24th AIAA International Communications Satellite Systems Conference (ICSSC) and 4t
11 - 14 June 2006, San Diego, California

AIAA 2006-5362

Flexible and Dynamic Communication Architecture for Lunar Exploration

Hui Zeng[*], Ayan Roy-Chowdhury[*], Nicolas Rentz[*], Michael Hadjitheodosiou[†] and John S. Baras[‡]
HyNet Center, Institute of Systems Research,
University of Maryland, College Park, MD, 20742, USA
E-mail: {zengh, ayan, nrentz, michalis, baras}@umd.edu

We present work related to designing a communication network for lunar exploration. We discuss the important requirements for future space missions that influence the design of this network. We list some characteristics of the network, and highlight important issues and constraints related to performance, cost, and future network evolution. We address issues related to efficiency, fairness, end-to-end delay minimization and Quality-of-Service (QoS) framework in order to enable a flexible access and dynamic mission operation. To provide optimal or near-optimal efficient utilization and fair allocation of bandwidth of the downlink channel while guaranteeing specific QoS requirements for different service classes, we propose two-level (long-term static and short-term dynamic) sharing of a slotted high data rate satellite communication link. The long-term optimal bandwidth allocation is implemented to provide per-stream/per-user QoS guarantee and give the inputs to the next level, short-term dynamic bandwidth allocation. In our time-varying short-term bandwidth allocation with threshold regulation, a small portion of bandwidth is still assigned to all active spacecrafts in advance, but most bandwidth is dynamically allocated by the requests from them, per frame, by solving an optimal timeslot scheduling problem.

I.   Introduction

THE new phase of space exploration involves a growing number of human and robotic missions with varying communication and service requirements. These will include continuous, maximum coverage of areas of concentrated activities, such as in the vicinity of in-space planetary outposts and orbiting missions (single spacecraft or constellations) around the Earth, Moon or Mars. These nodes would be connected back to Earth through a broadband backbone and relay infrastructure. This Space Information Highway would serve the dual role of providing virtual presence to space, mission telemetry and control and coordination between missions but also broadband capability to download collected data back to Earth.

Several network topologies that involve a space component are possible. Most of the proposed topologies are for scientific interplanetary communication, with satellites acting as relays to connect remote networks on distant planets to networks on Earth. The resulting networks form hierarchical hybrid meshes and present interesting challenges to overcome the constraint of long propagation delay, ensure robustness against fluctuations in satellite channel conditions due to atmospheric changes, and to ensure secure communication between users.

In this report we present work related to designing a communication network for lunar exploration. We discuss the important requirements for future space missions that influence the design of this network. We list some characteristics of the network, and highlight important issues and constraints related to performance, cost, and future network evolution. The proposed network can be considered as a prototype for future missions to the Moon and beyond. This network shares similarities with terrestrial wireless network and sensor network architectures. However, the issues related to performance, robustness and security are different due to the long delay over the inter-satellite links, the limited power of the space nodes, the special hardware required to support functionality in space, and very different conditions on the lunar surface. Therefore solutions that are geared towards terrestrial

---

* Research Assistant, HyNet, University of Maryland College Park.
† Assistant Research Scientist, HyNet, Institute for Systems Research, University of Maryland, AIAA Member.
‡ Professor, HyNet, Institute for Systems Research and ECE Department, University of Maryland, AIAA Member.

wireless networks might not be suitable for the interplanetary network we consider. We simulated the lunar mission network in software and provide a few performance results. We also discuss the security issues for space exploration networks and suggest algorithms and protocols that can be implemented for ensuring secure communications while maintaining the performance of such networks.

We discuss the design requirements and characteristics of space exploration networks. Design and performance tradeoffs associated with the lunar mission network are discussed in detail and we focus on the problem of Dynamic Bandwidth Allocation in a network that includes space links.[1]

II. Characteristics of Lunar Exploration Network

In order to design and analyze the space exploration network, we make certain assumptions about the mission characteristics, since the specific details and requirements for such missions are not determined yet. We define the unique constraints and complexities associated with an inter-planetary space network, and subsequently we try to adapt solutions from comparable space-terrestrial networks, while addressing the specific requirements wherever necessary. In our view, a space network will consist of one or more small clusters of wireless networks on the remote planetary surface, which will be connected by long-distance broadband links to heterogeneous terrestrial networks. Some important design considerations for such a space exploration network include the following.

- The number of missions might grow and any mission might evolve, with varying communication and service requirements. Evolution of a mission would impact the size and/or topology of the network.
- On the remote planetary surface, the areas of concentrated activities would require continuous coverage by the satellites.
- There might be orbiting missions of single spacecraft or constellations.
- The long-distance broadband backbone should have the capability to upload mission telemetry and control data to the remote outposts, and download collected mission data to the command centers on Earth. There might also be the requirement for coordination between different missions using the satellite broadband backbone.
- The space network would consist of both stationary and mobile nodes. In case of the latter, we assume that the mobility would be highly predictable.
- The network backbone should be capable of supporting a wide range of data rates – from a few *kilobits per second* in case of command and telemetry traffic to several *gigabits per second* in case of collected science data downloaded to Earth centers.
- The utilization of the network links would be variable in time – there would be periods of idle time or low keep-alive interchanges, followed by periods of full utilization.
- The link delays would vary from a few milliseconds (for example, in case of surface wired/wireless links) to a few seconds or even minutes (for example, in case of inter-planetary link between Earth and a remote planet).
- The link error characteristics would be different between the wired/wireless links in the surface networks, and the space satellite links, requiring different error correction algorithms.

The security requirements might be different in different segments of the overall network, and on the different links, which would influence implementation of a heterogeneous set of security algorithms and policies.

III. Two-level Dynamic Bandwidth Allocation

This problem has been extensively studied for a variety of commercial satellite systems offering service to large numbers of users. In this paper however we focus on the allocation of bandwidth in a space relay network that supports several scientific spacecrafts with a number of different streams on-board sharing a broadband satellite channel to send traffic to the ground. Our system model includes a number of mobile spacecrafts (MS) in Lower Earth Orbit (LEO), a Geo-synchronous (GEO) relay satellite, and the ground network consisting of several ground stations (GS). The downlink channel of the relay satellite is shared by these spacecrafts, which we model as a set of streams with different priority levels going through a common queue and a router. The data will be delivered to the ground station through this relay.

To provide dynamic access with fairness and efficiency, a suitable hybrid-mode Time Division Multiple Access (TDMA) protocol along with frame-wise packet scheduling for bursty data flows was proposed for this network in Ref. 1. It was shown that a carefully designed time-varying bandwidth allocation based on the instantaneous or statistical traffic from all users/flows performs better in terms of throughput and end-to-end delay. However, short-term (time varying) bandwidth allocation alone may cause instability and will have difficulties for providing QoS

guarantees and for managing the long-term (average) behavior of all the users/flows. Hence, we propose a two-level bandwidth allocation in our implemented TDMA scheme.

For a well-coupled framework with per user/flow average bandwidth management, we derive our long-term bandwidth allocation problem formulation from the model discussed by Kelly[2], and draw some ideas from recent work[3, 4, 5]. In addition, for instantaneous bandwidth management, we incorporate ideas from some recent work[6, 7, 8] to formulate the short-term timeslot assignment problem and find the solution for optimal timeslot scheduling.

A. Hybrid-Mode MAC Protocol

In our hybrid-mode TDMA protocol, all active spacecrafts are using the common downlink channel to send packets to the ground station (NCC). There are request sub-channels in this downlink channel and feedback in the control link for bandwidth reservation. The whole data channel is divided into a number of identical sub-channels. A fixed number of these sub-channels are allocated for the static slots and the rest are used for reservation-based slots.

1) Triple Request

Every traffic source will be provided with a guaranteed QoS in terms of its triple request: LR (Lower Resource), TR (Targeted Resource) and UR (Upper Resource).

2) Access Control Algorithm

Reservation mini-slots are used for access requests from new users. In the demands, according to the different requirements, every traffic source will provide its triple request: LR, RR and UR, and its priority level and weight when trying to get access to the broadband satellite network. Certainly, for some types of traffic sources, three parameters might be redundant and therefore could be combined. This framework is similar to the studies presented by Hung[3] and BoD protocol[4], but has a different parametric model.

The access control algorithm will be performed as follows: a new user will be admitted only if the sum of the LRs of all active users is less than or equal to the total bandwidth of the broadband channel. After admitted to the network, every stream will be assigned a new triple: LR, PR (Projected Resource) and UR. The stream will be allocated its PR as a sum of LR and a best-effort share from the rest of the available bandwidth. The best-effort share will be assigned according to fairness and efficiency by solving an optimization problem. We will present the detailed problem formulation and results in the following Sections.

B. Two-Level Bandwidth Allocation

To provide per-user/per-flow QoS guarantees for different users with fairness consideration, an efficient bandwidth allocation process, along with a good admission control algorithm, is necessary. Because of the significant propagation delay, the satellite communications network needs a more adaptive bandwidth allocation and a more systematic admission control. To efficiently use the available channel, it is shown that a well-design time-varying bandwidth allocation based on the instantaneous or statistical traffic of all users/flows has better performance in terms of throughput and end-to-end delay[1]. However, short-term (instantaneous) bandwidth allocation alone may cause instability and will have difficulties for providing QoS guarantees and for managing the long-term (average) behavior of all the users/flows. Besides, the instantaneous and average behavior bandwidth managements need to be well-coupled with each other. Therefore, we propose a two-level bandwidth allocation.

As shown in Fig. 1, the two-level bandwidth allocation is performed by the scheduler at the ground station within the Network Control Center (NCC). To access the channel, a new user, or new flow, first sends a request to the scheduler. After performing the Access Control algorithm mentioned in Section III-A, the scheduler will broadcast its decision to the users. If the user/flow is accepted, a static initial bandwidth allocation is made by the long-term bandwidth allocator. Then the initial allocations are delivered to the short-term bandwidth allocator as control parameters for the next-level scheduling. Under some additional conditions, the long-term bandwidth allocation might be performed and updated by the next level. In the short-term scheduler, according to the continuous bandwidth requests from users/flows, the time-varying bandwidth allocation (or slots assignments) will be obtained and broadcast. This is another reason why we use the triple request model, which gives us more control for bandwidth management.
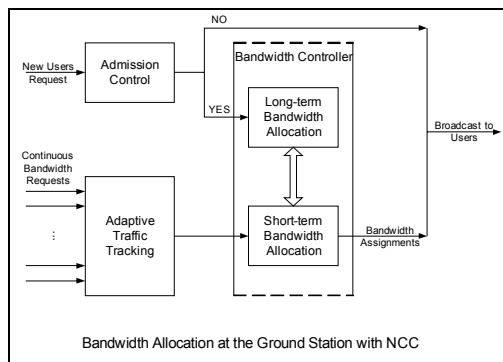

Bandwidth Allocation at the Ground Station with NCC

Figure 1. Two-level Bandwidth Allocation at the Ground Station

IV.   Long-Term Bandwidth Allocation

As we mentioned earlier, to provide certain per-stream (and per-user) QoS guarantee, for an access request from a new stream (or new user), the central scheduler will perform the admission control algorithm to ensure the sum of the contracted bandwidths (rates) of all the users/flows is less than or equal to the targeted bandwidth of the broadband channel. The stream will be allocated its PR as a sum of LR and a best-effort share from the available bandwidth, according to fairness and efficiency by solving an optimization problem. Long-term refers to a relatively long time range compared with the dynamic bandwidth allocation, which is performed per frame or on a multi-frame basis.

A. Kelly's Model

Consider a network with a set L of resources or links and a set I of users (or flows). Let B_l denote the finite capacity of link $l \in L$. Each user has a route r, which is a non-empty subset of L. Define a 0-1 matrix A, where $A_{l,r} = 1$ if $l \in r$, and $A_{l,r} = 0$ otherwise. Suppose that if a rate (bandwidth) x_i is allocated to the user then $U_i(x_i)$ represents its utility. Here, the utility $U_i(x_i)$ is an increasing, strictly concave and continuously differentiable function of x_i over the range $x_i \geq 0$ (i.e., elastic traffic). Also, utilities are additive so that the aggregate utility of rate allocation $x = (x_i, i \in I)$ is $\sum_{i \in I} U_i(x_i)$. Let $B = (B_l, l \in L)$ and $U = (U_i(\cdot), i \in I)$. The rate-control optimization problem is formulated as follows:

SYSTEM (U, A, B):

$$\max \sum_{i \in I} U_i(x_i)$$

$$subj. \ to \ Ax \leq B, x \geq 0. \tag{1}$$

From the convexity of the feasible region for x and the strict concavity of the utility function, it follows that the solution of (1) is unique and proportionally fair. In Ref. 5, it is shown that the optimum solution associated with the logarithm utility function is also a Nash Bargaining Solution. We are interested in proportional fairness, or its variations, because of its simplicity and popularity, although there are also other fairness criteria.

B. Utility Functions Discussion

Denote $c = $ LR, $b = $ UR, $a = $ TR. To incorporate LR (or c), we modify the utility function to $\log(x-c)$. To incorporate TR (or a), we want the optimal solution associated with the modified utility function to have the following property: below its TR, the traffic source is very likely to get more bandwidth assignment if the price is payable; while some way beyond the TR, more bandwidth assignment is not desirable considering the price. In other words, the TR is a measure to describe the starting point of the turning zone for the trade-off between the resource and the price.

$U(x)$	$U'(x)$	$U''(x)$	Region
$\log(x-c)$ $-(x-a)/k$	$1/(x-c)$ $-1/k$	$-1/(x-c)^2$	$k \geq b - c$
$\log(x-c)$ $-(x-a)^2/k$	$1/(x-c)$ $-2(x-a)/k$	$-1/(x-c)^2$ $-2/k$	$k \geq 2(b-a)(b-c)$
$\log(x-c)$ $-\lvert x-a \rvert/k$	$1/(x-c)$ $-1/k$	$-1/(x-c)^2$ except $x=a$.	$k \geq b - c$

Table 1.   Comparison of Utility Functions

Considering simple variations of the logarithm function, we list some candidates in Table 1, where $k > 0$, is the desired attenuation parameter for the designated source.

Recall that the utility function $U(x)$ is an increasing, strictly concave and continuously differentiable function of x over the range $x \in (c, b]$ for elastic traffic. So, to keep the strict concavity, the second derivative of $U(x)$ needs to be negative, which is clearly correct, except for one point $(x = a)$, for the 3rd candidate utility function. To keep the utility function to be increasing, the first derivative of $U(x)$ is nonnegative over the range $x \in (c, b]$, which leads to the regions specified in Table 1 respectively.

For detailed comparisons of the above utility functions, we draw all of them and the truncated logarithm function in Fig. 2, with $c = 1$, $b = 8$, $a = 6$. As shown in Fig. 2, before the point $x = a$, the 2nd line is the most steep one among all the utility functions; while after the point $x = a$, it is the most flat one except the 3rd one. However, the 3rd line is the least steep one before the point $x = a$. Therefore we select the 2nd one, which is the associated utility function

$$\log[(x-c) \times \exp(-(x-a)/k)] = \log(x-c) - (x-a)/k.$$

Also note that after the point $x = a$, the 2nd line and the 4th line coincide with each other.
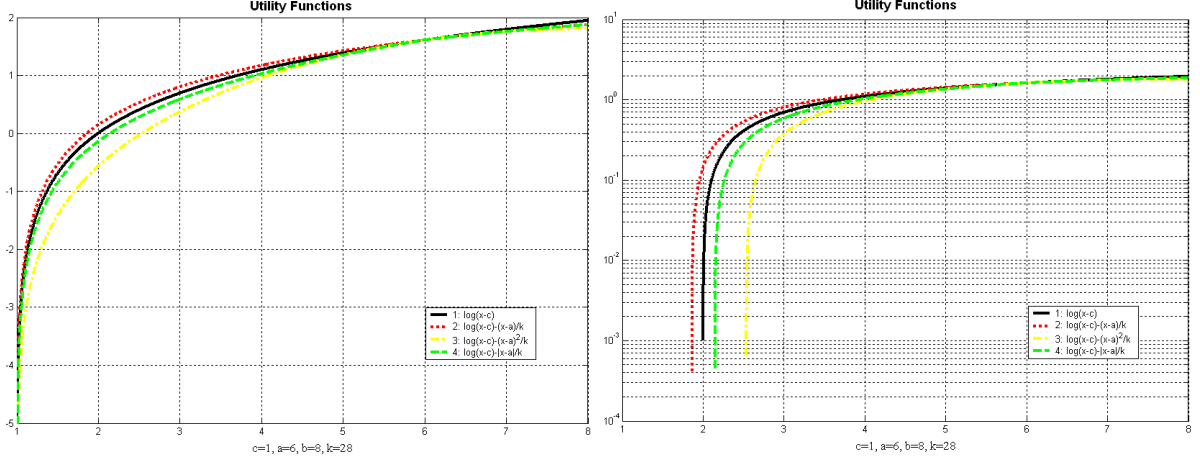
Figure 2. Comparison of the Utility Functions
Left: linear scale, Right: semi-logarithmic scale.

1st : $log(x\text{-}c)$, 2nd: $log(x\text{-}c)\text{-}(x\text{-}a)/k$, 3rd: $log(x\text{-}c)\text{-}(x\text{-}a)^2/k$, 4th: $log(x\text{-}c)\text{-}|x\text{-}a|/k$.

C. Problem Formulation

Assume that the available bandwidth for each node is greater than the sum of the LRs in the same node. If for one specific node this assumption does not hold, the long-term bandwidth allocation problem is trivial. The feasible rate vector space X is defined as:

$$X = \left\{ x\colon x \in R^N, x > LR, x \le UR \text{ and } Ax \le B \right\},$$

and has at least one nonempty interior point. $LR = [LR_1, LR_2, \ldots, LR_N]^T$ is the vector of lower resource requests of N users, $UR = [UR_1, UR_2, \ldots, UR_N]$ is the vector of upper resource requests of N users. Recall that for simplicity we will use c, a, b to denote LR, TR, UR in the equations respectively.

Now our centralized bandwidth allocation problem is:

$$\max \sum_{i=1}^{N} [m_i \cdot \log(x_i - c_i) - \frac{x_i - a_i}{k_i}]$$

$$\text{subj. to: } x_i \ge c_i, x_i \le b_i \tag{2}$$

$$Ax \le B$$

$$i = 1, 2, \ldots, N$$

with the assumption: $Ax_c < B$, where $x_c = [c_1, c_2, \ldots, c_N]^T$. Here, m_i is the weight for the source i, while x_i, c_i, a_i, b_i, k_i are defined for the source i.

Before solving the problem, we will first investigate the objective function, or namely our chosen utility function. According to our earlier discussion, the optimal point x of the above problem has the property that for any perturbation, when $m_i = 1$,

$$\sum_{i \in I} \delta x_i / (x_i - c_i) \le \sum_{i \in I} \delta x_i / k_i ,$$

which is similar to proportional fairness. Then nearby the optimum point, the aggregation of the relative changes of all the sources will be upper-bounded, although not zero. We call this property "pseudo-proportional fairness". When $(k_i, i \in I)$ are large enough, the upper bound will be small, even close to zero. We will see that this property is well-coupled with the short-term time-varying bandwidth allocation in section V.

D. Problem Solution

In (2) with linear constraints and our assumptions, the first-order Kuhn-Tucker conditions are the sufficient and necessary conditions for optimality.[9] Now we consider the Lagrangian form:

$$L(x, \lambda, \beta, \mu) = \sum_{i=1}^{N} [m_i \cdot \log(x_i - c_i) - (x_i - a_i)/k_i] - \sum_{i=1}^{N} \lambda_i (c_i - x_i) - \sum_{i=1}^{N} \beta_i (x_i - b_i) - \sum_{l=1}^{L} \mu_l [(Ax)_l - B_l],$$

$$x_i \ge 0, \lambda_i \ge 0, \beta_i \ge 0, \mu_i \ge 0, i = 1, \ldots, N,$$

where λ_i, μ_i, β_i, i=1, …,N, are slack variables associated with LRs, URs and capacity constraints.

And by considering the sufficient and necessary conditions under our assumptions, we obtain the unique solution as follows:

$$\forall i = 1, ..., N, \quad \forall l = 1, ..., L,$$

$$x_i = c_i + \min\left[(b_i - c_i), \ \frac{m_i}{\frac{1}{k_i} + \sum_{l=1}^{L} \mu_l A_{l,i}} \right] \tag{3}$$

$$Ax \leq B, \ (Ax - B)_l \cdot \mu_l = 0, \ \mu_l \geq 0.$$

We have several useful remarks for the obtained optimal solution:

1. The Lagrange multiplier μ_l is the implied cost of unit flow through link l, or the shadow price of additional unit capacity for link l.
2. For one specific user, the assigned bandwidth is explicitly dependent on the link cost and its own parameters, while implicitly dependent on the users in other nodes.
3. m_i is the weight for user i. The user with higher m_i has better opportunity to get more bandwidth than the user with lower one in the same node.
4. k_i is the desired attenuation parameter for the source i. Assume that k is proportional to *(a-c)* and *(b-c)*, while inversely proportional to *(b-a)*, then we have that $k \propto \frac{a-c}{b-a}(b-c)$. Again, the user with higher k_i has better opportunity to get more bandwidth than the user with lower one.

Now consider asymptotics of k_i, and the relation between our model and the model associated with proportional fairness. Recall that $k_i \geq b_i - c_i, \forall i = 1,...,N$. As all k_i go to ∞, it follows that our objective function

$$\max \sum_{i=1}^{N} [m_i \cdot \log(x_i - c_i) - (x_i - a_i)/k_i]$$

$$\rightarrow \max \sum_{i=1}^{N} [m_i \cdot \log(x_i - c_i)],$$

which is exactly the one with proportional fairness discussed in Ref. 2 and Ref. 5 when $m_i = 1$. As a result, our optimal solution here is exactly the one in Ref. 5 as all k_i go to ∞. Also, with k_i increasing, the attenuation for the source i is decreasing, and then the possibility for the source i to get more bandwidth after certain point is increasing. This just shows the relation between our model, associated with pseudo-proportional fairness, and the model associated with proportional fairness.

V. Short-Term Bandwidth Allocation

In this section we formulate the general time-varying dynamic bandwidth allocation problem for slotted TDMA protocol in a space communications network based on the parameters determined by the long-term bandwidth allocation, and then find its solution.

Let M denote the complete set of all MS, and M_a denote the set of active MS generating traffic. The central ground station performs the burst-level scheduling, i.e., the scheduling only occurs once during each frame, and allocates timeslots to a stream within a frame in a contiguous fashion. The scheduler generates a bandwidth allocation table (BAT) and sends it back to all the MS in the set M_a. Then each active MS knows its assigned timeslots after reading the BAT.

A. Problem Definition

Let N denote the number of the total available data slots. We consider the penalty weights v_{kl}, $k \in M_a$, $l \in C$ for the service class l of the MS k to reflect the QoS and different requirements. For different slot assignments, the aggregated penalty can be calculated with the definition of these penalty weights and the utility function. Our objective for optimal scheduling is to find the solution that minimizes the total penalty.

Every time, before making the BAT, the scheduler collects the updated information, including the number of MS and active MS, the bandwidth demands (D) of the active MS and the penalty weights. To present the different types of traffic, let C denote the set of service classes. Thus, D is a two-dimensional matrix $\{D_{kl}\}$, $k \in M_a$, $l \in C$. D could be directly given by the MS or computed by the collected information from the MS. The latter is more practical while more complicated since an estimation step is a must. The PR, i.e., (x_i, $i = 1, ..., N$), are used as parameters for

calculations. We use a matrix $s = \{s_{kl}\}$ to denote the amount of assigned data slots for service class $l \in C$ of the MS $k \in M_a$.

B. Problem Formulation

Now we can formulate our problem as:

$$\text{Minimize} \quad \sum_{k \in M_a} \sum_{l \in C} v_{kl}\,(D_{kl} - s_{kl})^{+} \tag{4}$$

subject to:

$$s_{kl} \leq \min(U_{kl}, D_{kl}), \qquad k \in M_a,\ l \in C$$
$$s_{kl} \geq L_{kl}, \qquad\qquad\quad k \in M_a,\ l \in C$$
$$\sum_{k \in M_a} \sum_{l \in C} s_{kl} \leq N,$$
$$\forall\, s_{kl} \in \{0,\, 1,\, 2,\, ...,\, N\}$$

If a MS requests more timeslots than the available data slots that can be assigned to it, only a portion of its requested slots will be actually admitted and the residual packets must wait for the next scheduling. Let U_{kl} and L_{kl} denote the upper bound and lower bound of the capacity for the service class $l \in C$ of MS $k \in M_a$, respectively. The LR and UR from the user via the long-term bandwidth allocation could be used directly here. Some mappings from LR and UR are also allowed. The PR, i.e., $(x_i, i = 1, ..., N)$, are used as parameters for bounded assignments. The upper bound of the waiting time (delay) for the service class $l \in C$ of MS $k \in M_a$ is set and used in the decision of penalty weights to guarantee the maximum delay if necessary.

C. Problem Solution

The solution for this linear problem can be found by the following steps:

1. *Sorting*: Sort the penalty matrix $\{v_{kl}\}$ and re-list the elements in a vector V in descending order.
2. *Lower Bound assignment*: Determine the number of data slots for the active MS to satisfy the lower bound requirement.
3. *Additional Amount assignment*: After the 2nd step, assign the available slots to the active MS according to their order in the vector V until the demand or upper bound is fulfilled.
4. *Final assignment*: Allocate timeslots to each stream within a frame in a contiguous fashion.
5. *Create the BAT*.

Our problem formulation is based on two assumptions: 1) The demands D_{kl}, upper bounds U_{kl} and lower bounds L_{kl} are known or could be determined by the scheduler. 2) The penalty weights v_{kl} are very important and distinct. Another concern is that our problem should consider the multi-frame condition in the space communications network with long propagation delay.

We make some improvements for allowing for these concerns. Usually the U_{kl} and L_{kl} can be assigned according to the service requirements of the streams and the practical condition of the whole channel, and can be viewed as two adjustable parameters. Let t_0 and t denote the time the request was created in the MS and processed in the scheduler respectively. Between t_0 and t, the total assigned timeslots for the service class l of the MS k is called "credit" and denoted by $Ct_{kl}(t_0, t)$. Similarly, the total incoming packets between t_0 and t plus the number of packets in queue at time t_0 for the service class l of the MS k is called "debit" and denoted by $Dt_{kl}(t_0, t)$. Then, the "balance", which is $[Dt_{kl}(t_0, t) - Ct_{kl}(t_0, t)]^{+}$, is a very practical determination of the demand D_{kl}. The cumulative bandwidth assignment for one user is upper-bounded by its (TR × frames + U) and lower-bounded by its (TR × frames − L). Notice that we have considered the multi-frame condition for the long propagation delay. The penalty weights v_{kl} are assigned discrete values based on the relations between the "balance" and some prescribed thresholds. When the v_{kl} of some streams are the same, the calculated demands D_{kl} are used to determine their order in the first step.

VI.   Configuration and Simulation Results

A. Network Configuration

We use OPNET to model the MAC protocol and network scenario. For simplification, we use only one relay satellite, and consider four LEOs in its coverage zone with the altitude range of 701-716 km. The RTD is more than 0.6 seconds in this case. We set $M \cdot T_f = 0.68$ seconds and the number of data slots per frame to 64. The total data rate of the link is 200Mbps and we consider a sub-channel with 2Mbps capacity and accordingly 1/100 of the data rate for all spacecrafts. Hence, by combining these parameters and the length of source packet (512 bytes), we set $M = 5$, and get $T_f = 0.1372$ sec. The downlink channel is assumed to be error-free.

We are particularly interested in the total throughput, which in a sense provides an indicator of the level of bandwidth efficiency; and ETE delay, which is the time interval between traffic generation on-board the spacecrafts and its arrival at the ground station.

The network traffic is diverse, i.e., the traffic loads are unevenly distributed among the spacecrafts. Also, the source traffic data rate in a specific spacecraft varies considerably. Those properties match the unpredictable and dynamic traffic pattern in this environment. Also, specific spacecrafts could probably be inactive for a period of time, and an adaptive protocol would be capable to accommodate that.

B. Simulation Results

Our simulation is run for several minutes to reach steady-state. We try to adjust the simulation time to account for the fact that the time for having the spacecrafts inside the common coverage zone under one GEO relay satellite is limited. Note that the spacecrafts are orbiting rapidly (typically their orbit periods are around 95 minutes). Also note that these LEOs have an altitude range of 701-716 km. The propagation delay would vary from 0.24s to 0.30s. The variation of the propagation could be 0.06s, almost half the length of one frame. For more details on the time-varying propagation delay we refer to Ref. 1.

We first present the performance of our hybrid protocol under unevenly distributed traffic loads. Then the ETE-delay and successful throughput performances of a conventional (static) TDMA solution will be compared with our new protocol. As shown in Fig. 3, the ETE delay is ranging from 0.26 seconds to less than 0.5 seconds under different traffic loads with different numbers of active spacecrafts. Considering the large propagation delay, and its large variation due to spacecraft mobility, this is very good. A major portion of the ETE delay in this case is introduced by the error in time synchronization. Suppose, when the spacecraft has packets to transmit, it will calculate the propagation delay, say t_1, and determine whether it owns the current data slot (or control slot). If the answer is yes, then it sends the packet, otherwise it determines the time for its data slot, say t. However, at time t, the propagation delay is t_2, not t_1. The direct result is that the spacecraft misses its data slot. And when the packets are delayed inappropriately in the spacecraft, the ETE delay and the throughput would be affected, obviously. From Fig. 3, we can obviously see that, the less active spacecrafts we have, the better performance the protocol will have.

Now we fix the ratio of expectations of traffic loads of four users (spacecrafts) as 3:2:2:1, and study the performance of our hybrid protocol in terms of ETE delay, successful throughput and the fairness under this special scenario. As shown in Fig. 4, the hybrid protocol outperforms the fixed TDMA in terms of ETE delay and successful throughput. This is because the hybrid protocol can utilize the data slots once belonging to the inactive spacecrafts or spacecrafts with low data rate in a short range, while in the fixed TDMA, these data slots are just wasted. Another reason is that in the hybrid protocol, the data slots are dynamically assigned based on the behavior of their traffic, and therefore achieve better bandwidth utilization. The more bursty and unpredictable the traffic sources are, the more the hybrid protocol will outperform the fixed TDMA solution.

To study the (long-term) fairness among all users, the successful average throughputs of the total channel and every individual user are shown in Fig. 5. As we can see, although
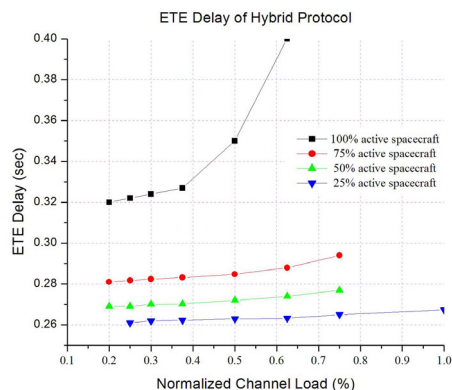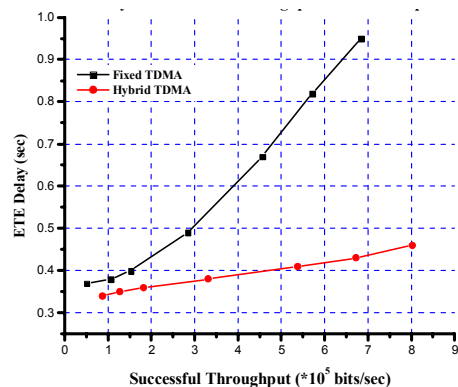

Figure 3. ETE Delay of Hybrid Protocol

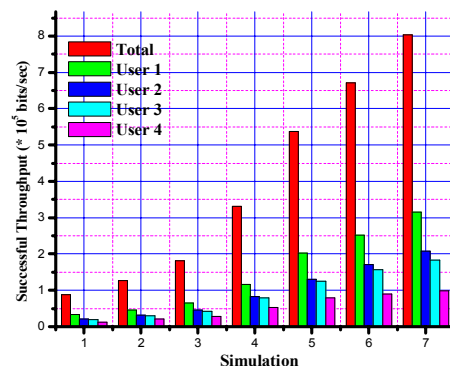**Figure 4. ETE Delay v.s. Throughput**

Figure 5. Fairness among Users

obviously proportional fairness is not achieved, pseudo-proportional fairness is obtained in some sense. In other words, the average share of the channel for every user is close to its proportional portion according to the expectations of traffic loads of four users, i.e., 3/8, 2/8, 2/8 and 1/8 respectively. Since we use the order of the users to break the tie sometimes in the dynamic bandwidth allocation, the average throughput of user 2 is always a little higher than that of user 3, despite the fact that they have same traffic loads.

VII.   Summary

To provide optimal or near-optimal efficient utilization and fair allocation of bandwidth of the downlink channel while guaranteeing specific QoS requirements for different service classes, we propose a two-level (long-term static and short-term dynamic) bandwidth allocation for a slotted TDMA high data rate satellite communication link. The long-term bandwidth allocation is implemented to provide per-stream/per-user QoS guarantee and shape the average behaviors. In our time-varying short-term bandwidth allocation with threshold regulation, a dynamic allocation is performed by solving an optimal timeslot scheduling problem according to the user requests and other parameters. By using simulation, the performance of a suitable Medium Access Control (MAC) protocol with timeslot scheduling is analyzed and compared with that of the existing static fixed-assignment scheme in terms of ETE delay and successful throughput. We also studied the fairness among all users under a special scenario and find that pseudo-proportional fairness is achieved by our hybrid protocol.

Acknowledgments

References

[1] Michael Hadjitheodosiou, Hui Zeng, Alex Nguyen and Brenda L. Ellis, "Flexible Access for A Space Communications Network with IP Functionality", Special Issue of Computer Networks Journal on "Extending the Internet to Space", Vol. 47, Pages 679-700, 2004.

[2] Frank Kelly, "Charging and Rate Control for Elastic Traffic", European Transactions on Telecommunications, Vol. 8, Pages 33-37, January 1997.

[3] Anthony Hung and Marie-Jose Montpetit, et al., "A Framework for ATM via Satellite", Proc. IEEE GLOBECOM, London, UK, 1996.

[4] G. Acar and C. Rosenberg, "Performance Study of End-to-End Resource Management in ATM Geostationary Satellite Networks with On-Board Processing", Space Communications Journal, Invited Paper, Vol. 17, No. 1-3 (2001), pp. 89-106.

[5] Haïkel Yaïche, Ravi R. Mazumdar and Catherine Rosenberg, "A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks", IEEE/ATM Transaction on Network, Vol. 8, No. 5, October 2000.

[6] Ki-Dong Lee, "Throughput-Maximizing Timeslot Scheduling for Interactive Satellite Multiclass Services", IEEE Communications Letters, Vol. 7, No. 6, June 2003.

[7] Marc Emmelmann and Hermann Bischl, "An Adaptive MAC Layer Protocol for ATM-based LEO Satellite Networks", Proc. VTC 2003, 2003.

[8] Paul Mitchell, David Grace, and Tim Tozer, "Analytical Model of Round-Robin Scheduling for a Geostationary Satellite System", IEEE Commun. Letters, Volume: 7, Issue: 11, pg. 546 – 548, Nov. 2003.

[9] Michel Minoux, "Mathematical Programming: Theory and Algorithms." Wiley, Chichester, 1986.