TECHNICAL RESEARCH REPORT

Multicast-aware Power Allocation in Multiple Spot-Beam Satellite Communication Systems

by Gun Akkor, John S. Baras, Michael Hadjitheodosiou

CSHCN TR 2004-19
(ISR TR 2004-38)

The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site http://www.isr.umd.edu/CSHCN/

Multicast-aware Power Allocation in Multiple Spot-Beam Satellite Communication Systems

Gun Akkor, John S. Baras, and Michael Hadjitheodosiou
Electrical and Computer Engineering Department, and
Center for Satellite and Hybrid Communication Networks,
University of Maryland, College Park, MD 20742, USA.
e-mail: {akkor, baras, michalis}@isr.umd.edu

***Abstract*— We address the problem of optimizing resource sharing and flow control in a multiple spot-beam broadband satellite system that supports both unicast and multicast flows. Satellite communication systems, with their wide-area coverage and direct access to large number of users, clearly have an inherent advantage in supporting multicast applications. In order to remain competitive against other broadband technologies, however, next generation satellite systems will be required to support both unicast and multicast flows and offer optimal sharing of system resources between these flows. We show that, in a multiple spot-beam system, a high load variation across spot-beam queues may force lower allocated session rates for active flows, and be perceived as unsatisfactory by potential users when both unicast and multicast flows are active in the system. We propose an optimization framework for balancing the spot-beam queue service rates such that the sum of the rate variances of all active multicast flows is minimized. This is achieved through the re-distribution of system power among spot-beam queues, by taking into account the load on the queues and the channel states. We conclude that it is possible to increase the average session rates of multicast flows by up to 16%, and the rates of unicast flows by up to 4% after this optimization is applied.**

***Index Terms*— System design, multicast delivery, satellite networks, power allocation, mathematical optimization.**

I. Introduction

The role of satellite systems in today's communication infrastructure is changing rapidly. This change is fueled by two main ingredients. The first one is the technological advances in the design of new satellite systems. Next generation satellite communication systems that utilize higher frequency bands, such as the Ka-band, and support spot-beam technology and on-board packet processing are currently under development [1]. These new systems will offer higher data rates and will enable the use of small, low-power, and low-cost user terminals. Therefore, they are likely to become more competitive against other broadband communication solutions in providing integrated voice, data, and multimedia communications.

The second component is the set of new applications, such as on-demand multimedia content delivery, distance learning, and distributed software updates, that have recently emerged in the Internet. These applications are distributed in nature and require concurrent transmission of the same content to multiple users. Satellite communication systems, with their wide-area coverage, direct and ubiquitous access to large number of users, clearly have an inherent advantage in supporting such services.

Despite the potential for multicast content delivery over satellite networks, however, such services remain largely unavailable due to the lack of an incentive to deploy them. From the network service providers' point of view, there will be an incentive to use multicast delivery only if it results in considerable bandwidth savings and allows deployment of new applications. The problem of providing users with an incentive to use multicast delivery is more difficult. From a user's point of view, a high service satisfaction (as perceived speed or performance) is required whether the provider uses unicast or multicast to deliver content. In order to make multicast delivery rewarding to both parties, next generation satellite systems should take into account that both unicast and multicast flows will co-exist in the system, and make sure that system resources are shared optimally between these flows. The latter issue is particularly important, since satellite bandwidth is scarce and satellite systems have to make the most out of the available resources to remain competitive against other broadband technologies.

In this paper, we address this problem from the perspective of resource sharing and flow control in a multiple spot-beam satellite system that supports both unicast and multicast flows. We show that a high load variation across the spot-beam queues may force lower allocated session rates for active flows, and be perceived as unsatisfactory by potential users when both unicast and multicast flows are active in the system. We propose an optimization based-approach to balance the load in the system, and in doing so, take into account that both multicast and unicast flows will co-exist and compete for the system resources.

The rest of the paper is organized as follows. In the next section, we outline the problem in the context of our target satellite system architecture, and identify the key issues. In Section III, we formulate our problem in an optimization framework. Section IV provides the solution, and Section V discusses the analysis framework we have developed for

This material is based upon work supported by NASA under award number NCC8235. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

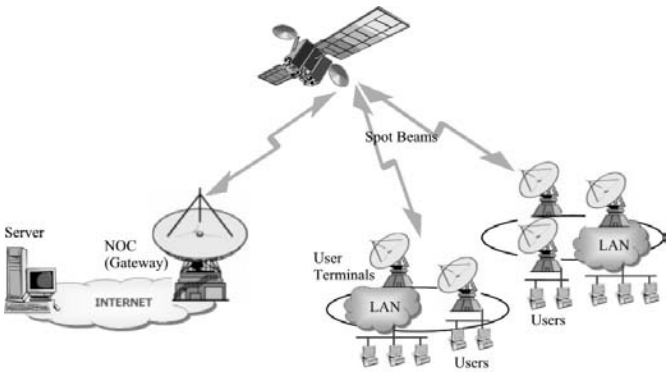

Fig. 1. Satellite communication system architecture. The satellite provides broadband access to users across multiple spot-beam locations.

testing the performance of our approach. In Section VI, we present numerical performance results. Last section concludes the paper and draws attention to future work on this subject.

II. Motivation

In this paper, we look at the problem of resource sharing and flow control in a multiple spot-beam broadband satellite system that supports both unicast and multicast flows. The system we consider is a star topology satellite network, where a Ka-band, geo-synchronous satellite provides broadband services to a large number of users located inside its footprint. In this scenario, users that are equipped with two-way direct communication terminals, access the terrestrial backbone network through a gateway node referred to as the network operations center (NOC). The satellite supports multiple spot-beams and on-board packet switching technologies that allow transmission of data to multiple users in multiple beam locations (Fig. 1).

The choice of the frequency band is not restrictive for our problem setting, but we believe that, next generation systems are moving in the direction of using higher frequency bands, because higher bands offer wider bandwidth segments that are not available at more crowded lower frequency bands. Therefore, we use a Ka-band channel model in our evaluations. The use of multiple spot-beams allows satellite power to be concentrated into densely populated areas, and enables the use of low-power, low-cost user terminals that offer two-way direct communication. It also provides efficient utilization of the available satellite bandwidth by high frequency reuse across spot-beam locations. An on-board processor and switch forward packets to one or more spot-beam queues.

In this multiple spot-beam system, packets of several active flows are queued at the NOC. The NOC forwards the packets to the satellite at a rate limited by the uplink capacity of the system. The on-board processor and switch forward the packets to one or multiple spot-beam queues, duplicating the packets in the latter case. A packet belonging to a unicast flow is forwarded to a single spot-beam queue, corresponding to the spot-beam location, in which the end user resides. In case of a multicast flow, however, receivers of the multicast session may reside in multiple spot-beam coverage areas, and therefore, packets need to be duplicated and forwarded to multiple spot-beam queues on-board the satellite. Therefore, while the packets of a unicast flow affect the load on only one spot-beam queue, in case of a multicast flow, a single session may affect the load on several spot-beam queues. This may have direct implications on the rate each flow is served, as well as the user satisfaction.

At every queue, multiple flows (unicast and multicast) share the total service rate of the queue. The rate-share of a flow belonging to a particular queue depends on the number of flows currently active in the queue, the type of the flows, and the rate allocation policy between different type of flows, — i.e. unicast and multicast. In order to avoid over-flowing of any of the on-board queues, the input rate of a flow at the NOC queue have to be determined by the minimum rate the flow can be served at the spot-beam queues. For a unicast flow, the *maximum sustainable session rate* at the NOC queue is equal to the rate-share of the flow at the spot-beam queue that it has been forwarded to. However, for a multicast flow, the maximum sustainable session rate is equal to the *minimum* of the *supportable* rate-shares the flow gets across multiple spot-beam queues. This requirement would cause all receivers of a multicast session to adjust their rates to this minimum, and would negatively effect user satisfaction if there is a high variation among the supportable session rates.

In this system, a high variation may be the result of several factors, such as the distribution of users across geographical spot-beam locations, uneven downlink channel rates due to climatic variations, and time of the day. In this paper, we propose an optimization-based approach for load balancing across spot-beam queues in order to minimize the rate variance multicast flows experience across multiple spot-beam queues. We show that this type of load balancing could result in higher rate allocations for most active flows, improving the total utilization of the system. In the following section, we describe this approach in an optimization framework and specify the parameters of interest.

III. Problem Formulation

In this system, M on-board spot-beam queues are served by K on-board antennas in a time-divided manner. The downlink transmission is organized into bursts, each of which occupies a fixed time interval. During a burst, an antenna serves only one spot-beam queue. We define the time it takes to serve each spot-beam queue only once with no antenna idling as a *transmission round*. A transmission round can be viewed as a frame of K rows, each corresponding to an on-board antenna, and $L = M/K$ columns, where we assume, without loss of generality, that L is an integer. We denote by $\mathcal{A}_l,\ l = 1, 2, \ldots, L$, as the set of spot-beam queues that are served simultaneously (corresponding to a column of the frame).

The transmission rate r_j of spot-beam $b_j,\ j = 1, 2, \ldots, M$, at the time of its burst interval, depends on the allocated power p_j, and the current channel state s_j, according to a general concave rate-power curve $\mu_j(p_j, s_j)$. For any state s_j of the

downlink channel, rate-power curve represents the rate, under a specific set of coding schemes, that achieves a target bit error rate (BER) as a function of the allocated power. The power levels of all beams satisfy:

$$0 \leq l_j \leq p_j \leq P_{tot}, \quad j = 1, 2, \ldots, M, \quad \text{and} \tag{1}$$

$$\sum_{j \in \mathcal{A}_l} p_j = P_{tot}, \quad l = 1, 2, \ldots, L, \tag{2}$$

where P_{tot} is the total available system power and $\{l_j\}_{j=1}^{M}$ is a set of lower bounds on the power levels of the queues.

A flow (connection) f_i, for $i = 1, 2, \ldots, N$, which is forwarded to spot-beam queue b_j is assigned a rate-share w_{ij} of the service rate of the queue, depending on the load of the queue, and the type of the flows forwarded to it, such that

$$w_{ij} = 0 \qquad \text{if } i \notin \mathcal{B}_j, \tag{3}$$

$$0 < w_{ij} \leq 1 \qquad \text{if } i \in \mathcal{B}_j, \tag{4}$$

$$\sum_{i \in \mathcal{B}_j} w_{ij} = 1 \quad j = 1, 2, \ldots, M, \tag{5}$$

where, $\mathcal{B}_j$ is the set of all flows that are forwarded to the spot-beam queue b_j. Therefore, the packets of flow f_i could be served at a *supportable session rate* of

$$\lambda_{ij} = w_{ij} \cdot r_j = w_{ij} \cdot \mu_j(p_j, s_j), \tag{6}$$

at the spot-beam queue b_j. However, the *maximum sustainable session rate* of the flow at the NOC queue is limited to the minimum rate that the flow could be served across spot-beam queues, — i.e.

$$\lambda_i = \min_{j:\, i \in \mathcal{B}_j} \{\lambda_{ij}\}, \tag{7}$$

in order to avoid overflowing of the spot-beam queues.

For unicast flows, there exists a single spot-beam queue index j for which $i \in \mathcal{B}_j$, corresponding to the beam where the destination user resides. However, for multicast flows, there are several indices for which this may be true. The variation in $\{\lambda_{ij}\}_{j=1}^{M}$ can be minimized by adjusting the service rates of spot-beam queues, — i.e. $\{r_j\}_{j=1}^{M}$. The service rates, in turn, depend on the allocated power levels and the channel states. Therefore, our goal is to minimize this variation by arranging the power level allocated to each queue, subject to a total power constraint, and a set of given channel states. In other words, we would like to find the optimal vector of power levels $\mathbf{p}^* = [p_1^* \ldots p_M^*]$ that would minimize the sum of the rate variances of all multicast flows across spot-beam queues:

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \sum_{i=1}^{N} \sigma_i^2, \tag{8}$$

subject to constraints

$$0 \leq l_j \leq p_j \leq P_{tot} \quad j = 1, 2, \ldots M, \tag{9}$$

$$\sum_{j \in \mathcal{A}_l} p_j = P_{tot} \quad l = 1, 2, \ldots, L, \tag{10}$$

$$\text{given} \quad \mathbf{s} = [s_1 \ldots s_M], \tag{11}$$

where,

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{M} x_{ij} \cdot (\lambda_{ij} - m_i)^2, \tag{12}$$

$$m_i = \frac{1}{N_i} \sum_{j=1}^{M} x_{ij} \cdot \lambda_{ij}, \tag{13}$$

$$x_{ij} = \begin{cases} 1, & \text{if } i \in \mathcal{B}_j \\ 0, & \text{if } i \notin \mathcal{B}_j \end{cases}, \tag{14}$$

$$N_i = \sum_{j=1}^{M} x_{ij}. \tag{15}$$

Note that for unicast flows, $N_i = 1$, and $\sigma_i^2 = 0$. Therefore, unicast flows do not contribute to the cost function, but they affect the solution since they change the total load on the system and consequently the rate-shares of every flow, — i.e. $\{w_{ij}\}$. In the remainder of this paper, the rate-power curve is assumed to be of the form $r_j = \beta(s_j) \cdot p_j, \ \forall j$. This assumption is later validated in Section V. In the next section, we provide the solution to (8).

IV. Solution

When no distinction is made among the spot-beam queues, the simplest assignment would be to set equal power levels for all, such that

$$p_j = \frac{P_{tot}}{K}, \ j = 1, 2, \ldots, M. \tag{16}$$

We call this assignment, the *equal-antenna-share* (EAS) policy and denote it by the vector $\mathbf{p}^{\text{EAS}}$. Given the channel state vector $\mathbf{s}$, the power vector $\mathbf{p}^{\text{EAS}}$ completely determines the service rate of each spot-beam queue and the sustainable session rate of each active flow:

$$r_j^{\text{EAS}} = \min\left(\beta(s_j) \cdot p_j^{\text{EAS}}, r_{max}\right), \quad j = 1, 2, \ldots, M, \tag{17}$$

$$\lambda_i^{\text{EAS}} = \min_{j:\, i \in \mathcal{B}_j} \left\{w_{ij} \cdot r_j^{\text{EAS}}\right\}, \quad i = 1, 2, \ldots, N, \tag{18}$$

where $r_{\max}$ is the maximum system downlink rate determined by the set of available modulation and coding methods. In the remainder of this paper, we use the EAS($\mathbf{p}^{\text{EAS}}, \lambda^{\text{EAS}}$) policy as the basis for comparison.

Equation (17) states that, the EAS policy power assignments may be in excess of the power levels required to achieve the maximum system downlink rate for a given channel state. From (18), we can conclude that the supportable session rates of a multicast flow, which are higher than the minimum in (18) could be reduced without effecting the maximum sustainable session rate of that flow. Consequently, it may be possible to maintain the same session rate at a lower queue service rate, resulting in a lower power level requirement for a given channel state. Combining these two observations, it is possible to calculate the set of *minimum power levels* that will maintain the EAS session rates:

$$p_j^{\min} = \max_{i:\, i \in \mathcal{B}_j} \left\{\frac{\lambda_i^{\text{EAS}}}{\beta(s_j) \cdot w_{ij}}\right\}, \ j = 1, 2, \ldots, M. \tag{19}$$

Observe that $p_j^{\min}$ is always less than or equal to p_j^{EAS} for all spot-beam queues. Therefore, the power difference between the two levels can be re-distributed to other spot-beam queues to possibly improve the session rates of active flows. The power vector with power levels given as in (19) is denoted by $\mathbf{p}^{\text{MIN}}$.

We refer to the solution to (8) as the *balanced-antenna-share* (BAS) policy and denote it by the vector $\mathbf{p}^{\text{BAS}}$. We consider two different solutions to (8). In the first case, which we will refer to as the BAS-I policy, the lower bounds on the power levels are set to zero for all beams — i.e. $l_j = 0,\ \forall j$. The total available system power is re-distributed among the spot-beam queues. This setting allows some queues to get power level assignments that are lower than the minimum power levels given in (19). Consequently, some flows may be served at rates lower than their rates under the EAS policy. In this case, the fairness of the algorithm becomes an issue, since it does not have control over which flow rates are reduced as a result of the optimization performed. In the second case, which we will refer to as the BAS-II policy, we set lower bounds on the power levels, such that no flow gets a lower sustainable session rate than their rates under the EAS policy — i.e. set $l_j = p_j^{\min}\ \forall j$. This choice, however, restricts the optimization space since only the excess power could be distributed, but guarantees that the optimization will return rates that are no worse than the EAS policy rates for all active flows. We will elaborate more on the fairness and the effects of having lower bounds in Section VI. In the next two sections, we present the solution under both policies.

A. Solution under BAS-I policy

Before proceeding with the solution, we classify spot-beam queues into three sets: (i) $\mathcal{E}$, the set of empty queues for which $\mathcal{B}_j = \emptyset$, (ii) $\mathcal{U}$, the set of spot-beam queues with only unicast flows, and (iii) $\mathcal{U}^c$, the set of beam queues with both unicast and multicast flows. Based on this classification, the solution power vector can be re-arranged, without loss of generality, as

$$\mathbf{p}^{\text{BAS-I}} = [\mathbf{p}_{\mathcal{E}}^{\text{BAS-I}} | \mathbf{p}_{\mathcal{U}}^{\text{BAS-I}} | \mathbf{p}_{\mathcal{U}^c}^{\text{BAS-I}}]^{\text{T}}. \tag{20}$$

Under BAS-I policy, empty spot-beam queues are removed from the calculation by setting $p_j = 0,\ \forall j \in \mathcal{E}$. The queues with only unicast flows are excluded from the calculations as well, because, independent of their service rates, the unicast flows that are forwarded to such queues will have zero rate variance. Therefore, we assign minimum power levels for such queues in order to guarantee that the session rates of the unicast flows are no worse than the EAS session rates, and set $p_j = p_j^{\min},\ \forall j \in \mathcal{U}$.

Having determined the power levels for the first two components of the solution vector, where $\mathbf{p}_{\mathcal{E}}^{\text{BAS-I}} = \mathbf{0}$ and $\mathbf{p}_{\mathcal{U}}^{\text{BAS-I}} = \mathbf{p}_{\mathcal{U}}^{\text{MIN}}$, the values for the power vector $\mathbf{p}_{\mathcal{U}^c}^{\text{BAS-I}}$, of cardinality $|\mathcal{U}^c|$ can be calculated as

$$\mathbf{p}_{\mathcal{U}^c}^{\text{BAS-I}} = \mathbf{G} \cdot \alpha + \mathbf{Z} \cdot \mathbf{d}, \tag{21}$$

where

$$\mathbf{G} = \mathbf{X}^{\text{-1}} - \mathbf{X}^{\text{-1}} \cdot \mathbf{B}^{\text{T}} \cdot (\mathbf{B} \cdot \mathbf{X}^{\text{-1}} \cdot \mathbf{B}^{\text{T}})^{-1} \cdot \mathbf{B} \cdot \mathbf{X}^{\text{-1}}, \tag{22}$$

$$\mathbf{Z} = \mathbf{X}^{\text{-1}} \cdot \mathbf{B}^{\text{T}} \cdot (\mathbf{B} \cdot \mathbf{X}^{\text{-1}} \cdot \mathbf{B}^{\text{T}})^{-1}, \tag{23}$$

and $\mathbf{X}$ is a $|\mathcal{U}^c|$x$|\mathcal{U}^c|$ matrix, $\mathbf{B}$ is a Lx$|\mathcal{U}^c|$ matrix, $\mathbf{d}$ is a Lx1 vector, and α is a $|\mathcal{U}^c|$x1 vector of Lagrangian multipliers.

The matrix $\mathbf{X}$ is given by $(\mathbf{A} - 2 \cdot \mathbf{V}^{\text{T}} \cdot \mathbf{V})$, where $\mathbf{A}$ is a $|\mathcal{U}^c|$x$|\mathcal{U}^c|$ diagonal matrix with entries

$$a_{jj} = \sum_{i=1}^{N} \frac{2}{N_i} \cdot \beta(s_j)^2 \cdot w_{ij}^2,\ \forall j \in \mathcal{U}^c, \tag{24}$$

and $\mathbf{V}$ is a Nx$|\mathcal{U}^c|$ matrix with entries

$$v_{ij} = \frac{1}{N_i} \cdot \beta(s_j) \cdot w_{ij},\ \forall j \in \mathcal{U}^c, \tag{25}$$

and $i = 1, 2, \ldots, N$. The entries of the matrix $\mathbf{B}$ represents the mapping of spot-beam queues to antenna groups and given by

$$b_{lj} = \begin{cases} 1, & \text{if } j \in \mathcal{A}_l \\ 0, & \text{if } j \notin \mathcal{A}_l \end{cases},\ \forall j \in \mathcal{U}^c, \tag{26}$$

and $l = 1, 2, \ldots, L$. The vector $\mathbf{d}$ represents the remaining power available for distribution to the spot-beam queues in set $\mathcal{U}^c$ following the power assignments to queues in set $\mathcal{U}$, and given by

$$d_l = P_{tot} - \sum_{j \in (\mathcal{U} \bigcap \mathcal{A}_l)} p_j^{\text{MIN}}, \tag{27}$$

for $l = 1, 2, \ldots, L$. In the solution of (21), a non-zero Lagrangian multiplier implies that the corresponding power level must be zero, and we have p_j strictly greater than zero when the multiplier α_j vanishes in (21).

The service rate vector $\mathbf{r}^{\text{BAS-I}}$ is determined by $\mathbf{p}^{\text{BAS-I}}$ and the channel state vector as

$$r_j^{\text{BAS-I}} = \min(\beta(s_j) \cdot p_j^{\text{BAS-I}}, r_{\max}), \tag{28}$$

for $j = 1, 2, \ldots, M$.

B. Solution under BAS-II policy

Following a classification similar to that of the previous case, we assign $\mathbf{p}_{\mathcal{E}}^{\text{BAS-II}} = \mathbf{0}$, and $\mathbf{p}_{\mathcal{U}}^{\text{BAS-II}} = \mathbf{p}_{\mathcal{U}}^{\text{MIN}}$. Under BAS-II policy, all queues that are not empty are guaranteed the minimum power level assignment given in (19). Therefore, we start the solution with a base power vector, $\mathbf{p}^{\text{BASE}}$, such that

$$\mathbf{p}^{\text{BASE}} = [\mathbf{0} | \mathbf{p}_{\mathcal{U}}^{\text{MIN}} | \mathbf{p}_{\mathcal{U}^c}^{\text{MIN}}]^{\text{T}}. \tag{29}$$

Total power used by each subgroup of spot-beam queues under this assignment is given by

$$d_l^{\text{BASE}} = \sum_{j \in \mathcal{A}_l} p_j^{\text{BASE}}, \tag{30}$$

and the remaining power for redistribution is equal to

$$d_l = P_{tot} - d_l^{\text{BASE}}, \tag{31}$$

for $l = 1, 2, \ldots, L$. Therefore, if $\exists l$ such that $d_l = 0$, there remains no additional power to distribute to spot-beam queues in $\mathcal{A}_l$ while satisfying the minimum requirement, and those queues will have to remain at their minimum power level assignments. Based on this observation, we can further classify the queues in $\mathcal{U}^c$ as those in set

$$\mathcal{H} \triangleq \bigcup_l \{j : j \in (\mathcal{U}^c \cap \mathcal{A}_l) \text{ and } d_l > 0\} \tag{32}$$

and its complement, $\mathcal{H}^c$ such that $\mathcal{U}^c = \mathcal{H} \cup \mathcal{H}^c$. Also, let $\mathcal{L}$ be defined as

$$\mathcal{L} \triangleq \{l : d_l > 0\}. \tag{33}$$

Then, the power level assignments for $\mathbf{p}_{\mathcal{H}^c}^{\text{BAS-II}} = \mathbf{p}_{\mathcal{H}^c}^{\text{MIN}}$, and the power level assignments for $\mathbf{p}_{\mathcal{H}}^{\text{BAS-II}}$ of cardinality $|\mathcal{H}|$ is given by

$$\mathbf{p}_{\mathcal{H}}^{\text{BAS-II}} = \mathbf{p}_{\mathcal{H}}^{\text{MIN}} + \mathbf{G} \cdot \alpha + \mathbf{Z} \cdot \mathbf{d}, \tag{34}$$

where G and Z are as defined in (22) and (23), respectively. $\mathbf{X}$ is a $|\mathcal{H}|$x$|\mathcal{H}|$ matrix, $\mathbf{B}$ is a $|\mathcal{L}|$x$|\mathcal{H}|$ matrix, $\mathbf{d}$ is a $|\mathcal{L}|$x1 vector, and α is a $|\mathcal{H}|$x1 vector of Lagrangian multipliers. Equation (34) represents the power that can be distributed in addition to minimum power assignments.

The matrix $\mathbf{X}$ is given by $(\mathbf{A} - 2 \cdot \mathbf{V}^{\text{T}} \cdot \mathbf{V})$, where $\mathbf{A}$ is a $|\mathcal{H}|$x$|\mathcal{H}|$ diagonal matrix with entries,

$$a_{jj} = \sum_{i=1}^{N} \frac{2}{N_i} \cdot \beta(s_j)^2 \cdot w_{ij}^2, \ \forall j \in \mathcal{H}, \tag{35}$$

and $\mathbf{V}$ is a Nx$|\mathcal{H}|$ matrix with entries,

$$v_{ij} = \frac{1}{N_i} \cdot \beta(s_j) \cdot w_{ij}, \ \forall j \in \mathcal{H}, \tag{36}$$

and $i = 1, 2, \ldots, N$. The entries of the matrix $\mathbf{B}$ represents the mapping of spot-beam queues to antenna groups and are given by

$$b_{lj} \quad = \quad \begin{cases} 1, & \text{if } j \in \mathcal{A}_l \\ 0, & \text{if } j \notin \mathcal{A}_l \end{cases}, \ \forall l \in \mathcal{L}, \ \forall j \in \mathcal{H}. \tag{37}$$

The vector $\mathbf{d}$ represents the remaining power available for distribution to the spot-beam queues in set $\mathcal{H}$ following the minimum power assignments to the queues and is given by (31) for all $l \in \mathcal{L}$. In the solution of (34), a non-zero Lagrangian multiplier implies that the corresponding power level must remain at the minimum power level, and we have $p_j > p_j^{\min}$ when the multiplier α_j vanishes in (34). The final solution vector is represented as

$$\mathbf{p}^{\text{BAS-II}} = [\mathbf{0} | \mathbf{p}_{\mathcal{U}}^{\text{MIN}} | \mathbf{p}_{\mathcal{H}}^{\text{BAS-II}} | \mathbf{p}_{\mathcal{H}^c}^{\text{BAS-II}}]^{\text{T}}. \tag{38}$$

The service rate vector $\mathbf{r}^{\text{BAS-II}}$ is determined by $\mathbf{p}^{\text{BAS-II}}$ and the channel state vector as

$$r_j^{\text{BAS-II}} = \min(\beta(s_j) \cdot p_j^{\text{BAS-II}}, r_{\max}), \tag{39}$$

for $j = 1, 2, \ldots, M$.

In the next section, we describe our analysis framework for evaluating the effectiveness of this approach.

V. Evaluation

In order to evaluate the effectiveness of our approach, we first have to define several components that directly affect its performance. The first component is the *rate-power curve* that determines the rate that achieves a target BER, given the allocated power level and the channel state. The next component is the *channel model* that the channel states are based up on. In order to realistically reflect the distribution of flows across spot-beam queues and to determine the queue-antenna mappings, we have to describe the *spot-beam configuration* of our architecture. Lastly, we have to determine the *rate-allocation policy* between the unicast and multicast flows that share the same spot-beam queue. The following sections describe these components in detail.

A. The rate-power curve

The rate-power curve is based on the following link power-budget calculation adapted from an application [1]–[3] for a commercial satellite system. For a given transmit power P_t in decibel Watts (dBW), the Equivalent Isotropically Radiated Power (EIRP) for the antenna system in dBW is given by

$$\text{EIRP} = P_t + G_t - L_t, \tag{40}$$

where G_t and L_t are the antenna gain, and the losses in the transmitting equipment in decibels (dB), respectively. The losses due to signal propagation through the atmosphere and rain attenuation are calculated as

$$L_o = L_p + L_r, \tag{41}$$

where, L_p and L_r are the losses due to propagation, and rain attenuation, respectively, both in dB. Then, the ratio of signal power to noise power spectral density in decibel Hertz (dBHz) follows as

$$C/N_o = \text{EIRP} - L_o + G/T - k, \tag{42}$$

where G/T in decibels per Kelvin (dB/K) is called the *figure of merit* of the receiver determined by the antenna gain G (dB) and its overall noise temperature T in Kelvin (K), and k is the Boltzmann constant in dBW/K/Hz. For a bit rate of R_b in dBHz, the ratio of bit energy to noise power density becomes

$$E_b/N_o = C/N_o - R_b \text{ in dB}. \tag{43}$$

The rain attenuation becomes substantial at Ka-band frequencies, and is the most important factor. Therefore, we assume that — all other effects remaining constant — we can express the rate as a function of the transmit power P_t and the rain attenuation level L_r for a given E_b/N_o value that guarantees a target BER for a given coding and modulation scheme. Consequently, one can rewrite (43), to determine the rate that achieves the target BER for a given power and rain attenuation level:

$$R_b = P_t + \beta(L_r), \tag{44}$$

where $\beta(L_r) = G_t - L_t - L_p + G/T - k - E_b/N_o - L_r$. It is possible to express (44) in linear terms:

$$R_b = \beta(L_r) \cdot P_t \text{ in bps}. \tag{45}$$

G_t (dB)	L_t (dB)	L_p (dB)
46.50	0.50	210.75
G/T (dB/K)	k (dB/K/Hz)	E_s/N_o (dB)
16.37	-228.60	3.56

TABLE I

NUMERICAL VALUES FOR LINK-BUDGET PARAMETERS AS TAKEN FROM REFERENCE [3]

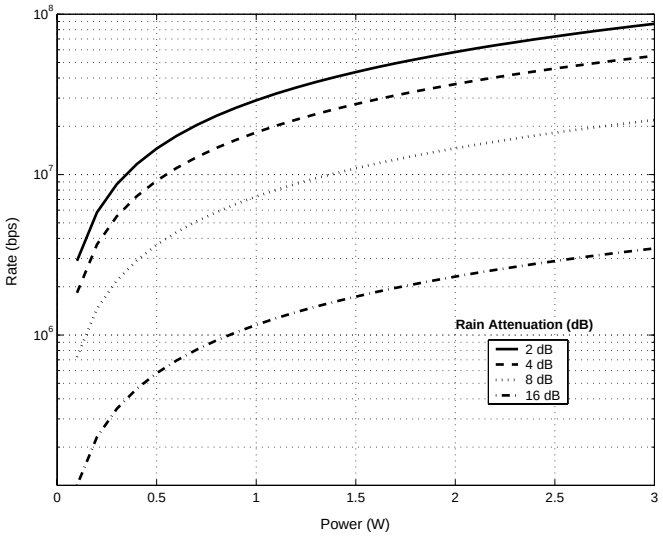

Fig. 2. Rate-power curves for different rain attenuation levels.

We will use (45) in calculating the rate-power relationship for a given rain attenuation level of the channel. Fig.. 2 shows rate-power relationship for different levels of rain attenuation. In this paper, we assume that rate is a continuous function of power, even though, in real systems, not all rates are achievable depending on the set of modulation and coding schemes available for implementation. The numerical values for the link-budget parameters are given in Table I.

B. Channel model

In order to determine the rain attenuation levels for the Ka-band channel, we use a model that is based on the simulator developed at DLR (German Aerospace Center), Institute for Communications and Navigation [4]. The model is based on specific channel model parameters from the DLR measurement campaign carried out at Oberpfaffenhofen near Munich, Germany, in the years 1994 till 1997 with the 40 GHz beacon of the Italian satellite ITALSAT. The channel simulator generates a time-series of attenuation, and calculates the cumulative distribution of attenuation. It is also possible to extract the probability of being in a fade exceeding a given duration and exceeding a fading depth given as parameter. The simulator generates a time-series with 68 seconds resolution. Each attenuation level sample in decibels is input to (44), which through the link-budget calculation gives the downlink rate as a function of allocated antenna power. Fig. 3 shows a sample realization of the rain attenuation time series and the

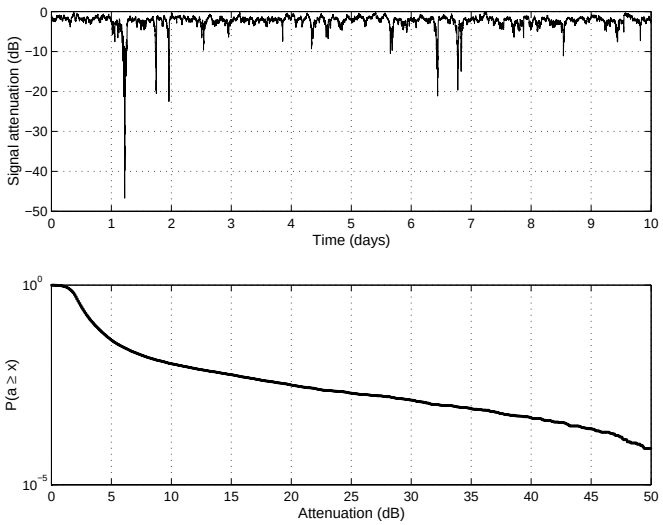

Fig. 3. A sample attenuation time series and the cumulative distribution function of rain attenuation.

corresponding cumulative distribution function for the channel model simulator.

C. Beam and antenna configuration

In order to evaluate the performance of our approach, we need to create unicast and multicast flows between the NOC and the spot-beam locations. However, the number of unicast and multicast flows forwarded to each spot-beam location and the distribution of the multicast users across these locations should reflect the possible load imbalance in a real multiple spot-beam satellite system. Therefore, we first consider the beam locations and the antenna assignments of a geostationary satellite proposed for a commercial satellite system [2], [3]. Fig. 4 shows the approximate locations of the $M = 48$ spot-beams in two polarizations over the United States for this system as indicated by 24 circles. In each circle, the upper and lower identifiers denote the left- and right-polarized spot-beam signals, respectively.

This 48 spot-beams share the access to $K = 4$ on-board antennas. The antenna assignments are as shown in Table II. Next, based on the approximate geographical area covered by each spot-beam, we have calculated the approximate population illuminated by each spot-beam, using the most recent U.S. Census Data [5]. Assuming that a flow f_i is more likely to be forwarded to spot-beam queue b_j if the spot-beam illuminates a larger fraction of the total population, we calculated the probability distribution plotted in Fig. 5. This distribution gives the probability of a flow being forwarded to a spot-beam for all 48 spot-beams and is used to create flows between the NOC and the spot-beam locations.

D. Rate allocation policy

Finally, we have to determine how the service rate of each spot-beam queue is shared among the unicast and multicast flows forwarded to the beam. The policy determines how multicast flows are treated compared to unicast flows sharing

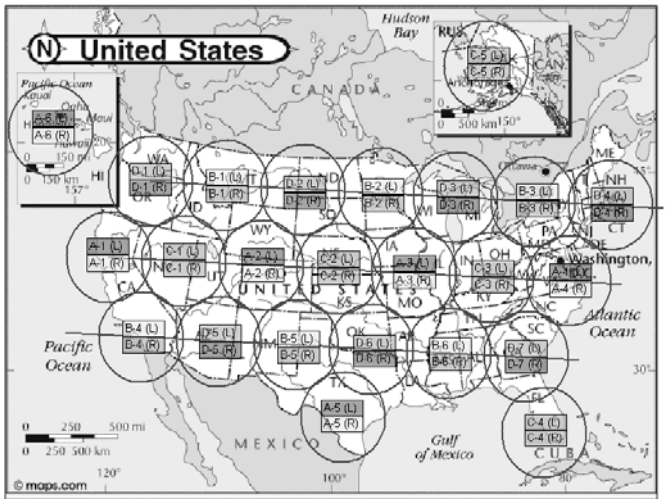

Fig. 4. Locations of the 48 beams in two polarizations over the United States for the example satellite system

ANT1	ANT2	ANT3	ANT4
D1-L	D1-R	B1-L	B1-R
D2-L	D2-R	B2-L	B2-R
D3-L	D3-R	B3-L	B3-R
D4-L	D4-R	B4-L	B4-R
D5-L	D5-R	B5-L	B5-R
D6-L	D6-R	B6-L	B6-R
D7-L	D7-R	A1-R	C1-L
C1-R	A1-L	A2-R	C2-L
C2-R	A2-L	A3-R	C3-L
C3-R	A3-L	A4-R	C4-L
C4-R	A4-L	A5-R	C5-L
C5-R	A5-L	A6-R	A6-L

TABLE II

SPOT-BEAM VS ANTENNA ASSIGNMENTS

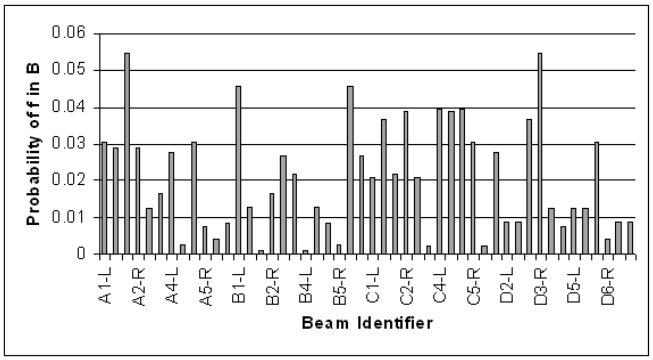

Fig. 5. Connection probability distribution

the same bottleneck, in this particular case, the same spot-beam queue. In [6], authors propose a policy that allocates resources as a logarithmic function of the number of users downstream of the bottleneck, and show that it achieves the best tradeoff between user satisfaction and fairness among unicast and multicast flows. In this paper, we adopt the same policy.

The rate-share w_{ij} of a flow f_i in spot-beam queue b_j is determined by n_{ij}, which is the number of receivers of the flow that resides in the area illuminated by the beam:

$$w_{ij} = \begin{cases} 0, & \text{if } n_{ij} = 0 \\ \frac{1+\log(n_{ij})}{\sum_{i \in \mathcal{B}_j} 1+\log(n_{ij})}, & \text{if } n_{ij} \neq 0 \end{cases} \quad (46)$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, M$.

In the next section, we calculate the optimal power levels of all spot-beam queues and the maximum sustainable rates of every flow under BAS policy and compare our results to the values under EAS policy.

VI. RESULTS

In this section, we will present numerical results on the performance of our approach. The results on BAS policy are given in comparison to the performance under the EAS policy — i.e when power levels are equally distributed. In each case, the system is loaded with L_u unicast connections, and L_m multicast connections that are generated according to the distribution function given in Fig. 5 between the NOC and the spot-beam locations. The multicast group size G_m is assumed to be distributed log-normally such that mean and standard deviation of $\log(G_m)$ is $\log(25)$ and 0.5, respectively. In the numerical results, the number of active unicast connections is set $L_u = 250$, such that in the absence of multicast connections, and under perfect channel conditions, the session rate of a unicast connection forwarded to the most crowded beam queue is approximately $5 - 6$Mbps. The maximum downlink rate is set to $r_{\max} = 92$Mbps. Each set of calculations are repeated 100 times to obtain the statistical results. In the following sections, we discuss our results under both BAS-I and BAS-II policies.

A. Results under BAS-I policy

In this section, we start looking at the results under BAS-I policy. The first set of results looks at the performance of the algorithm for a fix number of unicast and multicast connections while channel conditions change over time. In this scenario, there are $L_u = 250$ active unicast connections, and $L_m = 30$ active multicast connections. At every unit time, the channel states for all 48 spot-beams are sampled and power distribution levels under the BAS-I policy are re-calculated. Using the rate-power curve, the service rates of all spot-beam queues, and the sustainable session rates of all active flows are calculated for each time instance. In Fig. 6(a), we plot the percent change in the sustainable session rates of all active flows averaged over the test duration of $T = 100$ time units,

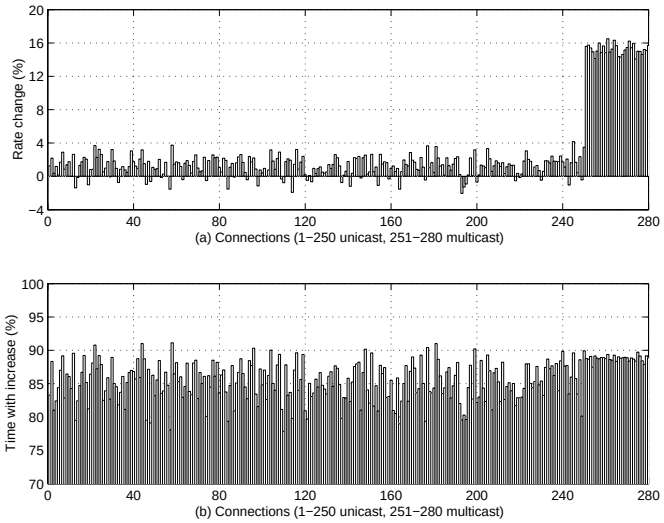

Fig. 6. (a) Percent rate change in the sustainable session rates of all active flows averaged over the test duration under BAS-I policy, compared to the session rates under EAS policy. (b) Percent of total test time BAS-I session rates are equal to or better than EAS session rates for all active flows.

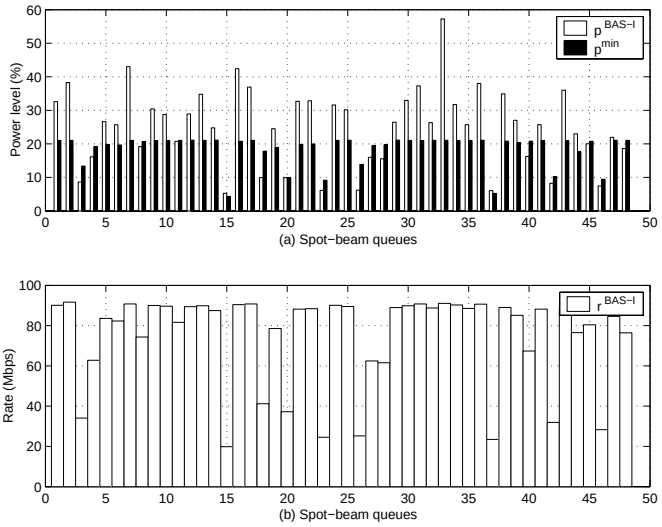

Fig. 7. (a) Average power assigned to each spot-beam queue as percentage of total system power over the test duration under BAS-I policy. (b) Average service rate of the spot beam queues over the test duration under BAS-I policy.

given by

$$\overline{\eta}_i^{\text{BAS-I}} = 100 \cdot \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\lambda_i^{\text{BAS-I}}[t] - \lambda_i^{\text{EAS}}[t]}{\lambda_i^{\text{EAS}}[t]} \right). \qquad (47)$$

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

We observe that all of the active multicast connections and 85% of all active unicast connections are served, on the average, at higher session rates compared to the EAS policy case. The multicast flows experience an average increase of up to 16% in their sustainable session rates, while unicast flows experience more moderate gains of up to 4%, with some having a lower average rate than their EAS rates. There are two factors behind this behavior. First, the optimization policy tries to minimize the rate variance experienced by all multicast flows, without taking into account the rates of the unicast flows sharing the same queues as the multicast flows. As a result, multicast flows benefit the most from the re-arrangement of power levels across spot-beam queues. Therefore, the fairness of the BAS-I policy at a per-flow level is an issue, even though, the net system throughput is increased.

Secondly, BAS-I policy allows power levels to go to zero, therefore, the service rates of some spot-beam queues drop down to levels that are lower than their EAS rates at the end of the optimization. Consequently, the flows incident to them have lower session rates. As a result, the instantaneous rate of an active flow may drop down to a level lower than the EAS rate, even though the average rate of the flow remains higher than the EAS rate. Therefore, it is important to look at the percentage of total time, the sustainable session rates of all active flows remain at a level equal to or higher than their EAS session rates. In Fig. 6(b), we look at this metric given by

$$\overline{\tau}_i^{\text{BAS-I}} = 100 \cdot \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\left(\lambda_i^{\text{BAS-I}}[t] \geq \lambda_i^{\text{EAS}}[t]\right), \qquad (48)$$

for $i = 1, 2, \ldots, N$, where $\mathbf{1}(\cdot)$ is the indicator function. We observe that for unicast connections, the flow rates are below the EAS rates $10-25\%$ of the time, while the number is $10-15\%$ for multicast flows over the same duration. Therefore, for a significant percent of the time, the instantaneous flow rates drop below EAS rates. From a user point of view, this fluctuation in the session rate of an active connection may not be desirable for some applications, even thought the perceived rate is higher on the average.

In Fig. 7(a), we plot the percentage of total power assigned to each spot-beam queue, and in Fig. 7(b), the corresponding service rates, given by

$$\overline{p}_j^{\text{BAS-I}} = 100 \cdot \frac{1}{T} \sum_{t=1}^{T} p_j^{\text{BAS-I}}[t], \text{ and} \qquad (49)$$

$$\overline{r}_j^{\text{BAS-I}} = \frac{1}{T} \sum_{t=1}^{T} r_j^{\text{BAS-I}}[t], \qquad (50)$$

respectively, for $j = 1, 2, \ldots, M$. Observe that several spot-beam queues have average power levels that are below the minimum power levels that would maintain the EAS session rates. Consequently, all unicast flows forwarded to such queues have lower average session rates than their EAS rates giving rise to the behavior we observe in Fig. 6(a).

The second set of results look at the average performance of the BAS-I policy under changing group dynamics. In these experiments, there are $L_u = 250$ unicast flows, while the number of active multicast flows is varied between $L_m = 10$ to 50. In Fig. 8(a), we plot the change in the sustainable session

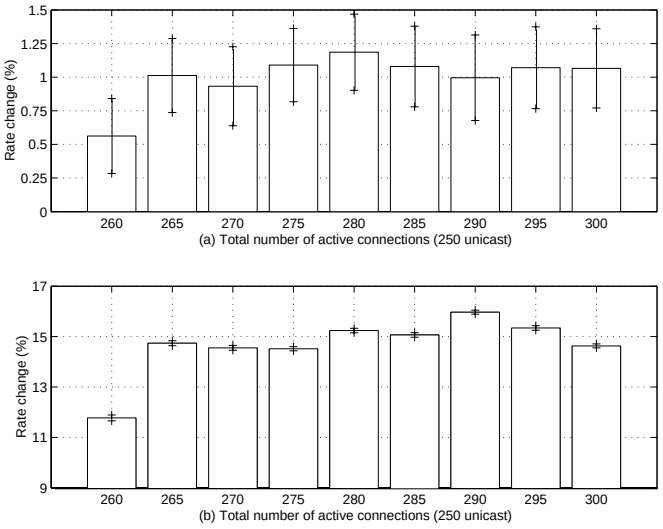

Fig. 8. (a) Average rate change in the sustainable session rates of all unicast flows average over all active unicast flows under the BAS-I policy. (b) Average rate change in the sustainable session rates of all multicast flows averaged over all active multicast flows under the BAS-I policy.

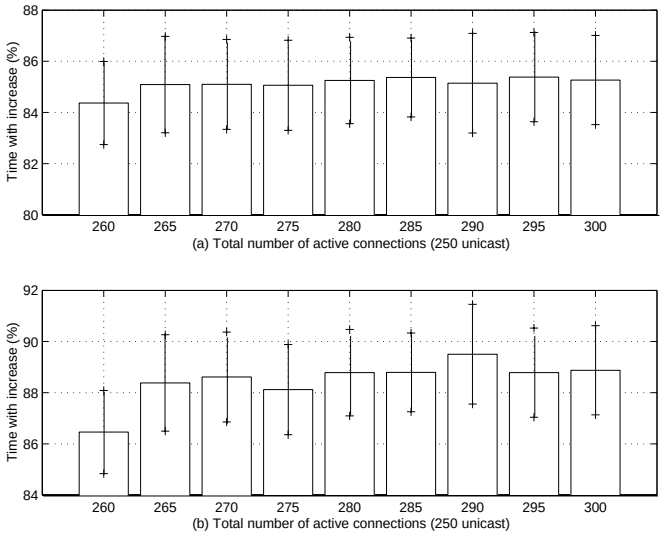

Fig. 9. (a) Percent of total test time BAS-I session rates of unicast flows are equal to or better than EAS session rates averaged over all active unicast flows. (b) Percent of total test time BAS-I session rates of multicast flows are equal to or better than EAS session rates averaged over all active multicast flows.

rates averaged over all active unicast flows, — i.e.

$$\overline{\eta}_u^{\text{BAS-I}}(L_m) = \frac{1}{L_u}\sum_{i=1}^{L_u}\overline{\eta}_i(L_m), \tag{51}$$

as the number of active multicast connections is varied. Fig. 8(b) plots the same metric for multicast flows:

$$\overline{\eta}_m^{\text{BAS-I}}(L_m) = \frac{1}{L_m}\sum_{i=L_u+1}^{(L_u+L_m)}\overline{\eta}_i(L_m). \tag{52}$$

Both figures also show the 95% confidence intervals on the mean values. We observe that unicast flows do not experience a significant increase in their average session rates while the number of active multicast connections is varied. The average value is low, because some unicast flows actually experience a decrease in their average session rates (compare to Fig. 6(a)). However, multicast flows experience $12-16\%$ increase in their average sustainable rates. Finally, in Fig. 9, we plot the percentage of time flows have higher session rates than their EAS rates as the number of active multicast connections is varied.

From the results of this section, we conclude that BAS-I policy increases the average session rates for both types of flows, however, (i) the policy is unfair against unicast flows, and some unicast flows may see a decrease in their rates, and (ii) although the session rates are higher compared to EAS session rates on the average, instantaneous values may drop below the EAS values. BAS-II policy elevates these issues by imposing bounds on the power levels to guarantee that session rates remain at or above EAS policy values at all times for both types of flows. In the next section, we look at the performance of the BAS-II policy under similar test settings.

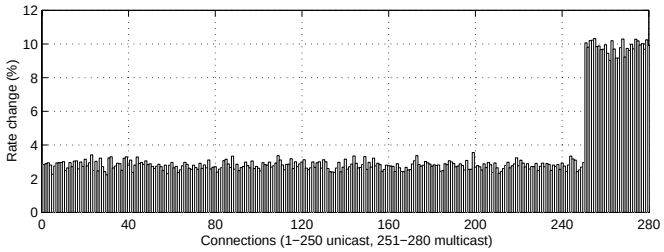

Fig. 10. Percent rate change in the sustainable rates of all active flows averaged over the test duration under BAS-II policy, compared to the session rates under EAS policy.

B. Results under BAS-II policy

In this section, we look at the same set of metrics under the BAS-II policy. In Fig. 10, we plot the percent change in the sustainable session rates of all active flows averaged over the test duration of $T = 100$ time units for the BAS-II policy case. Under BAS-II policy, multicast flows experience an average increase of up to 10% in their sustainable session rates, while unicast flows experience more moderate gains of up to 3.5%. We observe that, compared to BAS-I policy, the increase in the session rates of active multicast flows is down by approximately 5%, however, all unicast flows have average rates that are higher than their rates under the EAS policy, and the average increase is more uniform across all unicast flows. Moreover, instantaneous session rates remain above the EAS session rates at all times, — i.e. $\overline{\tau}_i^{\text{BAS-II}} = 100\%,\ \forall i$.

In Fig. 11(a), we plot the percentage of total power assigned to each spot-beam queue, and in Fig. 11(b), the corresponding service rates. Note that under this policy, all power levels are at or over the minimum power levels required to maintain EAS policy session rates for all active flows. We observe that, the

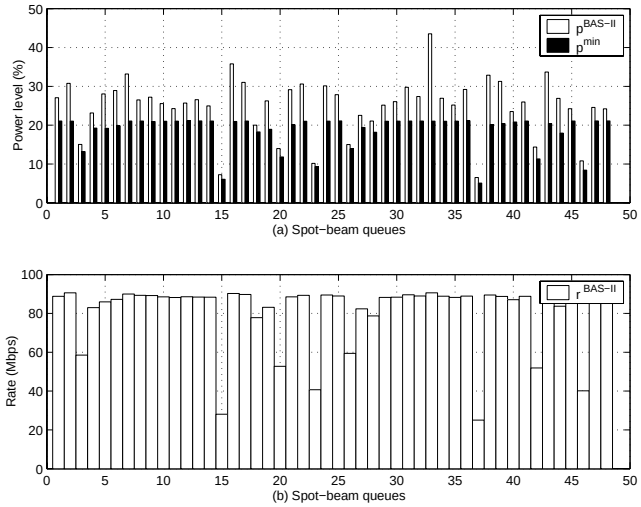

Fig. 11. (a) Average power assigned to each spot-beam queue as percentage of total system power over the test duration under BAS-II policy. (b) Average service rate of the spot beam queue over the test duration under BAS-II policy.

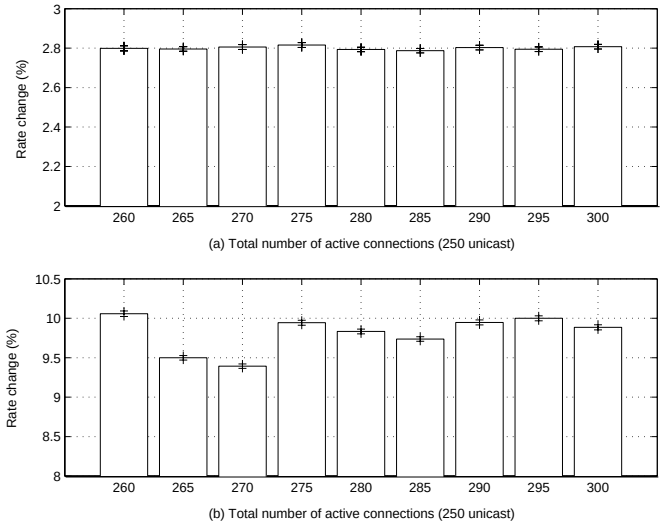

Fig. 12. (a) Average rate change in the sustainable session rates of all unicast flows average over all active unicast flows under the BAS-II policy. (b) Average rate change in the sustainable session rates of all multicast flows averaged over all active multicast flows under the BAS-II policy.

average service rate across spot-beam queues are now more uniform compared to the levels under the BAS-I policy.

The second set of results look at the average performance of the BAS-II policy under changing group dynamics. In these experiments, there are $L_u = 250$ unicast flows, while the number of active multicast flows are varied between $L_m = 10$ to 50. In Fig. 12(a), we plot the change in the sustainable session rates averaged over all active unicast flows as the number of active multicast connections is varied. Fig. 12(b) plots the same metric for multicast flows. The figures also show the 95% confidence intervals for the mean values. Under BAS-II policy, unicast flows favor much better than under the BAS-I policy and experience an average increase of up to 2.8%. Compared to BAS-I policy, multicast flows benefit less from the optimization, however, their average rates remain $9 - 10\%$ above their EAS session rates.

We can conclude that BAS-II policy still attains desirable performance improvements in terms of the average sustained session rates, while solving the fairness related issues of the BAS-I policy. In the next section, we sum up our observations and provide future directions on this work.

VII. Conclusion

In this paper we have introduced an optimization framework for balancing the spot-beam queue service rates such that the sum of the rate variances of all active multicast flows is minimized. This is achieved through the re-distribution of system power among spot-beam queues, taking into account the load on the queues and the channel states. The rate variance metric effectively captures the fact that multicast flows affect the load distribution of multiple spot-beam queues, and is used to achieve performance improvements from it. We provide two alternative policies, BAS-I and BAS-II, respectively. BAS-I policy does not impose any lower bounds on the minimum power level to be assigned to each spot-beam queue, and therefore may be unfair at a per-flow level. However, the policy increases the sustainable session rates of multicast flows by up to 16% when averaged over all active multicast flows. BAS-II policy imposes tight lower bounds on the power levels and, therefore, the multicast flows experience an increase of up to 10%. However, the policy also guarantees that the unicast flows do not see a performance degradation in their rates at the end of the optimization.

Depending on the application, the type of the flows, and the service rate guarantees provided by the service provider, it may not be necessary to require a strict minimum rate for all active flows. Therefore, an extension to the current policies is under study to provide a quality of service (QoS) or priority based minimum rate requirement to all active flows. In this alternative policy, a QoS level is attached to all flows to determine which flows are allowed a reduced rate, and what are the minimum rate requirements. This information is used to determine the minimum power levels for each spot-beam queue. On the overall, we conclude that it is possible to improve the performance of the system by careful tuning of system parameters to the requirements of the flows supported by it, and provide an example to the fact that future systems must be designed to support multiple types of flows.

References

[1] E. Lutz, M. Werner, and A. Jahn, *Satellite systems for personal and broadband communications*. Springer-Verlag, 2000.

[2] Hughes Communications Galaxy Inc., "Application of Hughes Communication Galaxy, Inc. before the Federal Communications Commission for two Ka-band domestic fixed communication satellites," December 1993.

[3] E. J. Fitzpatrick, "SPACEWAY system summary," *Space Communications*, vol. 13, no. 1, pp. 7–23, 1995.

[4] U.-C. Fiebig, "A time-series generator modeling rain fading and its seasonal and diurnal variations," in *1st International Workshop of COST-Action 280*, Malvern, UK, 2002.

[5] Population Division U.S. Census Bureau, "Table NST-EST2003-01, Annual estimates of the population for the United States, and for Puerto Rico: April 1, 2000 to July 1, 2003," release date: December 18, 2003. [Online]. Available: http://eire.census.gov/popest/data/states/tables/NST-EST2003-01.php

[6] A. Legout, J. Nonnenmacher, and E. W. Biersack, "Bandwidth allocation policies for unicast and multicast flows," *IEEE/ACM Trans. Networking*, vol. 9, no. 4, pp. 464–478, August 2001.