Integrated Connection Admission Control and Bandwidth on Demand Algorithm for a Broadband Satellite Network with Heterogeneous Traffic

Yi Qian[†a)], *Member,* Rose Q. Hu [††]*, Nonmember,* and Catherine Rosenberg[†††], *Nonmember*

Summary There are many system proposals for satellite-based broadband communications that promise high capacity and ease of access. Many of these proposals require advanced switching technology and signal processing on-board the satellite(s). One solution is based on a geo-synchronous (GEO) satellite system equipped with on-board processing and on-board switching. An important feature of this system is allowing for a maximum number of simultaneous users, hence, requiring effective medium access control (MAC) layer protocols for connection admission control (CAC) and bandwidth on demand (BoD) algorithms. In this paper, an integrated CAC and BoD algorithm is proposed for a broadband satellite communication system with heterogeneous traffic. A detailed modeling and simulation approach is presented for performance evaluation of the integrated CAC and BoD algorithm based on heterogeneous traffic types. The proposed CAC and BoD scheme is shown to be able to efficiently utilize available bandwidth and to gain high throughput, and also to maintain good Grade of Service (GoS) for all the traffic types. The end-to-end delay for real-time traffic in the system falls well within ITU's Quality of Service (QoS) specification for GEO-based satellite systems.

Key words: *connection admission control, bandwidth on demand, broadband satellite network, QoS, GoS.*

1. Introduction

Already very successful in broadcasting entertainment services, digital satellite systems are viewed as viable service vehicles. The demand for Internet and multimedia services and the subsequent need for higher bandwidth drive network operators to seek mechanisms to cost-effectively provide broadband access. Satellite-based networks can supplement existing wire-line and legacy networks to bring broadband and multimedia services to end-users (e.g., [2], [8], [11], and [15]). Many proposals have been made to national and international regulatory agencies for allocation of spectrum for broadband applications using low earth orbit (LEO) satellites, medium earth orbit (MEO) satellites, and geo-synchronous (GEO) satellites. This paper concentrates on GEO based satellite networks, which play an ever-increasing role in the public and private Internets, due mostly to their large geographic coverage, inherent broadcast capabilities and fast deployment. They are attractive to support data, audio and video streaming; bulk data transfer such as software update or dissemination of web caches; and applications involving limited interactivity such as distance learning. They are also attractive to provide broadband access to users who are either beyond the reach of the terrestrial network, or have particular needs for broadcast/multicast applications or fast deployment.

Medium access control (MAC) protocols enable communicating access units at diverse locations within a beam to regulate the sending of their packets over the multiple access uplink and manage network resources as efficiently and fairly as possible. The MAC protocols have generally a dominant effect on the ability of the system to deliver on a QoS contract. For MAC protocols in satellite communications, the space environment possesses some major constraints that eliminate a large number of possible MAC protocols from consideration, even though these protocols may work well in terrestrial wireless systems. Some of the space environment constraints include long GEO satellite delays, poor air interface bit error rates, and low available spectrum. In [12], a literature survey is made on MAC layer protocol performance in satellite communications. Five classes of MAC layer protocols are investigated with respect to their applications in satellite communications. These classes include fixed assignment, demand assignment, random access, hybrid of random access and reservation, and adaptive protocols. It is concluded from [12] that despite the fact that there is no protocol that performs better than the others for all traffic scenarios and applications, some protocols have certain characteristics that make them more suitable for satellite communications. The literature survey in [12] also implies that current simulation and analysis of MAC layer protocol performance in satellite networks use either simple traffic scenarios or small configurations (i.e., a limited number of applications and access units). In this paper, an integrated connection admission control (CAC) and bandwidth on demand (BoD) MAC algorithm is proposed and analyzed for satellite network with heterogeneous traffic. Simulation modeling and analysis

---

[†] The author is with Department of Electrical and Computer Engineering, University of Puerto Rico, Mayagüez, Puerto Rico 00681.
a) E-mail: yqian@ece.uprm.edu
[††] The author is with Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762-9571, USA.
[†††] The author is with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

are presented in order to evaluate the performance of the integrated CAC and BoD algorithm for uplink multiple access of the high capacity broadband satellite communication system with On-Board Processor (OBP). A multi-frequency time division multiple access (MF-TDMA) scheme is used for uplink multiple access in the system. The simulations are based on the detailed models of heterogeneous traffic such as voice, video, and data. The MAC layer protocol proposed here is a hybrid of fixed assignment, reservation and demand assignment, which can be categorized as a different class from the five classes examined in [12].

The rest of the paper is organized as follows. In Section 2, a GEO-based high capacity satellite network with on-board switching is presented, and the uplink MF-TDMA scheme is described. In Section 3, an integrated CAC and BoD scheme for MF-TDMA access is proposed. Section 4 presents the performance metrics and the detailed modeling approach for the source traffic and the integrated CAC/BoD algorithm. Section 5 discusses the simulation results. Section 6 presents the conclusions of this study.

2. Network Architecture and Uplink Multiple Access Structure

Fig. 1 illustrates the architecture of the GEO-based high capacity broadband satellite network investigated in this paper. The satellite OBP can be considered as a simplified packet switch with fixed packet size. There are *n* spot beams and *n* uplinks/downlinks to/from the satellite. Two types of access units are defined in the system: Subscriber Access Units (SAUs) connect individual users to the network, and gateways connect the satellite network to other networks.

As it is shown in Fig. 1, user links support access to individual SAUs, while gateway-links (G-links) support access to gateways. A G-link is a very high-speed point-to-point bi-directional TDM link between the satellite and a gateway, whereas the user links are MF-TDMA links on the uplink while they are TDM on the downlink. The Network Control Center (NCC) is in charge of most of the signaling and management functions in the satellite network. The performance study in this paper is focused on SAU uplink access since the multiple access uplink is usually the bandwidth bottleneck of the entire satellite system. A mix of frequency and time sharing access, i.e., MF-TDMA, is baselined for the user uplink, where the SAUs access narrowband carriers of a fixed bandwidth (*b kHz*) on a time sharing basis, with a time slot (marked in *X* in Fig. 2) dimensioned on the basis of one fixed size packet transmission. For each spot beam, the user uplink MF-TDMA frame structure is shown in Fig. 2. There are *m* time slots in a frame of a TDMA channel, and *B* channels in the MF-TDMA scheme. In this paper, the fixed packet size of 48 bytes or 384 bits is used, which is based on the assumption of an on-board ATM-like switch.

SAUs are assumed to be agile enough to access any time slot (TS) in the frame. Because a SAU is only equipped with one antenna, a SAU is not allowed to transmit simultaneously on two different frequencies in the same timeslot. If two SAUs attempt to use the same TS, then the transmitted information in the TS is garbled and then lost. Hence, a MAC scheme is needed to assign each TS to only one SAU. Because the uplink time slots are a scarce resource, and there can be tens of thousands of active SAUs, a MAC scheme is needed to assign TSs to SAUs to satisfy their QoS requirements without wasting precious time slots. An integrated CAC and BoD scheme is presented in the next section.

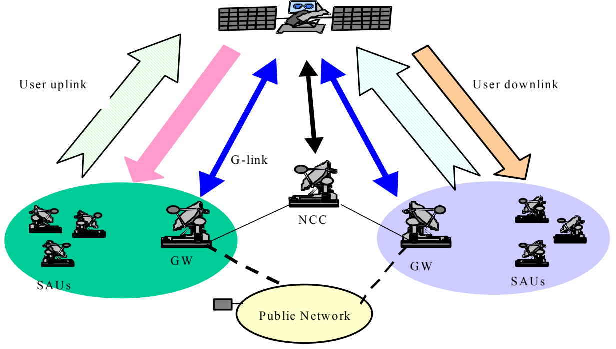


Fig. 1 The GEO-Based High Capacity Broadband Satellite Network.

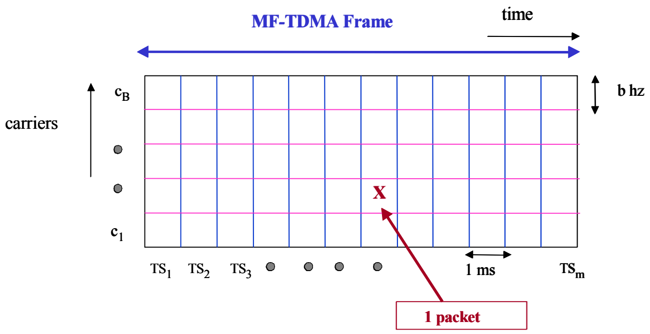


Fig. 2 User Uplink MF-TDMA Frame Structure for a Spot Beam

3. Integrated CAC and BoD Algorithm

This section describes the integrated CAC and BoD algorithm. The aim is to develop an end-to-end resource management scheme that can be implemented in a large scale satellite system and hence several issues linked to scalability, complexity and cost trade-offs are tackled. As it is mentioned in [13], there are mainly two factors that shape the operation principles of a BoD scheme within the satellite network. The first one is the long propagation delay between the satellite and the access units. This implies that the close loop control between the BoD controller (either located on the satellite or in the NCC) and the SAU will be difficult. The one-way propagation delay time between a GEO satellite and an access unit is 125 msec. The second one is the integration of CAC and BoD. End-to-end resource management for broadband satellite systems integrating multiple access, BoD and CAC is the key to deliver acceptable QoS to services while providing adequate efficiency (i.e., a level of Grade of Service (GoS) that entails the use of such systems).

3.1 CAC Definition

CAC is a network process that receives as an input, a connection request that specifies the traffic descriptor and QoS (quality of service) requirements of the connection and returns a response granting or denying the admission request. The objective of the CAC is to ensure that the network meets its end-to-end QoS guarantees to connections that are admitted into the network. The CAC process is responsible for deciding whether a new connection request can be accepted, and if so, then how much resource should be allocated to it.

3.2 BoD Definition

On the other hand, *BoD* is defined here as a set of MAC protocols and algorithms that allow a connection to request resources on a demand basis, while the connection is already in progress, in an environment where many bursty connections share a common medium access link. Hence, BoD is needed because of the multiple access user uplink. BoD will be invoked many times during the progress of some types of connections (and will not be invoked for other types of connection), while CAC is usually invoked only once at the connection set-up for every connection (except in the case of re-negotiations). BoD is the process by which SAUs can request resources on top of other resources that have been statically allocated to them, on a periodic time frame basis, during the CAC process. The proposed BoD process consists of the following 5 steps, as shown in Fig. 3.

1) Computing the needs in the SAU;
2) Signaling the needs from the SAUs to the BoD controller;
3) Computation, by the BoD controller, of the allocation of time-slots (TS), i.e., the creation of the Burst Time Plan (BTP), which is a table containing the assignment of each TS for the next period.
4) Signaling the response from the BoD controller to the SAUs (broadcast of the BTP);
5) Allocating the TS among the different connections in the SAU.

One of the main issues with BoD is linked with timing since the period is roughly only in a few tens of milli-seconds, i.e., a SAU will have to compute a request every few tens of milli-seconds, and the controller will only have this time for computing the allocation and preparing the BTP. Hence one of the main issues is to develop algorithms that are fast and scalable.

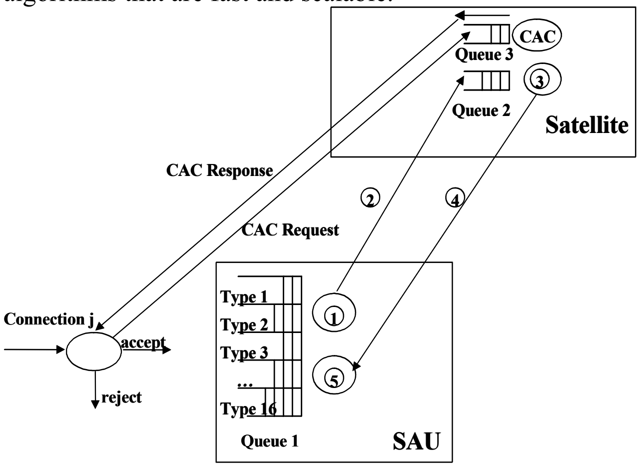


Fig. 3 Integrated CAC and BoD Model

Using BoD means that, on a need basis in each beam, a request for resource (RR) will be made on behalf of each active BoD connection periodically by each active SAU. In the context of a GEO based satellite network, BoD will not be used for real time connections because of the long response time involved. BoD is not necessarily connection-based, i.e., a SAU may perform some kind of aggregation in order not to send a separate request for resource for each of its connection. This is key to the problem because, most probably, connection-based BoD will not be implemented due to the large overhead it requires. However, to explain the method it will assume in the following that RR is connection based.

3.3 Integrated CAC and BoD Algorithm

This study shows that CAC and BoD are interconnected intricately, and careful integration of the design of CAC and BoD is crucial for the design of viable and efficient broadband satellite networks offering QoS guarantees to connections. The integrated algorithm of uplink CAC with BoD is described in the following.

A connection *j* of a given traffic type will be allocated at call set-up by the CAC for the call duration the following amounts of resource:

1. A static amount of resource SR_j on the uplink. Depending on the connection type, the traffic descriptor and requested QoS, SR_j could be equal to

zero or up to the Peak Rate of the connection. The connections that are of a type that does not use BoD will only receive SR_j amount of resource on the uplink.

2. A booked amount of resource BR_j that is allocated by the CAC for the call duration but managed by the BoD (that is where the combination or coupling between uplink CAC and BoD occurs). Depending on the connection traffic type, the traffic descriptor and requested QoS, BR_j could be equal to zero or up to *PeakRate* – SR_j amount of resource. In order to get any of the resource booked for it, the connection has to use BoD. BR_j is reserved for connection *j* so that if it needs it is sure to get it. However if it does not ask for it (or at least not completely), it is available on a best-effort basis for any other connections within the same beam. Having this booked rate allows the delivery of QoS to services by making sure that they will always get what they need. The advantage of not allocating this rate statically is that when not needed, it can be made available to other connections in the same beam.

The CAC on the uplink will accept a call *j* only if it has enough resources, i.e., if the sum of what is to be statically allocated to *j* (i.e., SR_j) and what is to be booked for *j* (i.e., BR_j) is less than the total amount of resource of the multiple access link, minus the sum of the already allocated resource and minus the sum of the booked resource for all ongoing calls *k* on the uplink, i.e., only if:

$$SR_j + BR_j + \sum_k SR_k + \sum_k BR_k \le C_T \qquad (1)$$

where C_T is the total amount of resource available for the multiple access uplink traffic. In fact, some SAUs have constraints of their own and the above condition is only a necessary condition. Another constraint of SAU is linked to the MF-TDMA scheme as follows:

SAUs are not allowed to transmit simultaneously on the same timeslot in different frequencies. (2)

Note that a connection is not restricted to ask only for BR_j. Any BoD connection *j* can ask for RR_j that is greater than its BR_j (RR_j and BR_j have to be understood as values on top of SR_j). What happens is that if RR_j is greater than BR_j, the connection will get at least BR_j. The best-effort need of connection *j* for the given period is: $\mathrm{BE}_j = \max(0, \mathrm{RR}_j - \mathrm{BR}_j)$. These BE_j are managed completely by the BoD controller. What a connection will really get is $BR_j + BE_j'$ where BE_j' is its fair share of the best effort capacity available for this period that the BoD controller will compute by knowing the best-effort need of every connection. The only interaction of CAC with BE assignment is that the CAC could admit new connections or release existing ones so that the total amount of resource available for best-effort for a given period, say C_A, depends on: the total amount of resource statically allocated for this period, i.e.: $\sum_k SR_k$; and the amount of resources that had been booked and have been requested for this period, i.e., $\sum_k \min(RR_k, BR_k)$.

Indeed,

$$C_A = C_T - \sum_k \min(RR_k, BR_k) - \sum_k SR_k \qquad (3)$$

Hence the BoD controller has to share for the given period C_A among all the connections accessing the same uplink (i.e., in the same beam) that have a non-zero BE_j for this period. If C_A is large enough, all the connections could get what they ask for and the leftover capacity could be freely assigned. Otherwise, the BoD controller will allocate C_A among all competing connections with fairness.

3.4 Fair and Efficient Share of C_A within a Beam

At the beginning of each period the BoD controller has to decide the distribution for the next period the available amount of resource C_A among all the connections requesting best-effort resources, i.e., for which BE_j is nonzero. It needs a solution that will share the available capacity in a fair and efficient manner. Game theory ([10] and [13]) suggests what to do: if for the next period, there are *n* connections that have a non zero BE_j, then allocate to connection *k* ($1 \le k \le n$) BE_k* solution of the following optimization problem:

Maximize $\prod_{k=1}^{n} BE_k^*$

Subject to

$$BE_k^* \le BE_k,\ \forall k \ \ and \ \ \sum_{k=1}^{n} BE_k^* \le C_A$$

This problem can be solved in a very fast way since the computational complexity is in *O(n)*, and the solution scales very well with the number of connections and SAUs.

3.5 The BTP Jitter Management

Every period, the BoD controller broadcasts on each downlink the BTP for the corresponding beam. The BoD controller sends a BTP every period *P* where *P* is a multiple of the frame duration (*F*). The BTP is said to be frame-based if *P* = *F*. If the BTP is frame-based and *F* is relatively small, the BTP can be filled independently from one BoD period to another because the real time (RT) connections will not be too affected in terms of jitter (i.e., RT traffic can tolerate the jitter implied by rebuilding completely the BTP from scratch at each period). Then a simple way to create a BTP at the beginning of a period while respecting the MF-TDMA constraint would be to fill it up row by row where a row corresponds to a frequency and the CAC does not need to be involved in the creation of the BTP, it just sends information about newly admitted or released connections to the BoD controller. On the other hand, when *F* is too large or *P*>*F* (for instance the BoD controller cannot compute a BTP every *F* seconds), then it needs to creating a BTP in such a way that does not increase the jitter of RT connections.

A hierarchical management of the Burst Time Plan is proposed for improved scalability and simplified coupling of BoD and CAC. This is really about the partitioning of physical responsibilities to fill-up and manage the BTP between the CAC and the BoD and the corresponding exchanges of information. The concept of a CAT (CAC Allocation Table) is introduced, which is a masking table of the BTP filled up by the CAC as described below. In creating the BTP, the BoD Controller uses the CAT that the CAC sends to it periodically.

Whenever a new real-time connection i requesting admission has been admitted by the CAC, the CAC computes, using SR_i, the number of statically allocated TSs (SATS) that must be assigned to connection i. Next, the CAC updates the CAC allocation table (CAT) by placing the RT SATS in a way that does not violate the SAU constraint. Note that the CAT does not include the SATS for non real-time connections since they are not jitter sensitive. Updating the CAT requires a non-trivial placement algorithm. Note that a connection could be rejected if the algorithm cannot find a suitable placement. The CAC also erases from the CAT the SATS of the newly released real-time connections. Hence, the CAT comprises at a given time, the current view of the uplink CAC on the status of the TS with respect to real-time connections. Each TS in the CAT can be either unused or SATS. The CAT TSs that are SATS have been assigned by the CAC to RT connections for the next period and have an identifier identifying their connection. Note that the CAC only works on connection (or session) level, so that the CAT is connection-based. Hence, the CAC can use the CAT to deallocate SATS for a real-time connection that has been released, allocate SATS for new real-time connections, and, if deemed necessary, rearrange the table while taking into account the Packet Delay Variation constraints of the real-time connections. Since it is the BoD that sends the BTP to the SAUs, it means that any TS allocation is only effective once the BoD has received the corresponding CAT and has integrated the information into its current BTP. The CAC also sends information to the BoD controller about newly released or admitted non-real time connections, i.e., the corresponding number of SR and BR.

The assignments in the CAT satisfy the constraints of the SAUs. Thus, a new real-time connection is accepted by the uplink CAC only if the connection's required SATS can be placed properly to the currently unused TSs in the CAT, while respecting to the constraints on the SAUs. Note that some connections may be admitted, if the CAT is re-arranged, to remove fragmentation. Hence, the CAT may require re-arrangement from time to time. Having separated the impact of CAC and BoD on the allocation in the above way allows placing the BoD controller and the CAC controller in different elements.

4. Performance and Simulation Modeling

This section describes performance and simulation modeling approach. This includes two parts, namely source traffic modeling and simulation modeling on the integrated CAC and BoD algorithm.

4.1 Source Traffic Modelling

Many different applications are identified as possible GEO satellite services. By using the detailed traffic modeling methods described in [1], [3], and [4], a list of 16 traffic types for most possible applications is identified for the satellite network. Five QoS classes are defined for the satellite system depending on how delay-sensitive the traffic is, with Class 1 the most sensitive and Class 5 least sensitive to the delay.

- Class 1: ***real-time traffic***, e.g., Custom Calling Service traffic;
- Class 2: ***conversational traffic***, e.g., VoIP, packet video traffic, etc.;
- Class 3: ***interactive traffic 1***, e.g., data dissemination;
- Class 4: ***interactive traffic 2***, e.g., web access, telnet;
- Class 5: ***best effort traffic***, e.g., ftp, email;

A two level traffic model composed of a Session Level and a Burst/Packet Level is used. Session Level is modeled as a Poisson process, i.e., session inter-arrival follows an exponential distribution. Session duration is also exponentially distributed. Table 1 gives an example of the session level traffic parameters. Detailed description of the Burst/Packet Level modeling was presented in [1], [3] and [4]. For most traffic types, the burst level is modeled inside each session as an ON/OFF discrete-time Markov model with exponential ON and OFF distributions. For some of the applications, special models are used during ON period. For example, for the Web Access application, the ON period follows a Weibull distribution while the OFF period follows a Pareto distribution. Inside each ON period, the inter-arrival time of packets follows another Weibull distribution. Video traffic burst model is captured as an MPEG model. For the long-range dependent traffic such as Business LAN Networking, self-similar traffic model is used. The traffic parameters are selected so that the traffic generated in the simulation matches the actual traffic traces.

Table 1: Session Level Traffic Parameters

Traffic Types	Mean Arrival Rate (Sessions/Second)	Mean Holding Time (Seconds)	QoS Class	Peak Uplink Rate (Kbps)
Type 1: Custom Calling Services	λ_1	30	Class 1	64
Type 2: VoIP, medium-quality	λ_2	180	Class 2	16
Type 3: VoIP, high-quality	λ_3	180	Class 2	64
Type 4: Fax	λ_4	180	Class 2	64
Type 5: Video	λ_5	1200	Class 2	384
Type 6: Online Shopping and Ordering	λ_6	1200	Class 2	64
Type 7: Business LAN Networking	λ_7	1200	Class 2	384
Type 8: Data Dissemission	λ_8	1200	Class 3	64
Type 9: Web Access	λ_9	900	Class 4	64
Type 10: Telnet	λ_{10}	600	Class 4	64
Type 11: FTP	λ_{11}	40	Class 5	64
Type 12: Email with Text Only	λ_{12}	60	Class 5	64
Type 13: Email with Image Attachment	λ_{13}	60	Class 5	384
Type 14: Email with File Attachment	λ_{14}	60	Class 5	64
Type 15: Email with Audio Attachment	λ_{15}	600	Class 5	64
Type 16: Email with Video Clip Attachment	λ_{16}	600	Class 5	384

The values for SR_j correspond to the following amount of resource [14]:

- Class 1: peak rate PR_j;
- Class 2: effective rate ER_j;
- Class 3: sustainable rate STR_j, which is between the effective rate & the minimum rate;
- Class 4: minimum rate MR_j, to keep the session running;
- Class 5: 0;

The values for BR_j correspond to the following amount of resource:

- Class 1: 0 (no BoD for real-time Class 1);
- Class 2: 0 (no BoD for real-time Class 2);
- Class 3: ER_j - STR_j;
- Class 4: 0 (best effort based);
- Class 5: 0 (best effort based);

Table 2 shows the list of the 16 traffic types and the statically allocated and booked resources used in the simulation. In the simulation a BTP frame period is assumed to be 192 *ms*. A time slot in each frame corresponds to 384/192 = 2 Kbps. *SATS* (respectively *BATS*) is the number of *Time Slots* corresponding to *SR* (respectively *BR*).

The normalized *offered load* ρ is expressed as:

$$\rho = \frac{\sum_{i=1}^{K} \lambda_i u_i (SATS_i + BATS_i)}{mB} \tag{5}$$

where K is the total number of types of traffic in the system (in this case, 16), λ_i is the mean arrival rate, u_i is the mean holding time, $(SATS_i + BATS_i)$ is the total number of reserved slots for each connection of type i traffic, and mB is the total number of time slots in an uplink MF-TDMA frame. $\lambda_i u_i$ is the average number of concurrent connections of type i and $\lambda_i u_i(SATS_i + BATS_i)$ represents the total time slots requested by type i. Notice that ρ is the load *offered* by the traffic sources. Since CAC may deny admission to some connections, the actual network load will be lower than ρ. Type percentage p_i is defined as the ratio of offered traffic by type i to the total offered traffic in the system and it is expressed as

$$p_i = \frac{\lambda_i u_i (SATS_i + BATS_i)}{\sum_{j=1}^{K} \lambda_j u_j (SATS_j + BATS_j)} \tag{6}$$

Similarly, the normalized *offered load* ρ_i of type i traffic is defined as

$$\rho_i = \frac{\lambda_i u_i (SATS_i + BATS_i)}{mB} \tag{7}$$

Table 2: Traffic Parameters for All Traffic Types

Traffic Types	QoS Class	SATS	BATS	Peak Uplink Rate (Kbps)
Type 1: Custom Calling Services	Class 1	32	0	64
Type 2: VoIP, medium-quality	Class 2	7	0	16
Type 3: VoIP, high-quality	Class 2	29	0	64
Type 4: Fax	Class 2	31	0	64
Type 5: Video	Class 2	142	0	384
Type 6: Online Shopping and Ordering	Class 2	27	0	64
Type 7: Business LAN Networking	Class 2	142	0	384
Type 8: Data Dissemission	Class 3	4	28	64
Type 9: Web Access	Class 4	1	0	64
Type 10: Telnet	Class 4	1	0	64
Type 11: FTP	Class 5	0	0	64
Type 12: Email with Text Only	Class 5	0	0	64
Type 13: Email with Image Attachment	Class 5	0	0	384
Type 14: Email with File Attachment	Class 5	0	0	64
Type 15: Email with Audio Attachment	Class 5	0	0	64
Type 16: Email with Video Clip Attachment	Class 5	0	0	384

4.2 CAC and BoD Simulation Modelling

Based on the integrated CAC and BoD scheme shown in Fig. 3, the following CAC/BoD modeling approach is used.

Step 0: As is shown in Fig. 4, the time scale is divided into equal length time intervals equivalent to a frame duration (192 *ms* in the simulation). Each of these time intervals is referred to as a Computation Cycle. T_{n-1} is referred to as the beginning of the (*n-1*)th BTP Execution Cycle in the SAUs and the *n*th BTP Computation Cycle in the BoD Controller. The BTP is updated periodically as shown in Fig. 5. For the slot allocations of the *nth* BTP, the BTP is updated during the *n*th *Computation Cycle* of the BTP, which starts at time $T_{(n-1)}$-t_4-t_5 and ends at time T_n-t_4-t_5 so that the *n*th BTP can be broadcasted by the satellite. Note that t_4 accounts for downlink propagation delay, t_5 accounts for the SAU slot assignment time, t_3 accounts for delay during BoD Controller Computation phase, and t_{3b} accounts for the time that the BoD controller spends on the assignment of best effort (BE) time slots.

Step 1: A two level traffic generator is used in the simulation: Session Level Traffic Generator, and Burst/Packet Level Traffic Generator.

Step 2: The Session Level Traffic Generator sends the connection requests to CAC. A CAC request is buffered in Queue 3 in Fig. 3 before the CAC Controller can process it. The CAC decides if a connection should be accepted or not. If the connection is accepted, the corresponding Burst/Packet Level Traffic Generator will start to transmit packets.

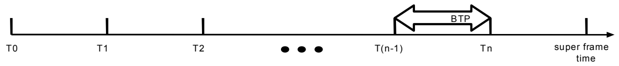
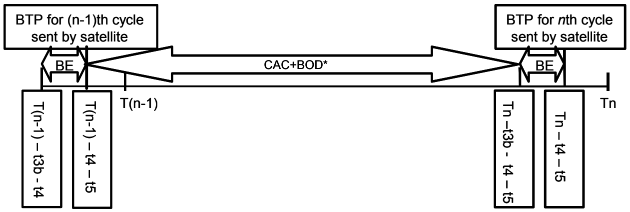


Fig. 4 BTP Computation Cycles



BOD*: static and booked allocation of time slots

Fig. 5 *n*th Computation Cycle

Step 3: The Burst/Packet Level Traffic Generator then generates fixed size packets for each connection *j* that is admitted by CAC. A subqueue in Queue 1 of Fig. 3 is maintained in the SAU to buffer packets for each traffic type. Based on the BTP, the packets are removed from these subqueues based on connection identifiers in the BTP. A packet level traffic generator is implemented as a child processes to its connection process in the simulation so that it can be dynamically generated once a connection is accepted and killed when the connection is released.

Step 4: A *BoD Computation of Needs* process is invoked for all packets in the sub-queues that do not have assigned time slots. The *BoD Computation of Needs* processing time is constant t_1, and the delay associated with *BoD Request Signaling Phase* is constant t_2.

Step 5: When a BoD request for some connection arrives at the BoD controller, the controller performs the following:

a. If the BoD request arrives during the time interval ($T_{(n-1)}$-t_3-t_4-t_5, T_n-t_3-t_4-t_5), then the BoD request will be processed during the *n*th computation cycle; otherwise, it will wait in Queue 2 of Fig. 3 and will be processed in *(n+1)*th computation cycle.
b. When the request is processed, the BoD controller assigns the requested BATS slots to the connection, as discussed in Section 3.
c. If Step b does not satisfy all of the request requirements, then the controller attempts to satisfy the excess requirements for the BoD request by using best effort time slots (BETS) as shown in Section 3.

Step 6: The BTP table is broadcasted to SAUs via the BoD Response Signaling Phase. Assume that the *BoD Controller Computation Phase* processing time is equal to t_3 for each BoD request, and the delay associated with *BoD Response Signaling Phase* is equal to t_4.

Step 7: When the SAU receives the BTP, the *SAU Slot Assignment* removes from each sub-queue in Queue 1 the number of packets that corresponds to the BATSs and BETSs for each connection. The delay associated with *SAU Slot Assignment* is a constant t_5.

The simulation is implemented in OPNET™ [7]. The following performance measures are collected from the simulation:

- *Connection Blocking Probability* --- percentage of number of blocked connections for each type of traffic to the total number of sessions requesting connection (regardless of type) in the system;
- *BoD uplink throughput* --- percentage of used timeslots in a BTP period;
- *SAU buffer sizes for each traffic type*;
- *End-to-end delay for each type of traffic* --- packet delay between the time when a packet enters queue 1 and the time the packet is received by a destination SAU on the downlink air interface.

5. Simulation Results and Discussions

Extensive simulation experiments have been done for multiple spot beams with various traffic mixes. In the following, the simulation results are presented for a selected typical spot beam with some typical traffic scenarios.

5.1 Blocking Probabilities

Notice that Class 5 traffic (Types 11 – 16) is *best effort* traffic, and no CAC is needed. Table 3 shows two different traffic mixes used in the simulation. In Case 1, the offered load for voice traffic (Types 2-3) dominates, and, in Case 2, the offered load for Video, Online Shopping and Ordering, and Business LAN Networking traffic dominates. For each of the two cases, the traffic arrival rates (sessions/second) in terms of the offered load are obtained from Equations (5), (6) and (7) and are shown in Table 4 and Table 5.

Table 3: Traffic Mixes in Simulation Experiments

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
Case 1	0.1	0.25	0.25	0.1	0.05	0.05	0.05	0.05	0.05	0.05
Case 2	0.05	0.05	0.05	0.05	0.2	0.2	0.2	0.1	0.05	0.05

Table 4: Case 1 Traffic Arrival Rates

ρ	0.5	1.0	1.5	2.0	2.5
λ_1	1.10000	2.20000	3.30000	4.40000	5.50000
λ_2	2.09524	4.19048	6.28571	8.38095	10.47619
λ_3	0.50575	1.01149	1.51724	2.02299	2.52874
λ_4	0.18925	0.37849	0.56774	0.75699	0.94624
λ_5	0.00310	0.00620	0.00930	0.01239	0.01549
λ_6	0.01630	0.03259	0.04889	0.06519	0.08148
λ_7	0.00310	0.00620	0.00930	0.01239	0.01549
λ_8	0.01375	0.02750	0.04125	0.05500	0.06875
λ_9	0.19556	0.39111	0.58667	0.78222	0.97778
λ_{10}	0.29333	0.58667	0.88000	1.17333	1.46667

Table 5: Case 2 Traffic Arrival Rates

ρ	0.5	1.0	1.5	2.0	2.5
λ_1	0.55000	1.10000	1.65000	2.20000	2.75000
λ_2	0.41905	0.83810	1.25714	1.67619	2.09524
λ_3	0.10115	0.20230	0.30345	0.40460	0.50575
λ_4	0.09462	0.18925	0.28387	0.37849	0.47312
λ_5	0.01239	0.02479	0.03718	0.04958	0.06197
λ_6	0.06519	0.13037	0.19556	0.26074	0.32593
λ_7	0.01239	0.02479	0.03718	0.04958	0.06197
λ_8	0.02750	0.05500	0.08250	0.11000	0.13750
λ_9	0.19556	0.39111	0.58667	0.78222	0.97778
λ_{10}	0.29333	0.58667	0.88000	1.17333	1.46667

Fig. 6 and Fig. 7 show the steady state connection blocking probabilities for the two traffic mixes respectively. When the total offered traffic load is below 1, the blocking probability for each traffic type is very low, and there is no blocking for Type 2, Type 9 and Type 10 traffic due to their low requested bandwidth, i.e., SATs+BATs. When the total offered load increases beyond 1, the blocking rates for some traffic types are still low. For example, Types 5-8 consistently show small blocking probabilities due to their low offered load (see equation 7). By contrast, Type 1 blocking probability is consistently among the highest when the offered load is high because the Type 1 offered loads are among the highest offered loads. From the simulation, it is found that the connection blocking probability is proportional to the traffic arrival rate by that type and also proportional to its requested bandwidth SATS+BATS but has nothing to do with its call holding time, i.e., *blocking_probability (type i)* $\propto \lambda_i$ * *(SATS_i+BATS_i)*. Fig. 8 shows a sample snapshot of the number of concurrent connections of each traffic type admitted by the CAC for Traffic Mixes Case 1 and $\rho = 1$. For each type *i*, the theoretical number of concurrent connections should be $\lambda_i \mu_i (1-Pb_i)$. The simulation results match the theoretical value very well.

The simulation results show that the GoS for voice traffic in the satellite system falls well within Bellcore's GoS specifications ([5] and [6]), which is well below 3% blocking rate, when the traffic load is reasonable. There can be several solutions to reduce the high blocking probabilities for voice traffic when the traffic load is heavy. For example, one solution is to assign voice connections on a permanent virtual connection basis instead of a dynamic basis. This guarantees bandwidth to be available only for voice connections, and voice does not need to compete with other types of connections for bandwidth. Other solutions are to use reservation or to increase the capacity of the individual carriers. Alternatively, some applications, such as video, can be removed in favor of voice traffic. A final solution is to use compressed 32 Kbps or lower rate voice services. A methodology is presented in [3] that estimate network revenue given the total source load admitted into the network. The accepted source load can be calculated as follows:

$$\rho\_accepted = \frac{\sum_{i=1}^{K} \lambda_i u_i (1 - Pb_i)(SATS_i + BATS_i)}{mB}$$

where Pb_i is the call blocking probability for traffic type *i*. Thus the network revenue can be projected for different scenarios based on the methodology given in [3].

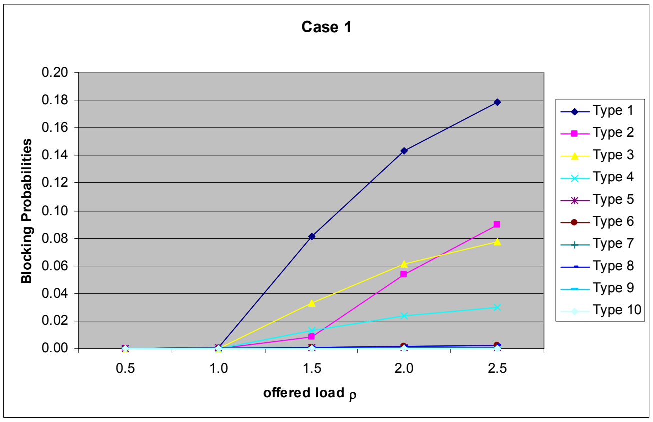


Fig. 6 Traffic Mix Case 1 Blocking Probabilities

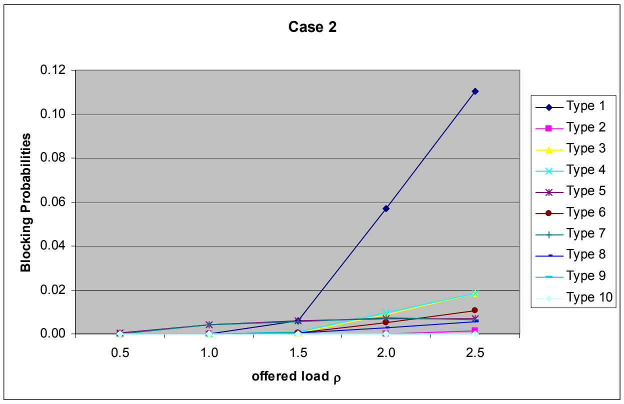


Fig. 7 Traffic Mix Case 2 Blocking Probabilities

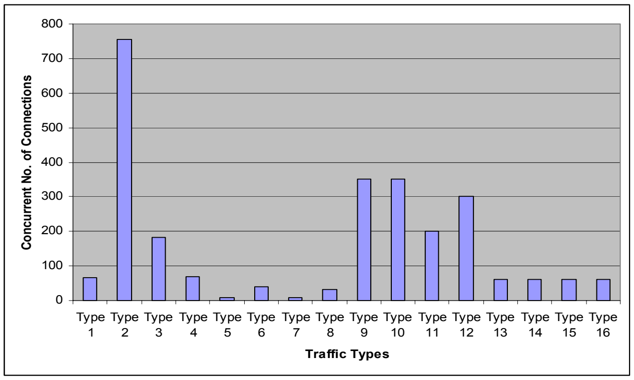


Fig. 8 A Snapshot of the Number of Concurrent Connections for Each Traffic Type in Steady State

5.2 BoD Throughput

Fig. 9 shows the satellite uplink throughput for Traffic Mixes Case 1 and ρ = 1 simulation. Without BoD, the source traffic considers the booked bandwidth the same as the statically reserved bandwidth. As it is shown in Fig. 9, the steady state uplink throughput with BoD is 68%, while the throughput without BoD process is only 53.5%. Hence, the BoD process helps increase the throughput by 25%. For the throughput with BoD, Fig. 9 also shows that the remaining 100%- 67% = 33% bandwidth is completely caused by unused SATS from Class 1, Class 2, Class 3, and Class 4 connections. Due to the statistical fluctuations at the packet level of these connections, the reserved SATS could be more than what is needed from time to time. The unused SATS cannot be made available to other connections via BoD process due to the signaling complexity on long delayed GEO satellite link. The wasted SATS is the price to pay for providing QoS in a GEO satellite network. Therefore, proper dimensioning SATs and BATs can improve the satellite network throughput. But the trade-offs also need to be considered between throughput and end-to-end delay for the traffic.

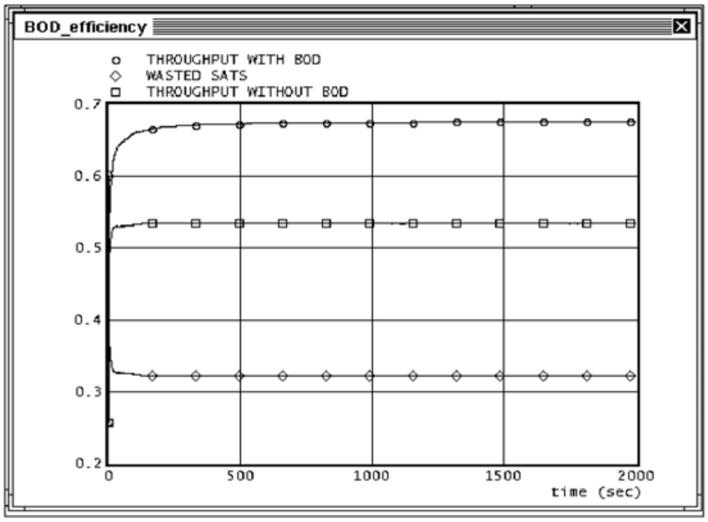


Fig. 9 Satellite Uplink Throughput

5.3 SAU Buffer Size and Packet Delay

Table 6 shows the average, minimum, and maximum buffer occupancy for each traffic type, where the measurement is in number of fixed size (48 bytes) packets. In Table 6, a medium-quality voice connection buffer needs no more than 13 packets (about 5000 bits). The Type 1 traffic has constant buffer occupancy of 21 packets. This constant occupancy is due to the time mismatch among the 64 Kbps Type 1 traffic rate (i.e., one packet every 6 ms), the GEO round trip delay, and the BTP period of *m* ms. The simulation results show that the "one second rule", i.e., the SAU buffer size is equal to the maximum amount of information that a SAU can transmit in one second, is sufficient for the SAU buffer dimensioning for most traffic types. For some Class 5 applications, e.g., Type 13, and Type 16, a "three second rule" is necessary. The buffer sizes from the simulations can serve as a guideline for SAU and satellite on-board memory design.

Table 6: Buffer Occupancy per Traffic Type

Traffic Types	Average Queue Size (packets)	Minimun Queue Size (packets)	Maximum Queue Size (packets)	Buffer Limit (1-second rule) (packets)
Type 1: Custom Calling Services	21	21	21	167
Type 2: VoIP, medium-quality	9	7	13	42
Type 3: VoIP, high-quality	27	15	42	167
Type 4: Fax	6	0	44	1000
Type 5: Video	27	9	47	167
Type 6: Online Shopping and Ordering	17	0	58	1000
Type 7: Business LAN Networking	124	40	256	167
Type 8: Data Dissemission	4	0	11	167
Type 9: Web Access	1	0	5	167
Type 10: Telnet	4	0	11	167
Type 11: FTP	6	0	76	167
Type 12: Email with Text Only	5	0	51	167
Type 13: Email with Image Attachment	493	0	1000	1000
Type 14: Email with File Attachment	7	0	112	167
Type 15: Email with Audio Attachment	84	0	167	167
Type 16: Email with Video Clip Attachment	517	0	1000	1000

Table 7 shows the end-to-end delay for all 16 traffic types which is measured as the SAU (BoD) queuing delay plus the round trip GEO satellite link delay (0.25 seconds). The maximum end-to-end delay for Class 1 (Type 1) traffic is below 400 ms. The mean end-to-end delay for Class 2 and Class 3 (Types 2-8) traffic is between 300 and 500 ms. The average end-to-end delay for Class 4 (Types 9-10) traffic is around 1 second, and its corresponding maximum delay is around 3 seconds, which is acceptable for interactive traffic such as web access and telnet. The average end-to-end delay for Type 11, Type 12, Type 14, or Type 15 is between 1 second and less than 3 seconds, while the average end-to-end delay for Type 13 or Type 16 is around 12 seconds or 17 seconds respectively. The maximum end-to-end delay for Class 5 (Types 11-16) traffic is between 4 seconds and 30 seconds, which is acceptable for the best effort traffic. The simulation results show that the maximum end-to-end delays for Class 1 traffic and the mean end-to-end delays for Class 2 traffic in this satellite system fall below the maximum ITU's QoS specification for GEO-based satellite systems which is 400 milli-seconds for GEO satellite system [9], except for medium-quality VoIP (Type 2) traffic, which is supposed to be a *lower-quality* cheaper service.

Table 7: Delays for Each Traffic Type

Traffic Types	Average Queuing Delay (seconds)	Minimum Queuing Delay (seconds)	Maximum Queuing Delay (seconds)	Average ETE Delay (seconds)	Minimum ETE Delay (seconds)	Maximum ETE Delay (seconds)
Type 1: Custom Calling Services	0.13	0.13	0.13	0.38	0.38	0.38
Type 2: VoIP, medium-quality	0.26	0.20	0.32	0.51	0.45	0.57
Type 3: VoIP, high-quality	0.18	0.10	0.28	0.43	0.35	0.53
Type 4: Fax	0.04	0.00	0.21	0.29	0.25	0.46
Type 5: Video	0.04	0.01	0.06	0.29	0.26	0.31
Type 6: Online Shopping and Ordering	0.12	0.00	0.46	0.37	0.25	0.71
Type 7: Business LAN Networking	0.16	0.06	0.34	0.41	0.31	0.59
Type 8: Data Dissemission	0.24	0.00	0.29	0.49	0.25	0.54
Type 9: Web Access	0.89	0.09	3.13	1.14	0.34	3.38
Type 10: Telnet	0.59	0.06	2.83	0.84	0.31	3.08
Type 11: FTP	1.36	0.25	12.03	1.61	0.50	12.28
Type 12: Email with Text Only	1.13	0.25	8.23	1.38	0.50	8.48
Type 13: Email with Image Attachment	11.65	0.25	16.83	11.90	0.50	17.08
Type 14: Email with File Attachment	1.41	0.25	16.35	1.66	0.50	16.60
Type 15: Email with Audio Attachment	2.40	0.25	3.82	2.65	0.50	4.07
Type 16: Email with Video Clip Attachment	16.48	0.25	27.34	16.73	0.50	27.59

6. Conclusions

Satellite systems are attractive for the transport of broadband and multimedia services. GEO satellites can transport traffic from long distances to the gateways with uniform delay that is independent of terrestrial distances. An important feature of these systems is allowing for a maximum number of simultaneous users, hence, requiring effective MAC layer protocols. This paper proposes and analyzes a MAC layer protocol --- an integrated CAC and BoD algorithm for a GEO-based high capacity broadband satellite network. A modeling and simulation method is developed to evaluate the performance of the integrated CAC and BoD algorithm. The simulation and analysis integrated model is based on realistic traffic scenarios and applications. Using detailed simulations, the developed CAC and BoD scheme is demonstrated to efficiently utilize available bandwidth and to gain high throughput, and also maintain good Grade of Service (GoS) for all the applications. The buffer sizes observed from the simulations can serve as a guideline for access unit and satellite on-board memory design. The end-to-end delays for real-time traffic in the system falls well within ITU's Quality of Service (QoS) specification for GEO-based satellite systems.

References

[1] Abaye, J. Babbitt, B. Best, R. Hu, and P. Maveddat, "Forecasting Methodology and Traffic Estimation for Satellite Multimedia Services," *Proceedings of IEEE ICC'99*, pp.1084-1088, Vancouver, Canada, June 1999.

[2] S. Alouf, E. Altman, J. Galtier, J.-F. Lalande, and C. Touati, "Quasi-optimal Bandwidth Allocation for Multi-spot MFTDMA Satellites", *Proceedings of IEEE INFOCOM05*, March 2005.

[3] J. Babbitt, Y. Qian, and H. Abu-Amara, "Global Traffic Generation, Modeling, and Characterization Methodology," *Proceedings of 8th International Telecommunication Network Planning Symposium*, pp. 321-326, Sorrento, Italy, October 1998.

[4] J. Babbitt, H. Abu-Amara, R. Hu, and Y. Qian, "Traffic Modeling for a High Capacity Multi-bean Satellite Network with On-board Cross-connectivity", *Proceedings of IEEE VTC'2004 Spring*, Milan, Italy, May 17-19, 2004.

[5] Bellcore, Document Number GR-1110, "Broadband Switching System (BSS) Generic Requirements," September 1994.

[6] Bellcore, Document Number SR-NWT-002480, "Broadband Switching System (BSS) Technical Analysis Description," November 1993.

[7] http://www.opnet.com/products/modeler/

[8] A. Jamalipour, *The Wireless Mobile Internet Architectures, Protocols and Services, Chapter13: Satellite in Wireless IP*, John Wiley & Sons Ltd., 2003.

[9] ITU-T, Recommendation G.114, "One-Way Transmission Time," 1996.

[10] R. Mazumdar and C. Rosenberg, "A Game Theoretic Framework for Rate Allocation and Charging of Available Bit Rate (ABR) Connections in ATM Networks," *IFIP Broadband Communications'98*, Stuttgart, March 1998.

[11] B. Fan, R. Tafazolli, and B. G. Evans, "Connection Management for Broadband Mobile Satellite Systems", *IEE Proceedings of Communications*, Vol.150, Issue 4, pp.298-303, August 2003.

[12] H. Peyravi, "Medium Access Control Protocols Performance in Satellite Communications", *IEEE Communications Magazine*, Vol.37, No.3, pp.62-71, March 1999.

[13] C. Rosenberg, "End-to-End Resource Management for ATM On Board Processor Geostationary Satellite Systems," *Proceedings of 4th Ka-Band Utilization Conference*, Venice, Italy, pp. 481-488, November 1998.

[14] K.W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, page 144, Springer-Verlag, London, 1995.

[15] A. Iera, A. Molinaro, P. Pace, and S. Marano, "Multimedia Traffic in Broadband Satellite Networks", Proceedings of IEEE ICC03, Vol.1, pp.428-432, May 2003.