

TECHNICAL RESEARCH REPORT

Load balancing in multi-beam satellite systems

by Gun Akkor, John S. Baras, Michael Hadjitheodosiou

CSHCN TR 2004-9
(ISR TR 2004-18)



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/>

Load Balancing in Multi-beam Satellite Systems

Gun Akkor, John S. Baras, and Michael Hadjitheodosiou
University of Maryland, College Park, MD 20742, USA.

We propose an approach to optimize resource sharing and in a multi-beam broadband satellite system that supports both unicast and multicast flows. We show that in this architecture, the load on every spot-beam queue could be different, depending on the type of the flows and the distribution of the receivers across spot-beam coverage areas. This load unbalance may significantly under-utilize the system resources and decrease the system throughput when both unicast and multicast flows are active in the system. In this paper, we formulate an optimization problem for intra- and inter-beam resource sharing such that the variance of the session rates experienced by users of a flow located in different beam coverage areas is minimized. The result of our resource allocation also determines the maximum sustainable rate of each flow. We calculate the beam utilization and maximum sustainable sessions rates with and without optimization and compare the results. We conclude that this method significantly improves the session rates and overall utilization of the system when both unicast and multicast flows are active.

I. Introduction

BROADBAND satellite systems are quickly becoming an integral component of communication systems. Next generation satellite communication systems utilizing higher frequency bands, such as the Ka-band, and supporting spot-beam technology and on-board packet processing are under development. These systems offer higher data rates and enable the use of small, low-power and low-cost user terminals. They are likely to play a greater and more important role in the broadband communication infrastructure.

In the recent past, a new set of applications, such as distributed computing, distributed software updates, and distance learning, have emerged in the Internet. These applications are distributed in nature and require concurrent transmission of the same content to multiple users. Satellite communication systems offer wide-area coverage and ubiquitous access to potentially large number of users. Therefore, they have an inherent advantage in supporting multicast services. However, the question of how multicast support could be efficiently integrated into the design of new satellite systems remains to be a challenge.

In this paper, we look at the problem of resource sharing and flow control in a multi-beam satellite system that supports both unicast and multicast flows. We show that a high load variation across the spot-beams queues may significantly under-utilize the system resources and decrease the system throughput when a mixture of unicast and multicast flows is active in the system. We propose an optimization based-approach to balance the load in the system, and in doing so, take into account the presence of both multicast and unicast flows.

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 problems. In Section III, we formulate our problem in precise terms. Section IV discusses the solutions and Section V evaluates the effectiveness of the propose approach. Last section concludes the paper.

II. Problem Outline

In this paper, we consider a satellite communication system, 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 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 spot-beams. A user terminal may serve a small home network with only a few user machines or may act as a gateway for a local area network. In general, we assume that there are as many terminals as there are users, and from our point of view, user terminals are the end-points. The structure of the network is shown in Figure 1.

In this multiple spot-beam system, packets of several active flows are queued at the NOC buffer. The NOC forwards the packets to the satellite at a rate upper-bounded by the uplink capacity of the system. An on-board processor and a 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 user resides. In a multicast flow, however, receivers of the multicast session may reside in multiple beam coverage areas, and therefore packets need to be duplicated and forwarded to multiple spot-beam queues. Depending on the type of the flows and the distribution of the users across spot-beam coverage areas, the load on each spot-beam queue could be different. For a unicast flow, the maximum sustainable rate at the NOC is equal to the rate at which the flow could be served at the spot-beam queue it belongs to. However, for a multicast flow, the maximum sustainable rate at the NOC is equal to the minimum of the rates it could be served at the spot-beam queues it has been forwarded to. Consequently, if the multicast flow experiences a high rate variation across the queues, this minimum requirement would significantly under-utilize the queues.

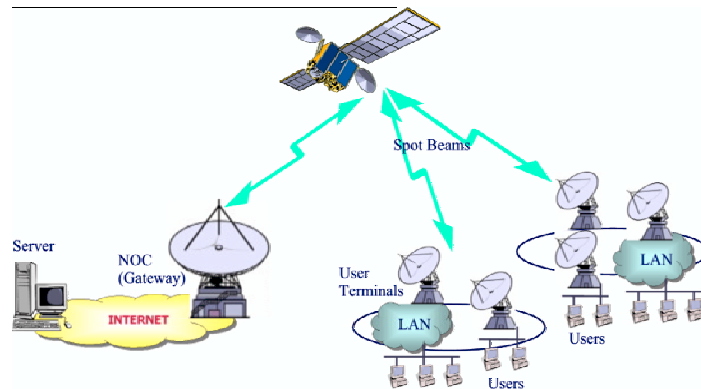


Figure 1: Satellite Network Architecture

In a typical system, the number of spot-beams queues are much larger than the number of on-board antennas for two reasons: (i) with a large number of antennas satellite becomes too heavy, and (ii) because of frequency re-use across beams, not all beams could be served simultaneously. Therefore, spot-beam queues normally share the access to the antennas, transmitting packets in the form of bursts when the connection is realized. In this architecture, if all spot-beam queues had equal burst durations then the amount of data that could be transmitted at a fixed downlink rate would be the same for all queues. However, it may be possible to optimize the burst duration of each beam based on the load of the queue and the type of the flows served at the queue such that the resulting system minimizes the rate variances experienced by multicast flows. Therefore, following problem has to be addressed:

- We need to find the optimal way to share the resources of the system among spot-beams queues (inter-beam) and among flows belonging to the same queue (intra-beam) such that the queues are utilized efficiently and the flows are served at the maximum sustainable rates at the NOC.

In the next section, we formulate this problem in detail and define the optimization problem. We discuss why existence of active multicast flows cause under-utilization of the system and how the solution of this optimization setting helps improve the system performance.

III. Problem Formulation

In this system, M spot-beam queues share the access to K on-board antennas, each transmitting in bursts only when its output is switched to an antenna. We assume that a fixed set of beams share the access to a given antenna at any time and denote the collection of indices representing the spot-beam queues that access antenna k for $k=1,2,\dots,K$ by set A_k . The time-share c_j denotes the fraction of time spot-beam b_j , for $j \in A_k$, accesses the antenna k , such that

$$0 \leq c_j \leq 1, \text{ for } j=1,2,\dots,M \text{ and} \\ \sum_{j \in A_k} c_j = 1. \quad (1)$$

A flow f_i , for $i=1,2,\dots,L$, which is forwarded to beam b_j is assigned a time-share w_{ij} of the fraction of the burst duration of the beam depending on the load of the beam and the type of the flows forwarded to it, such that

$$w_{ij} = 0, \text{ if } i \notin B_j \\ 0 < w_{ij} \leq 1, \text{ if } i \in B_j \text{ for } \forall i, j, \\ \sum_{i \in B_j} w_{ij} = 1 \quad (2)$$

where, B_j is the set of indices representing the flows that are forwarded to beam b_j . Therefore, the packets of flow f_i could be served at a *maximum rate* of

$$\lambda_{ij} = w_{ij} \cdot c_j \cdot R \quad (3)$$

at beam b_j , where R is the downlink rate of the antenna. However, the *maximum sustainable rate* of the flow at the NOC queue is limited to

$$\lambda_i = \min_{j: i \in B_j} \{ \lambda_{ij} \} \quad (4)$$

in order to avoid overflowing of any of the on-board beam queues. For unicast flows, there exists a single beam index j for which $i \in B_j$ corresponding to the beam where the destination user resides. Therefore, if all active flows in the system were unicast, the system would remain fully utilized, i.e.

$$\sum_{j=1}^M \sum_{i \in B_j} w_{ij} \cdot c_j = K. \quad (5)$$

When there are multicast flows, on the other hand, those queues for which $\lambda_{ij} > \lambda_i$ for some multicast flow f_i will remain under-utilized. The time-share, w_{ij} , of each flow is determined by the load on the queue and the type of the flows forwarded to the queue, and a high variation across beams would significantly under-utilize the system. Therefore, our goal is to minimize this variation across the beams by adjusting the time-share each beam gets of the antenna. In other words, we would like to find the optimal vector $\mathbf{c}^* = [c_1^*, \dots, c_M^*]$ that would minimize the normalized ($R=1$) rate variance a flow experiences across beams over all flows:

$$\mathbf{c}^* = \arg \min_{\mathbf{c}} \sum_{i=1}^L \sigma_i^2, \quad (6)$$

such that

$$l_j \leq c_j \leq 1 \quad \forall j \\ \sum_{j \in A_k} c_j = 1 \quad \forall k, \quad (7)$$

where $0 \leq l_j \leq 1$ is a lower-bound on the value of c_j , and

$$\sigma_i^2 \stackrel{\Delta}{=} \frac{1}{N_i} \sum_{j=1}^M s_{ij} \cdot (w_{ij} \cdot c_j - \mu_i)^2, \quad (8)$$

$$\mu_i \stackrel{\Delta}{=} \frac{1}{N_i} \sum_{j=1}^M w_{ij} \cdot c_j, \quad (9)$$

$$s_{ij} = \begin{cases} 1, & \text{if } i \in B_j \\ 0, & \text{if } i \notin B_j \end{cases}, \text{ and} \quad (10)$$

$$N_i \stackrel{\Delta}{=} \sum_{j=1}^M s_{ij}. \quad (11)$$

Equation (8) gives the sample variance of the time-share of flow f_i across the N_i beams it is forwarded to and Eq. (9) is the sample mean of the time-shares of the flow. Note that for unicast flows, $N_i = 1$ and $\sigma_i^2 = 0$. Therefore, unicast flows do not affect the solution of our optimization algorithm, but they change the total load on the system and the time-share each flow gets from the system. In the following section, we give the solution to Eq. (8) and define the antenna sharing policy we will use as basis for comparison.

IV. Solution

When no optimization is considered, the simplest assignment is to set

$$c_j = \frac{1}{|A_k|}, \forall j \in A_k, k = 1, 2, \dots, K. \quad (12)$$

We call this assignment, *equal-antenna-share* (EAS) policy and denote it by the vector \mathbf{c}_{EAS} . The solution to Eq. (6) is called the *balanced-antenna-share* (BAS) policy and denoted by the solution vector \mathbf{c}_{BAS} . We consider two different solutions to Eq. (6). In the first case, which we will refer to as BAS-I, the lower bounds on the solution values are set to zero for all beams — i.e. $l_j = 0, \forall j$. Therefore, in the solution vector some beams may get lower time-shares than their assignments under the EAS policy, which results in lower sustainable rates for some flows. In the second case, which we will refer to as BAS-II, lower bounds on the solution values are set such that no flow gets a lower sustainable rate than that of under the EAS policy, i.e.

$$l_j = \max_{i \in B_j} \left\{ \frac{\gamma_i}{w_{ij}} \right\}, \quad (13)$$

where γ_i is the maximum sustainable rate for flow f_i under EAS policy given by

$$\gamma_i = \min_{j: i \in B_j} \left\{ w_{ij} \cdot \frac{1}{|A_k|} \right\}, \text{ for } i \in B_j \text{ and } j \in A_k. \quad (14)$$

In the solution, we classify the beams into three sets: (i) E , the set of empty beams for which $B_j = \phi$, (ii) U , the set of beams with only unicast flows, and (iii) U^c , the set of beams with both unicast and multicast flows. Under

both BAS-I and BAS-II policies, $c_j = 0, \forall j \in E$, and such beams can be removed from the calculation. Beams with only unicast flows have to be excluded from the calculations as well, because, independent of the time-share they get, unicast flows that are forwarded to the those beams will have zero rate variance. Therefore, we keep the EAS policy time-shares for such beams, and set $c_j = 1/|A_k|$ for $j \in (U \cap A_k)$ and $k = 1, 2, \dots, K$. The assignments of the remaining beams are policy dependent and described in the following sections.

A. Solution under BAS-I policy

Under this policy, the optimum policy assignments for the beam time-shares are given by

$$c_j = \frac{1 - \sum_{l \in (U \cap A_k)} c_l}{a_j \cdot \sum_{l \in (U^c \cap A_k)} \frac{1}{a_l}}, \text{ for } j \in (U^c \cap A_k) \text{ and } k = 1, 2, \dots, K, \quad (15)$$

where, a_j are coefficients given by

$$a_j = \sum_{i=1}^L \left[\frac{2 \cdot (N_i - 1)}{N_i^2} \cdot w_{ij}^2 \right]. \quad (16)$$

B. Solution under BAS-II policy

Under this policy, the optimum assignments for the beam time-shares are more complicated due to the presence of lower bounds. In this case, the method given in Appendix is used to calculate the assignments.

V. Evaluation

A. Analysis setup

In order to evaluate the effectiveness of our approach, we need to create unicast and multicast flows between the

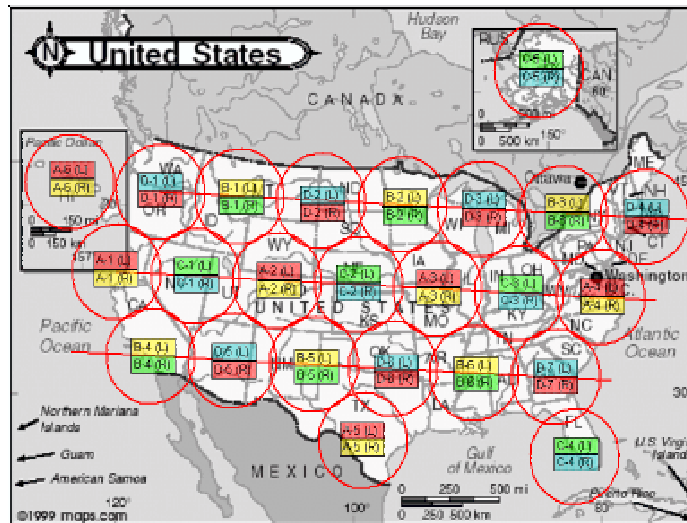


Figure 2: Locations of the 48 spot-beams in two polarizations over the United States for the satellite system.

NOC and the spot-beam locations. However, the number of the unicast and multicast flows forwarded to each spot-beam location and the distribution of multicast users across these locations should reflect the possible load unbalance in a real multi-beam satellite system. Therefore, first, we have looked at the beam locations and the antenna assignments of a geo-synchronous satellite proposed for a commercial satellite system^{1, 2}. Figure 2 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 1. Next, based on the approximate geographical area covered by each spot-beam, we have calculated the total population illuminated by each spot-beam using the most recent U.S. Census Data³. Assuming that a flow f_i is more likely to be forwarded to a spot-beam queue b_j if the beam illuminates a larger fraction of the total population, we calculated the probability distribution plotted in Fig. 3. This distribution gives the probability of f_i being forwarded to a beam b_j for all 48 spot-beams and is used to create flows between the NOC and the spot-beam locations. Finally, we have to determine how the burst duration of each beam 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 the same bottleneck, in this particular case, the same spot-beam queue. In Ref. 4 authors propose a

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 1: Assignment of spot-beams to on board antennas for the SPACEWAY system.

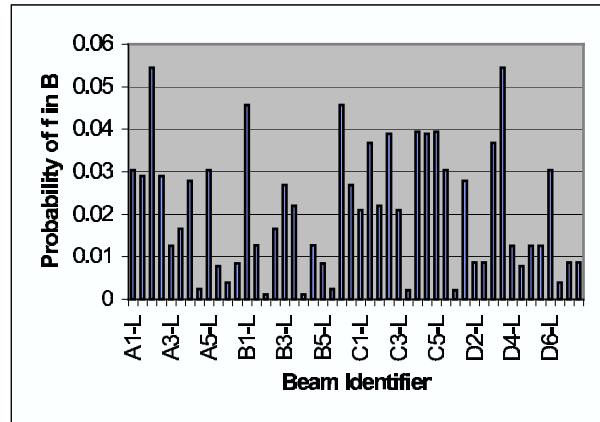


Figure 3: Probability distribution function of a flow being forwarded to a spot-beam location designated by the beam identifier.

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 time-share w_{ij} of a flow f_i in beam 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:

$$w_{ij} = \begin{cases} 0, & \text{if } n_{ij} = 0 \\ \frac{1 + \log(n_{ij})}{\sum_{i \in B_j} 1 + \log(n_{ij})}, & \text{if } n_{ij} \neq 0 \quad \forall i, j \end{cases} \quad (17)$$

In the next section, we calculate the optimal time-share of each spot-beam, and the maximum sustainable rates of every flow under both BAS-I and BAS-II. We compare our results to the EAS policy.

B. Analysis results

For comparison, we create 250 unicast flows between the NOC and the spot-beam locations and then look at the performance gains when the system resources are shared with 10 to 40 multicast flows. For each test instance, we create 500 different flow configurations using the test setup described above and look at the average gains over the

test set. In Fig. 4, we look at the maximum sustainable rate increase (at the NOC) or decrease experienced by a single flow and the average sustainable rate improvement over all flows as the number of active multicast flows is increased from 10 to 40. We observe that under BAS-I policy, an average of 40%-50% rate improvement is achievable over the test sets. As more multicast flows are added to the system relative improvement becomes more

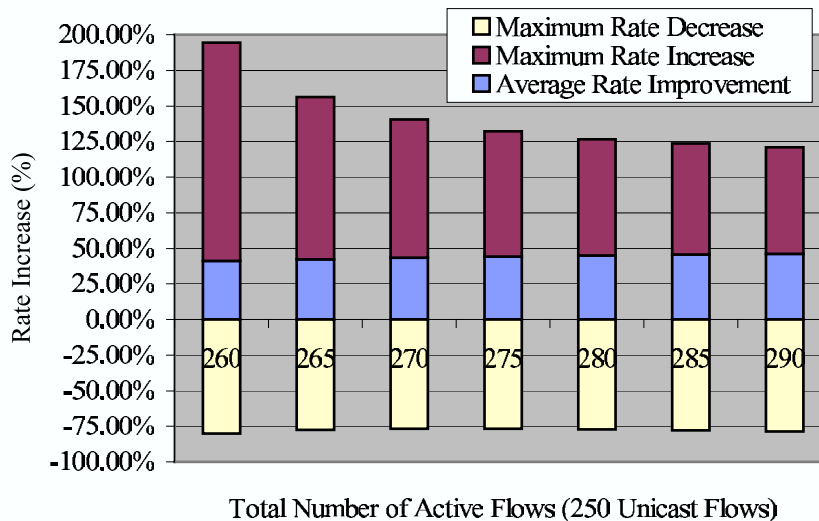


Figure 4: Maximum rate increase, decrease, and the average rate improvement experienced by the sustainable rates of active flows at the NOC for 250 unicast flows and 10 to 40 multicast flows under BAS-I policy with respect to the sustainable flow rates under EAS policy.

significant, since our approach does not yield any improvement over the EAS policy when there are only unicast flows. However, the maximum sustainable rate improvement experienced by a single flow decreases as more flows compete for the available resources. Although the system experiences considerable gains in terms of the average sustainable rate of a flow, the fairness is an issue with BAS-I policy, since some flows end up experiencing a rate decrease compared to EAS policy. In Fig. 5, we look at what percentage of the total number of flows experience a decrease in their sustainable rates under BAS-I policy. We observe that approximately 80% of the total flows experience an increase in their sustainable rates. Therefore, from a system performance point of view, the BAS-I policy improves utilization significantly.

In Fig. 6, we look at similar performance gains under BAS-II policy. BAS-II policy imposes lower bounds on the solution such that no flow experiences a decrease in their sustainable rates with respect to the EAS policy. This constrains the solution space. Therefore, we observe more moderate gains of 10%-20% on the average sustainable flow rates at the NOC. However, in this case there is no fairness problem, since all flows are served at equal or better rates compared to that of EAS policy.

In Fig. 7, we look at the improvement in the utilization of the NOC queue defined by

$$\eta = \frac{\left(\sum_{i=1}^L \lambda_i^{\text{BAS}} - \sum_{i=1}^L \lambda_i^{\text{EAS}} \right)}{\left(\sum_{i=1}^L \lambda_i^{\text{EAS}} \right)}. \quad (18)$$

We observe that, BAS-I outperforms the BAS-II policy for all test cases due to the more stringent constraints imposed by the latter. However, the difference between the respective performance gains becomes smaller as the number of multicast flows increases, since this time the algorithm becomes constrained due to the existence of too many competing flows.

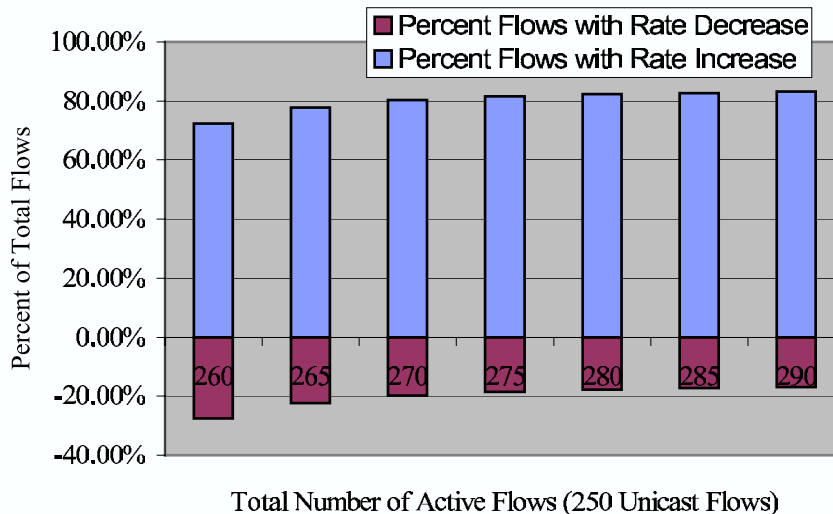


Figure 5: Percentage of flows experiencing a rate increase or a decrease in their sustainable rates at the NOC for 250 unicast flows and 10 to 40 multicast flows under BAS-I policy with respect to the sustainable flow rates under EAS policy.

Finally, in Figs. 8 and 9, we look at the utilization of individual spot-beam queues under both policies for a test case of 250 unicast flows and 20 multicast flows. We observe that a significant percent of all spot-beam queues is better utilized under BAS policies.

VI. Conclusions

In this paper, we have proposed an approach to optimize the allocation of system resources across multiple spot-beams that support both unicast and multicast flows. The proposed method attempts to balance the load variation among spot-beams by adjusting for the burst duration allocated to each spot-beam queue. The benefits are two fold. The optimization minimizes the rate variation a multicast flow experiences due to the load unbalance when its receiver set spans more than one spot-beam location, and allows a higher sustainable rate. Moreover, utilization of both on-board queues and the NOC queue are improved, resulting in a higher average throughput.

However, we only tested the performance gains under static configurations. In our tests, we showed the possibility of performance improvements over an average of 500-hundred test cases. We are currently investigating how this method can be combined with an on-board algorithm to dynamically check for the load on each spot-beam queue and the type of the flows to redistribute the resources.

A more extensive study is currently under way to see how the rate restrictions imposed by this method would interact with the flow control mechanisms of individual flows.

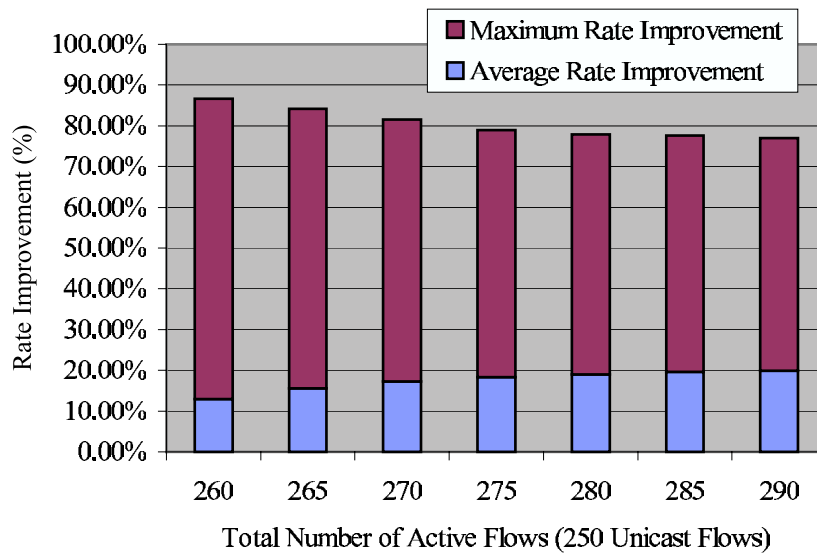


Figure 6: Maximum and average percent improvement experienced by the sustainable rates of active flows at the NOC for 250 unicast flows and 10 to 40 multicast flows under BAS-II policy with respect to the sustainable flow rates under EAS policy.

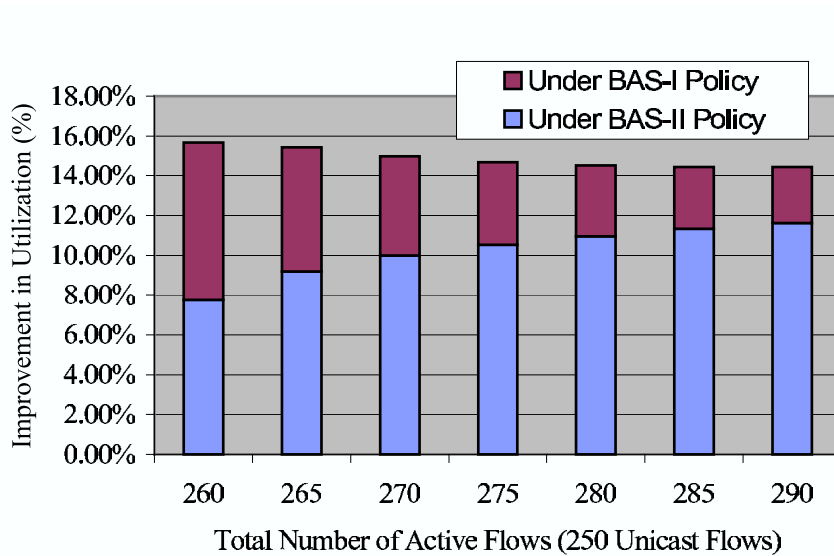


Figure 7: Average improvement in the utilization of the NOC queue for 250 unicast flows and 10 to 40 multicast flows under both policies with respect to the utilization under EAS policy.

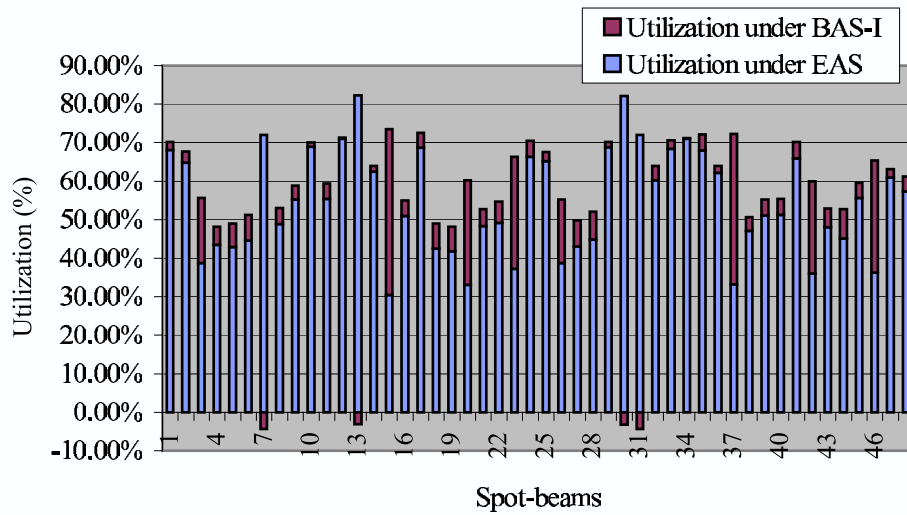


Figure 8: Utilization of individual spot-beam queues for 250 unicast flows and 20 multicast flows under BAS-I policy with respect to EAS policy.

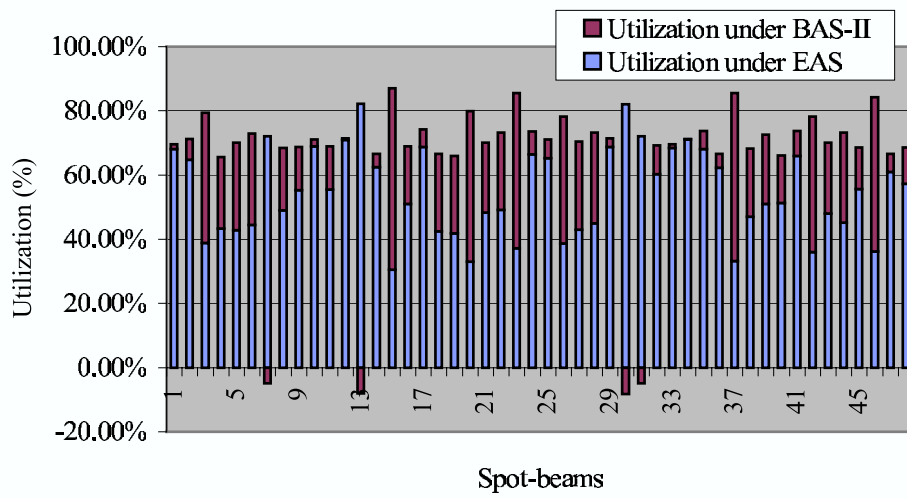


Figure 9: Utilization of individual spot-beam queues for 250 unicast flows and 20 multicast flows under BAS-II policy with respect to EAS policy.

Appendix

In this section, we give the method for calculating the $c_j, j \in (U^c \cap A_k)$ for $k=1,2,\dots,K$ under BAS-II policy:

CALCULATE a_j using Eq. 16 for $\forall j \in (U^c \cap A_k) \quad k=1,2,\dots,K$.

SORT a_j in descending order for $\forall j \in (U^c \cap A_k) \quad k=1,2,\dots,K$.

FOR $k=1,2,\dots,K$,

SET $d = 1 - \sum_{l \in (U \cap A_k)} c_l$.

SET $S = \sum_{l \in (U^c \cap A_k)} \frac{1}{a_l}$.

FOR $\forall j \in (U^c \cap A_k)$,

IF $\frac{d}{a_j \cdot \sum_{l \in (U^c \cap A_k)} \frac{1}{a_l}} \geq l_j$ THEN $c_j = \frac{d}{a_j \cdot \sum_{l \in (U^c \cap A_k)} \frac{1}{a_l}}$

ELSE

$c_j = l_j$.

$d = d - l_j$.

$S = S - \frac{1}{a_j}$.

END

END

END

References

¹Fitzpatrick, E. J., "SPACEWAY System Summary," Space Communications, Vol. 13, No. 1, 1995, pp. 7-23.

²Hughes Communications Galaxy, Inc., "Application of Hughes Communication Galaxy, Inc. Before the Federal Communications Commission for Two Ka-Band Domestic Fixed Communication Satellites," December 1993.

³Population Division, U.S. Census Bureau, "Table NST-EST2003-01 - Annual Estimates of the Population for the United States and States, and for Puerto Rico: April 1, 2000 to July 1, 2003," [online database], URL: <http://eire.census.gov/popest/data/states/tables/NST-EST2003-01.php> [Release Date: December 18, 2003].

⁴Legout, A., Nonnenmacher, J., and Biersack, E. W., "Bandwidth-Allocation Policies for Unicast and Multicast Flows," *IEEE/ACM Transactions on Networking*, Vol. 9, No. 4, 2001.