*, vol. 5, New York, NY, 2002, pp. 3010–3014.
- [3] W. R. Stevens, *TCP/IP Illustrated, Volume 1: The Protocols*. Reading, MA: Addison-Wesley, 1994, ch. 21.
- [4] S. Floyd, M. Mathis, J. Mahdavi, and R. Povolosky, An extension to the selective acknowledgment (SACK) option for TCP, IETF, RFC 2883 (Proposed Standard), July 2000.
- [5] S. Floyd and T. Henderson, The NewReno modification to TCP's fast recovery mechanism, IETF, RFC 2582 (Experimental), Apr. 1999.
- [6] M. Allman *et al.*, Ongoing TCP research related to satellites, IETF, RFC 2760 (Informational), Feb. 2000.
- [7] L. G. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *Proc. AFIPS Conf., National Computer Conf.*, 1973, pp. 711–716.
- [8] L. Kleinrock and S. Lam, "Packet switching in a slotted satellite channel," in *Proc. AFIPS Conf., National Computer Conf.*, 1973, pp. 703–710.
- [9] W. Crowther *et al.*, "A system for broadcast communication: Reservation aloha," in *Proc. Int. Sys. Sci. Conf.*, Jan. 1973, pp. 371–374.
- [10] J. E. Wieselthier and A. Ephremides, "Fixed- and movable-boundary channel-access schemes for integrated voice/data in wireless networks," *IEEE Trans. Commun.*, vol. 43, pp. 64–74, Jan. 1995.
- [11] A. Hung, M. J. Montpetit, and G. Kesidis, "A framework for ATM over satellite," presented at the IEEE GLOBECOM'96, London, U.K., 1996.
- [12] C. Rosenberg, "Broadband satellite systems: A network perspective," presented at the IEEE GLOBECOM 2000, San Francisco, CA, Nov. 2000. Tutorial.
- [13] H. Peyravi, "Medium access control performance in satellite communications," *IEEE Commun. Mag.*, pp. 62–71, 1999.
- [14] J. Neale and A. Mohsen, "Impact of CF-DAMA on TCP via satellite performance," in *Proc. IEEE GLOBECOM 2001*, San Antonio, TX, Nov. 2001, pp. 2687–2691.
- [15] M. Sooriyabandara and G. Fairhurst, "Dynamics of TCP over BoD satellite networks," *Int. J. Satell. Commun. Networking*, vol. 21, pp. 427–449, July 2003.
- [16] D. Connors, "Response initiated multiple access protocol," in *Proc. IEEE GLOBECOM 2000*, San Francisco, CA, 2000, pp. 1124–1129.
- [17] D. P. Heyman, T. V. Lakshman, and A. L. Neidhardt, "A new method for analyzing feedback-based protocols with applications to engineering web traffic over the Internet," *ACM Sigmetrics Performance Evaluation Rev.*, vol. 25, no. 1, pp. 24–38, 1997.
- [18] A. Berger and Y. Kogan, "Dimensioning bandwidth for elastic traffic in high-speed data networks," *IEEE/ACM Trans. Networking*, vol. 8, pp. 643–654, Oct. 2000.
- [19] K. Lindberger, "Balancing quality of service, pricing and utilization in multiservice networks with stream and elastic traffic," presented at the 16th Int. Teletraffic Congress, Edinburgh, U.K., 1999.
- [20] A. Feldmann. (1998, Dec.) Characteristics of TCP Connection Arrivals. [Online]. Available: <http://www.research.att.com/~anja/feldmann/papers.html>
- [21] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation," presented at the ACM SIGCOMM 1998, Vancouver, BC, Canada, 1998.
- [22] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The macroscopic behavior of the TCP congestion avoidance algorithm," *Comput. Commun. Rev.*, vol. 27, no. 3, pp. 67–82, July 1997.
- [23] N. Cardwell, S. Savage, and T. Anderson, "Modeling TCP latency," in *Proc. IEEE INFOCOM*, Seattle, WA, Mar. 2000, pp. 1742–1751.
- [24] M. Roughan, A. Erramilli, and D. Veitch, "Network performance for TCP networks. Part I: Persistent sources," presented at the 17th Int. Teletraffic Congress, Salvador, Brazil, 2001.
- [25] C. Casetti and M. Meo, "An analytical framework for the performance evaluation of TCP Reno connections," *Comput. Networks*, vol. 37, pp. 669–682, 2001.
- [26] R. J. Gibbens *et al.*, "Fixed-point models for the end-to-end performance analysis of IP networks," presented at the 13th ITC Specialist Seminar: IP Traffic Measurement, Modeling, and Management, Monterey, CA, 2000.
- [27] M. May, J. C. Bolot, A. Jean-Marie, and C. Diot, "Simple performance models of differentiated services schemes for the Internet," in *Proc. IEEE INFOCOM 1999*, New York, NY, Mar. 1999, pp. 1385–1394.
- [28] A. Misra, T. Ott, and J. Baras, "Predicting bottleneck bandwidth sharing by generalized TCP flows," *Comput. Networks*, vol. 40, pp. 557–576, Oct. 2002.
- [29] T. Bu and D. Towsley, "Fixed-Point approximation for TCP behavior in an AQM network," presented at the ACM SIGMETRICS 2001, Cambridge, MA, June 2001.
- [30] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.
- [31] T. Le-Ngoc and S. Krishnamurthy, "Performance of combined free/demand assignment multiple-access schemes in satellite communications," *Int. J. Satell. Commun.*, vol. 14, pp. 11–21, 1996.
- [32] T. Le-Ngoc and J. I. Mohammed, "Combined free/demand assignment multiple-access (CFDAMA) protocols for packet satellite communications," *Proc. 2nd IEEE Int. Conf. Univ. Pers. Communications (ICUPC)*, pp. 824–828, 1993.
- [33] The ns manual [Online]. Available: <http://www.isi.edu/nsnam/ns/doc>
- [34] M. Karaliopoulos, R. Tafazolli, and B. Evans, "An ns-derived GEO satellite network simulator: Features, capabilities, results," in *Proc. IEE Seminar Simulation Modeling Satellite Systems*, London, U.K., 2002, pp. 6/1–6/6.
- [35] G. Açar, "End-to-end resource management in geostationary satellite networks," Ph.D. dissertation, Dept. Elect. Electron. Eng., Imperial Coll., London, U.K., 2001.
- [36] G. Fairhurst and L. Wood, Advice to link designers on link automatic repeat request (ARQ), IETF, RFC 3366 (BCP), Aug. 2002.
- [37] M. Karaliopoulos, R. Tafazolli, and B. Evans, "On the interaction of TCP with BoD in GEO broadband satellite networks," in *Proc. IEEE GLOBECOM 2002*, Taipei, Taiwan, 2002, pp. 2945–2949.
- [38] V. Jacobson, R. Braden, and D. Borman, TCP extensions for high performance, IETF, RFC 1323, May 1992.
- [39] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end arguments in system design," *ACM Trans. Comput. Syst.*, pp. 277–288, 1984.
- [40] J. Border *et al.*, Performance enhancing proxies intended to mitigate link-related degradations, IETF, RFC 3135, June 2001.
- [41] Y. Zhang, *Internetworking and Computing Over Satellite Networks*. Norwell, MA: Kluwer, 2003.
- [42] M. Karaliopoulos, R. Tafazolli, and B. Evans, "TCP performance on split connection GEO satellite links," presented at the 19th ICSSC and Exhibit, Toulouse, France, 2001.
- [43] K. Kummerle, "Multiplexor performance for integrated line- and packet switched traffic," in *Proc. ICCS*, Sweden, 1974, pp. 508–515.
- [44] *Universal Mobile Telecommunications System (UMTS)*, June 2002. Interlayer procedures in connected mode, 3GPP TS 25.303, ver. 5.1.0, Rel. 5.
- [45] R. Jain, *The Art of Computer Systems Performance Analysis*. New York: Wiley, 1991, ch. 16–23.
- [46] N. Celandroni, F. Davoli, and E. Ferro, "Static and dynamic resource allocation in a multi-service satellite network with fading," *Int. J. Satell. Commun. Networking*, vol. 21, pp. 469–487, July 2003.
- [47] S. Floyd, "Connections with multiple congested gateways in packet-switched networks, Part 1: One-way traffic," *ACM Comput. Commun. Rev.*, vol. 21, no. 5, pp. 30–47, 1991.

- [48] F. Kelly, "Mathematical modeling of the Internet (extended version)," presented at the 4th Int. Congr. Industrial and Applied Mathematics, Edinburgh, U.K., 1999.
- [49] S. Low and D. E. Lapsley, "Optimization flow control I: Basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, pp. 861–874, 1999.
- [50] R. Mortier, I. Pratt, C. Clark, and S. Crosby, "Implicit admission control," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2629–2639, Dec. 2000.
- [51] A. Kumar, "Nonintrusive TCP connection admission control for bandwidth management of an Internet link," *IEEE Commun. Mag.*, pp. 160–167, 2000.
- [52] L. Massoulié and J. Roberts, "Arguments in favor of admission control for TCP flows," in *Proc. 16th Int. Teletraffic Congress*, Edinburgh, U.K., 1999, pp. 33–44.



Mercurios Karaliopoulos (S'99) was born in Thessaloniki, Greece, in 1975. He received the Diploma degree in electrical and computer engineering from Aristotle University of Thessaloniki, in 1998. In October 1999, he joined the Centre for Communication Systems Research (CCSR), University of Surrey, U.K., where he is working toward the Ph.D. degree in the area of broadband satellite networks.

His main research interests are in the area of mobile and wireless data networking and radio resource management. He is currently employed by CCSR as a Research Assistant in the Mobile Communications Research Group. During the last three years, he has been working for two European Commission projects, the IST SATIN (Satellite UMTS IP-based Network) and the IST MoDiS (Mobile Digital Broadcasting Satellite) projects and has participated in the European COST Action 272 (packet-oriented delivery over satellite).

Mr. Karaliopoulos is a Student Member of the Institution of Electrical Engineers (IEE) and a member of the Technical Chamber of Greece.



Rahim Tafazolli (M'89) is a Professor of satellite and terrestrial mobile communications and Head of the Mobile Communications Research Group with the Centre for Communication Systems Research (CCSR), School of Electronics and Physical Sciences, University of Surrey, Surrey, U.K.

His research activities focus on optimization techniques for mobile multimedia networks, mainly on advanced resource management, mobility management, and medium access control. He has published more than 300 research papers in refereed journals, international conferences, and as invited speaker. He currently has more than 15 patents in the field of mobile communications. He is an advisor and consultant to a number of mobile companies.

Prof. Tafazolli is a member of the Institution of Electrical Engineers (IEE) Committee on the U.K. Regulations on "Information Technology & Telecommunications," a Member of the EU Wireless Strategy Initiative (WSI) Think-Tank group, Chairman of "New Technologies" group of Wireless World Research Forum (WWRF), and Academic Coordinator of the U.K. Mobile Virtual Centre of Excellence (VCE). He is the Founder and the past Chairman of International Conference on "3G Mobile Technologies."



Barry G. Evans (M'86–SM'98) received the B.Sc. and Ph.D. degrees in electrical engineering and microwave systems from the University of Leeds, Leeds, U.K., in 1965 and 1968, respectively.

From 1969 to 1983, he was a British Telecom Lecturer-Reader in telecommunications systems at the University of Essex, Essex, U.K. In 1983, he was appointed to the Alec Harley Reeves Chair of Information Systems Engineering, University of Surrey, Surrey, U.K., and in 1990, he became the first Director of Postgraduate Centre for Satellite Engineering Research, which he built up to about 150 researchers and a spin-off company, Surrey Satellite Technology Ltd., Surrey. Since 1996, he has been Director of the new Centre for Communication Systems Research, Surrey, which is now 120 researchers strong and is Pro-Vice Chancellor for research and enterprise since 2001. He is the author of three books and over 400 technical papers.

Prof. Evans is a Fellow of the Royal Academy of Engineering, U.K., a Fellow of the Institution of Electrical Engineers (IEE), and a senior member of the American Institute of Aeronautics and Astronautics (AIAA). He is Editor of the *International Journal of Satellite Communications*. He has been Technical Advisor to the D.G. of OFTEL, a Member of the U.K. Foresight and MoD research committees, and an Advisor to the EU on framework and R&D programs.