Resource management for advanced transmission antenna satellites

Citation: Choi, J.P., and V.W.S. Chan. "Resource management for advanced transmission antenna satellites." Wireless Communications, IEEE Transactions on 8.3 (2009): 1308-1321. © 2009 Institute of Electrical and Electronics Engineers

As Published: http://dx.doi.org/10.1109/TWC.2009.071131

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/52353

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Resource Management for Advanced Transmission Antenna Satellites

Jihwan P. Choi, *Member, IEEE*, and Vincent W. S. Chan, *Fellow, IEEE*

***Abstract*—In satellite communications, narrow spotbeams can provide high power and data rates to the desired location while reducing spatial interference. Advanced transmission antenna technology is critical to generate and switch narrow beams rapidly among a large number of users under quality of service (QoS) constraints such as average delay. In this paper, we jointly optimize resource allocation and congestion control, and compare the performances of two types of satellite transmit antennas: a multiple beam antenna and a phased array antenna. For a multiple beam antenna with traveling wave tube amplifiers (TWTA), throughput is decided by either the most demanding user or the average of all user parameters. For a phased array antenna, joint antenna gain patterning and beam scheduling is given as a function of channel conditions, interference (depending on users' geographical distribution), and average delay requirements. We then develop a low-complexity on-line algorithm of choosing either interference suppression or sequential service for the active users who are closely located within the width of a spotbeam. Due to flexible power allocation, the phased array antenna can provide better performance than the multiple beam antenna when a small number of users are very demanding or many users are densely crowded in a small area.**

***Index Terms*—Satellite communication, resource allocation, congestion control, multiple beam antenna, phased array antenna.**

I. Introduction

A Future satellite will use many narrow spotbeams in its coverage area. Narrow spotbeams can project high power density and thus can support high data rates to small user terminals. In addition, the same frequencies can be reused in different cells, increasing the total system capacity. As higher frequency bands (20 GHz and beyond) are used to provide higher rates for data networking applications, it becomes more attractive architecturally to implement rapidly reconfigurable, agile and narrow beams. However, with narrow spotbeams, a large number of beams and transponders would be required. For example, consider a geosynchronous-Earth-orbit (GEO) satellite at an altitude of $L = 35,800$ km. Assume that the satellite has a transmit antenna with a diameter of $D = 10$ m at 20 GHz with carrier wavelength $\lambda = 0.015$ m. The diffraction-limited narrowest spotbeam size is then $\frac{\lambda L}{D} = 53.7$ km [1], [2]. To cover the whole United States ($\sim 5,000 \times 3,000$ km^2), the satellite would need to generate about 5,200 beams.[1] It will be impossible to carry corresponding transponders and on-board equipments. On-board resource is expensive and consumes considerable weight and power. An optimized method of agile antenna gain patterning and beam scheduling is required to greatly improve the efficiency of transmission and resource management.

In our previous work [3], we suppressed the issue of controlling excess traffic and its delay, and focused only on the long-term average gain in terms of Shannon capacity and power efficiency. The most challenging design task for maximizing the network efficiency is that the resource allocation and scheduling problem should be considered from the viewpoint of joint optimization over multiple network layers. It is still an open problem to solve all these issues together because of the complexity of the problem. Thus, one may only explore some of these crosslayer problems separately and develop a feel for how an efficient system must be designed. There have been prior works on the architecture of each separate network layer. Some examples for multibeam satellite systems include the design of antennas [4]–[6] and multiple access protocols [7]. Congestion control problems over satellite networks have been addressed by modifying the existing Transport Layer protocols or suggesting new protocols, to overcome the disadvantages of long propagation delay or bursty errors over satellite channels [8]–[10]. Congestion control prevents excessive packet loss and stabilizes the system with an acceptable queuing delay. In this paper, we use a form of rate congestion control by throttling incoming traffic, based on the average delay and available resource. We present a simplified formulation of the congestion control and satellite resource allocation problem. Then, we explore jointly optimized solutions, in order to focus on some aspects and obtain analytical results.

We compare two types of advanced transmission antennas for multiuser satellites. First, we examine the use of multiple beam antenna (Fig. 1(a)) with traveling wave tube amplifiers (TWTA). Traditionally, TWTAs have been widely used in satellite communications where a high power margin is required. Each multiple beam antenna feed is fed by its own TWTA, which results in a power constraint for each beam. Next, we consider the use of phased array antenna (Fig. 1(b)). A phased array antenna uses solid state power amplifiers (SSPA) and can linearly superimpose signals at array elements by controlling an antenna-patterning matrix. It

Manuscript received October 13, 2007; revised April 30, 2008, October 29, 2008, and December 11, 2008; accepted December 14, 2008. The associate editor coordinating the review of this paper and approving it for publication was A. Conti.

J. P. Choi is with Marvell Semiconductor, Inc., 5488 Marvell Lane M/S 2-401, Santa Clara, CA 95054 (e-mail: zhifuan@alum.mit.edu).

V. W. S. Chan is with the Dept. of EECS, MIT, 77 Mass Ave, Cambridge, MA 02139 (e-mail: chan@mit.edu).

This work was supported by DARPA NGI Grant. This paper was presented in part at the IEEE ICC 2004, Paris, France, at the IEEE MILCOM 2005, Atlantic City, NJ, and at the IEEE MILCOM 2006, Washington, DC.

Digital Object Identifier 10.1109/TWC.2009.071131

[1]Throughout the paper, we assume the use of transmission antennas generating square-shaped beams for simplicity. If antennas can shape circular beams, a factor of $\pi/4$ should be considered for calculation.

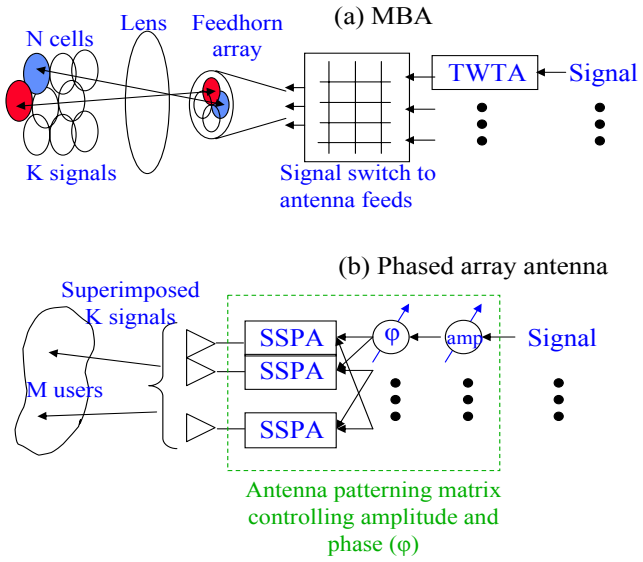


Fig. 1. Schematics of multiple beam antenna and phased array antenna

is reported that gallium nitride (GaN) [11] can achieve higher efficiency than other widely used materials such as silicon (Si), gallium arsenide (GaAs), and indium phosphide (InP). GaN is expected to be utilized for high-frequency high-power SSPAs with power efficiency up to 40 $\sim$ 60% (compared to 20% efficiency of TWTAs) [12]. SSPAs are frequently used to feed a large number of array antenna elements. Signal power can be divided among multiple channels up to the total power of the array. While the multiple beam antenna has a fixed beam size due to the fixed size of feedhorn for each signal, the phased array antenna can have any size and/or shape of beam by feeding many array elements with the same signal. Moreover, the phased array antenna together with transmission scheduling can cycle much more rapidly ($\ll$ msec) than the multiple beam antenna and is advantageous in meeting time deadlines via fast switching of resources. Flexible antenna gain patterning allows for simultaneous service of the users in a populated area by suppressing possibly significant interbeam interference. This paper will account for the impact of interference on the resource allocation problem. Each user with a satellite terminal is assumed to be fixed (i.e., no mobility) and operating at high frequency (e.g., Ka bands, 20 $\sim$ 30 GHz). The main factor for channel variation is weather-induced impairments due to rain [13], [14].[2]

This paper is organized as follows: In Section 2, for a multiple beam antenna, we formulate and solve the throughput maximization problem by controlling incoming traffic rates with average delay and transmitter-sharing constraints. In Section 3, we consider a resource allocation problem for a phased array antenna system, and derive an optimal solution of joint antenna gain patterning and beam scheduling. In Section 4, the throughput of the phased array antenna satellite is compared with that of the multiple beam antenna satellite. We give asymptotic closed-form solutions and numerical examples. In Section 5, we move our focus from the steady-state analysis to the real-time algorithm development for phased array antenna scheduling. Simulation results are compared with the steady-state solution in Section 6. Section 7 concludes the paper.

[2]For mobile terminals, multipath fading and shadowing effects should be considered as well. These processes are much faster than the weather-induced attenuation (10 ms vs. minutes/hours for the average duration of attenuation) [15]–[17].

II. Multiple Beam Antenna

A. Formulation

We want to allocate efficiently a limited amount of on-board transmission power and a small number of K active beams of a multiple beam antenna among many small N ($> K$) cells within a satellite coverage area. We also want to maximize the system throughput with reasonable queuing delays. For this purpose, we introduce a congestion control back-off parameter θ ($0 \leq \theta \leq 1$), which adjusts the incoming traffic rate, based on channel conditions and average delay constraints. We consider a multibeam satellite in a steady state. The incoming traffic of the average rate A_i (without congestion control) is presented to the i^{th} cell (Fig. 2). With the idealized assumption of infinite buffer size (i.e., no packet loss due to a full queue, for mathematical tractability), any type of general traffic pattern is acceptable. If A_i is too large for the system capacity, congestion control is triggered and incoming traffic should be backed off (i.e., slowed down) to θA_i. By maximizing θ while stabilizing the system, we can maximize the throughput. A single back-off parameter is used to achieve fairness for all users. Proportionally fair scheduling [18] maximizes $\sum_i w_i \log r_i$ subject to $\sum_i r_i = r_{total}$, where weight w_i determines how to allocate resource r_i within the total amount r_{total}. In our modeling, A_i plays the role of the weight for w_i, and the resource in consideration for r_i is the on-board power P_i allocated to the i^{th} cell. Solving the Lagrangian function $J(P_i) = \sum A_i \log P_i - \mu(\sum P_i - P_{total})$, where μ is a Lagrangian multiplier and P_{total} is the total on-board power, with respect to P_i gives the solution of $A_i/P_i = \mu$ for every i. Thus, backing-off every traffic rate by single θ still guarantees proportional fairness in terms of P_i proportional to A_i. We remark that the Transport Control Protocol (TCP) provides more bandwidth to small rate flows in general [19]. We may model the back-off parameter as $\theta_i(A_i)$, a decreasing function of the traffic rate, and use different θ_i for users. In an alternative way, A_i may be weighted by the price that users pay, so that the ill effect of greedy users can be minimized. However, these are out of scope of this paper. Instead, we will add per-user average delay constraints to the problem. The amount of allocated resource will not be simply proportional to the incoming traffic rate, but differentiated according to each user's quality of service (QoS) requirement.

The congestion control scheme that we propose here is very different from that of the TCP. According to the conventional classification of congestion control protocols [20], [21], our scheme is rate-control based and network-assisted while the TCP congestion control is window-control based and end-to-end implemented. Our rate-control scheme is implemented at the satellite on-board instead of end nodes. Upon triggering congestion control, the satellite throttles incoming traffic rates by explicitly feeding back the value of θ. The sources then slow down their traffic rates by multiplying θ. The packets

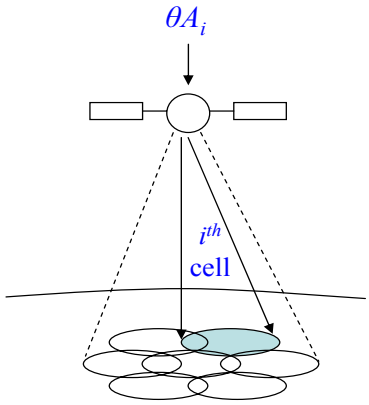


Fig. 2. A multiple beam antenna satellite with congestion-controlled incoming traffic

at the source should be backed off to meet the reduced incoming rate, but the satellite can deliver the packets that are admitted on-board within the average delay deadline. The use of single θ can simplify the feedback of the congestion control parameter; the satellite broadcasts only a single value of θ to all the sources. The decision of θ at the satellite reduces the feedback delay for congestion control, compared to end-to-end implementation. This can give significant advantage especially for long-latency satellite links. The parameter θ is bound to admission control function of packets entering the satellite on-board queues. One may extend the use of θ to the user-level admission control: e.g., if θ is less than some threshold, no more user is accepted into the system. This paper, however, does not consider this user-level admission control and assumes that the number of users in the system is already given.

Channel conditions over high frequency bands for fixed-location terminals are assumed to be quasi-static with constant signal attenuation over the period of interest. This assumption is reasonable because the packet processing time and transmission deadlines are in general much shorter than the coherence time of signal attenuation due to rain, which is of the order of minutes or hours. A multiple beam antenna, equipped with TWTAs, has a power constraint of $P_i^{Tx} \leq P_0$ for each beam, where P_i^{Tx} is the transmission power allocated to the i^{th} cell and P_0 is the maximum power that the TWTA can support before saturation. We assume that every TWTA is driven near saturation for efficiency, and thus frequency multiplexing is not viable to feed multiple signals into a single TWTA. In such a situation, it is optimal to use full power of $P_i^{Tx} = P_0$ for all K active beams all the time and to achieve full channel capacities. This results in constant channel capacity C_i for each quasi-static channel over the time interval of interest as follows, with the assumption of the band-limited Shannon capacity:

$$C_i = W \log\left(1 + \frac{D^2}{\lambda^2 L^2}\frac{\alpha_i^2 P_0}{N_0 W}\right) \quad \text{bits/s}, \tag{1}$$

where α_i^2 (≤ 1) is quasi-static signal power attenuation due to the weather effects, N_0 is the noise power density, and W is the bandwidth used. Note that diffraction theory [2] gives a factor of $\frac{D^2}{\lambda^2 L^2}$ to scale transmit power on the far-field of satellite-to-earth link. Every terminal is assumed to be equipped with the same unit size of receiver antenna with equal gain and efficiency, so the problem does not change even if they are taken into account.

For given A_i and C_i,[3] our problem is formulated as follows:

$$\text{maximize} \quad \theta \tag{2}$$

$$\text{subject to} \quad 0 \leq \theta \leq 1 \tag{3}$$

$$\bar{d}_i(\theta A_i, z_i(t)C_i) \leq \Delta_i \quad \text{for every} \quad i = 1, ..., N \tag{4}$$

$$\text{and} \quad \sum_{i=1}^{N} z_i(t) \leq K \quad \text{with } z_i(t) = 0 \text{ or } 1 \text{ for every } t \text{ and } i, \tag{5}$$

where $\bar{d}_i$ is the average queuing delay of the i^{th} queue, $z_i(t)$ is a binary variable to indicate whether the i^{th} cell is served (1) or not (0) at discrete time $t = 0, 1, 2, ...$, and Δ_i (> 0) is a given average delay deadline for the i^{th} cell. The objective is to maximize the back-off parameter θ of incoming traffic A_i while assuring constraints of average delay $\bar{d}_i$ within a given target deadline Δ_i in condition (4). Condition (5) is for the transmitter-sharing constraint. Note that $\bar{d}_i$ is a function of $\theta A_i, C_i$, and $z_i(t)$. Throughout the paper, we consider only an average delay constraint, not a hard deadline constraint for each packet, i.e., $d_i(t) \leq \Delta_i$. Time-varying traffic demand and channel conditions make it infeasible to apply the hard deadline constraint all the time. Deep fading events even for a short duration, which happen in reality from time to time, can prohibit any utilization of resource and lead to violation of the constraint. Instead, one may consider outage events, which give the probability that the delay (or any quality of service in general) exceeds the given threshold, but this is beyond the scope of the paper and may be referred to [22], [23].

Fig. 3 shows the block diagram of this beam allocation and congestion control scheme. Perfect information on fixed-user channels and the system is assumed for decision-making. In practice, the information can be inferred based on estimates on a reciprocal channel or from direct feedback in a return channel. In [13], [14], it is shown that signal attenuation due to rain can be estimated with good accuracy (within a 1 dB error in 4 seconds ahead) based on a simple one-pole model. Thus, even in the presence of long propagation delays over satellite-ground links, beam allocation and congestion control parameters can be determined in advance, for this scheme to operate well. An analytical framework for channel estimation error and the resulting performance degradation can be found in [24].

B. Analysis

Since a binary variable $z_i(t)$ makes the problem complicated, we assume that the average delay depends only on the

[3]As for A_i, in general, we are more interested in packets/sec than bits/sec. Since C_i is given in terms of bits/sec, we should compare A_i with $C_i/\bar{l_p}$ (or $A_i \cdot \bar{l_p}$ with C_i), where $\bar{l_p}$ is the average packet length in terms of bits. For simplicity of the analysis, we assume that $\bar{l_p} = 1$ here. Since our result is mainly a steady-state average, we can simply replace C_i with $C_i/\bar{l_p}$ to get the result in terms of packets or A_i with $A_i \cdot \bar{l_p}$ in terms of bits.

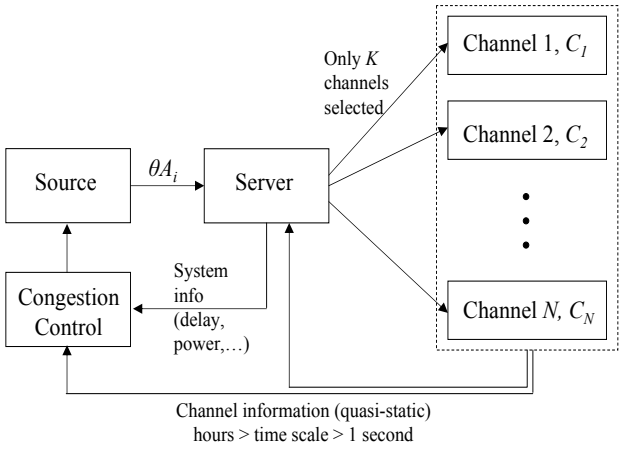


Fig. 3. Block diagram of the joint beam allocation and congestion control scheme

steady-state time-average of $z_i(t)$, $\bar{z}_i$. We replace the binary condition (5) with the following:

$$\sum_{i=1}^{N} \bar{z}_i \le K \quad \text{with} \quad 0 \le \bar{z}_i \le 1 \quad \text{for every } i. \tag{6}$$

This approximation is good when the active beams cycle through the cells more rapidly than the delay deadlines. When we assume that beam switching is very fast with no significant overhead cost, we can achieve $\bar{z}_i$ by time-sharing beams and changing $z_i(t)$ properly. The current switching technique can be as fast as milliseconds to feed dozens of active beams of a multiple beam antenna over microwave bands if a priori conditions such as traffic demand and channel conditions are provided. The faster switching technology is expected for a much larger scale of antenna (100 $\sim$ 1,000 beams) in the future, and the use of phased array antenna and solid state power amplifiers can be an alternative choice, which will be studied in Section 3.

We now solve the optimization problem in a closed form. Let us define $\bar{d}_i \triangleq g_i(\theta, \bar{z}_i)$. By nature of the delay function, g_i is an increasing function of θ and decreasing function of $\bar{z}_i$. If we assume that there exists an inverse function for $\bar{z}_i$, $g_i^{-1}(\Delta_i; \theta)$ with given θ, we have

$$\bar{z}_i \ge g_i^{-1}(\Delta_i; \theta) \tag{7}$$

from $g_i \le \Delta$. By combining this with condition (6), we have

$$g_i^{-1}(\Delta_i; \theta) \le 1 \quad \text{for every } i \text{ and} \quad \sum_{i=1}^{N} g_i^{-1}(\Delta_i; \theta) \le K, \tag{8}$$

and we can determine the optimum θ and $\bar{z}_i$. To get a meaningful insight for the analysis, we now simply assume that each queue of the satellite downlink beams resembles an $M/M/1$ queue. Suppose a Poisson arrival process of incoming traffic with average rate θA_i and an exponentially distributed traffic packet size with average transmission rate $\bar{z}_i C_i$. We have an average delay of the $M/M/1$ queue, given as

$$\bar{d}_i = g_i(\theta, \bar{z}_i) = \frac{1}{\bar{z}_i C_i - \theta A_i} \le \Delta_i \quad \text{for every } i. \tag{9}$$

By combining (8) and (9) and considering $0 \le \theta \le 1$, we give the optimum θ_{MBA} for joint beam allocation and congestion control with a multiple beam antenna as

$$\theta_{MBA} = \min\left\{ \min_i \left[\frac{1 - \frac{1}{C_i\Delta_i}}{\frac{A_i}{C_i}} \right], \frac{\frac{K}{N} - \frac{1}{N}\sum_i \frac{1}{C_i\Delta_i}}{\frac{1}{N}\sum_i \frac{A_i}{C_i}}, 1 \right\}, \tag{10}$$

and the corresponding $\bar{z}_i$ as

$$\begin{aligned} \bar{z}_i &\ge \theta_{MBA} \cdot \frac{A_i}{C_i} + \frac{1}{C_i\Delta_i} \\ &= \frac{A_i}{C_i} \cdot \min\left\{ \min_j \left[\frac{1 - \frac{1}{C_j\Delta_j}}{\frac{A_j}{C_j}} \right], \frac{\frac{K}{N} - \frac{1}{N}\sum_j \frac{1}{C_j\Delta_j}}{\frac{1}{N}\sum_j \frac{A_j}{C_j}}, 1 \right\} \\ &\quad + \frac{1}{C_i\Delta_i}, \end{aligned} \tag{11}$$

where the equality holds when $\theta_{MBA} < 1$ with $\bar{d}_i = \Delta_i$. When $\theta_{MBA} = 1$, congestion control is not needed and more $\bar{z}_i$ can be allocated with $\bar{d}_i < \Delta_i$. We observe that the multiple beam antenna provides different performance levels in the following two cases:

1) When no cell requires constantly a beam for itself (i.e., $\bar{z}_i < 1$ for every i), the multiple beam antenna system decides the congestion control parameter θ in terms of the average of expected parameter values and K/N, the ratio between the number of active beams and the total number of users.
2) When a cell requires constantly a beam for itself, the multiple beam antenna provides one whole active beam $\bar{z}_i = 1$ for the cell, but under-utilizes resources for others.

There is no feasible solution if $\frac{1}{C_i\Delta_i} > 1$ or $\sum_i \frac{1}{C_i\Delta_i} > K$, where the channel condition is too bad to support the required delay constraint. Since we have finite average delay deadline Δ_i, the additional price $\frac{1}{C_i\Delta_i}$ is imposed for each queue and $\sum_i \frac{1}{C_i\Delta_i}$ for the whole system. The smaller the deadlines and/or capacities are, the higher the price is. More urgent services and/or worse channel conditions require more beam allocation. Resource allocation is differentiated according to the different service requirements, even with a single back-off parameter θ.

III. Phased Array Antenna

A. Formulation

Since a phased array antenna can provide flexible beam size and shape (within the limitation of diffraction theory), we discard the concept of the cellular system illuminated by fixed size beams, but focus on individual user locations. We assume that there are M users on the Earth coverage area and each user expects to receive a different signal from a phased array antenna satellite. On the antenna aperture plane (ξ, η), which is assumed to be continuous over $|\xi| \le \frac{D}{2}$ and $|\eta| \le \frac{D}{2}$, the amplitudes and phases of array elements are controlled by a pattern-forming complex, to synthesize K ($\ll M$) active downlink signals (Fig. 4), whose number is limited by the number of on-board modulators. Denote the field distribution at the aperture as $V_i(\xi, \eta, t)$ for user i at time t. The field distributions for all users are linearly superimposed. We have

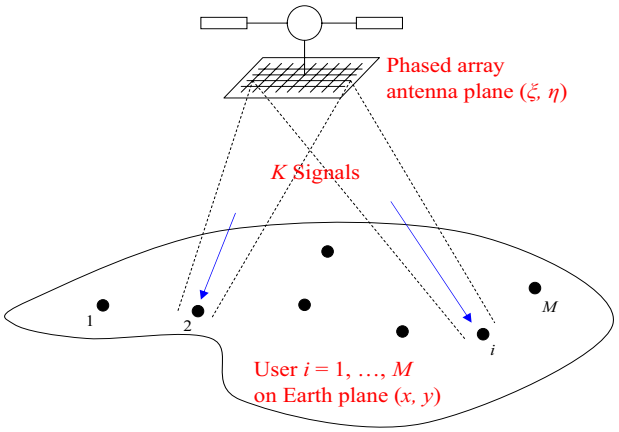


Fig. 4. A phased array antenna satellite, generating K active signals and serving M users on the Earth

the total field distribution of the antenna element at (ξ, η), given as

$$V_{sum}(\xi, \eta, t) = \sum_{i=1}^{M} V_i(\xi, \eta, t). \tag{12}$$

The aperture power density transmitted at (ξ, η) is $|V_{sum}(\xi, \eta, t)|^2$. On the antenna plane, each element has the maximum power density constraint of $|V_{sum}(\xi, \eta, t)|^2 \leq \rho_0$ if $|\xi| \leq \frac{D}{2}$ and $|\eta| \leq \frac{D}{2}$, for a constant ρ_0. We have $V_{sum}(\xi, \eta, t) = 0$ if $|\xi| > \frac{D}{2}$ or $|\eta| > \frac{D}{2}$. We can simplify this maximum power density constraint as follows:

$$\sum_{i=1}^{M} |V_i(\xi, \eta, t)|^2 \leq \rho_0. \tag{13}$$

The change rate of temperature in the SSPA due to heating is of the order of milliseconds while the modulation rate of V_i can be microseconds or smaller. Assuming that every V_i is independent of each other and its time average is equal to zero, we can ignore the heat accumulation and temperature change resulting from the linear terms $\sum_{i \neq k} V_i(\xi, \eta, t) V_k^*(\xi, \eta, t)$ in the power constraint of $|V_{sum}(\xi, \eta, t)|^2 = |\sum_i V_i(\xi, \eta, t)|^2 = \sum_i |V_i(\xi, \eta, t)|^2 + \sum_{i \neq k} V_i(\xi, \eta, t) V_k^*(\xi, \eta, t) \leq \rho_0$. This approximation suppresses the cross-products of different signals and thus decouples signals in amplitude/phase adjustment at each element. In practice, compared to the TWTA, the SSPA usually has a wider range of linearity before a sharp cut-off, so that linearity can be assured to allow superposition of signals at each element [25]. For the impact of nonlinearity on system performance, the detailed modeling and analysis can be referred to [26].

It is known that the received signal $U_i(x, y, t)$ on the Earth surface (x, y) is given by the two-dimensional Fourier transform of the field distribution $V_i(\xi, \eta, t)$ in case of far-field transmission [2]. The linearity of Fourier transform gives the received signal of

$$U_{sum}(x, y, t) = \sum_{i=1}^{M} U_i(x, y, t). \tag{14}$$

Every received signal waveform at (x, y) is assumed to have the identical propagation delay since the signal comes from the same satellite that is located very far. With a given signal-to-interference-and-noise-ratio (SINR), the capacity achievable for user i at (x_i, y_i) is

$$C_i(t) = W \log \left(1 + \frac{\alpha_i^2 P_i(x_i, y_i, t)}{W N_0 + \sum_{k \neq i} \alpha_i^2 P_k(x_i, y_i, t)} \right) \text{ bits/s}, \tag{15}$$

where $P_i(x_i, y_i, t) \triangleq |U_i(x_i, y_i, t)|^2$ is the received power of the i^{th} signal at (x_i, y_i), and all the other signals $P_k(x_i, y_i, t) \triangleq |U_k(x_i, y_i, t)|^2$ for $k \neq i$ at (x_i, y_i) are treated as interference.

We now solve the same throughput optimization problem as in (2). A phased array antenna has additional constraints of (12) $\sim$ (15). Only K users out of M can be served at one time (excluding the case of broadcasting same signals to different users). We have two time-varying control variables: (i) signal assignment $z_i(t)$ in constraints (4) and (5), and (ii) aperture field distribution $V_i(\xi, \eta, t)$, which is translated to power allocation $P_i(x, y, t)$. Different from the multiple beam antenna with binary (on-off) power allocation to each beam, the phased array antenna has multi-valued power allocation. Thus, the channel capacity in Eq. (15) is also multi-valued, which makes the problem complicated.

In subsection III-B we describe the optimum antenna gain patterning of $V_i(\xi, \eta, t)$, assuming that K users with $z_i(t) = 1$ are given. We focus on reducing interference to maximize the SINR with given transmit power. If interference is too severe between users, the sequential service should be deployed, which is in fact a scheduling problem. The dynamic scheduling policy will be discussed in subsection III-C. Then, we will show that the two decisions are eventually combined and made jointly depending on user location, requirements, and channel conditions.

B. Antenna Gain Patterning

Multiple spotbeams can give a higher total throughput than a single beam by concavity of the capacity function with respect to power [3] at the expense of possible interference from other active signal patterns. Interference is a monotonically decreasing function of the distance between active users inside the mainlobe and assumed to be negligible outside. That is, we mainly deal with interference due to the overlapping mainlobes for extremely close users. The sidelobe interference may not be insignificant, especially from the first sidelobes of adjacent beams that are much stronger than the beam of interest. However, the resulting degradation is smaller than that from the mainlobe, and easier to overcome at the smaller cost.[4]

If active users are scattered sparsely within the satellite coverage area, we can locate multiple narrowest spotbeams farther than the smallest beamwidth (of the mainlobe) $\frac{\lambda L}{D}$. The narrowest spotbeam results in the maximum gain in the

[4]When a beam is modeled as a sinc function, power difference is 13.5 dB between the main lobe and the first sidelobe, and 17.9 dB between the main lobe and the second sidelobe. One may locate nulls for interference without reducing the desired signal power much, or use error correction codes that can restore signals of up to 1 dB SINR.

direction of the main lobe, thus providing the maximum SINR and throughput. Each active signal pattern has a uniform amplitude over the entire aperture. Phases are adjusted to point the beam at the desired direction of (x_i, y_i). Every active signal is distributed over all antenna elements because a wide transmit antenna can make a narrow mainlobe. Multiple signals are linearly superimposed as

$$V_{sum}(\xi,\eta,t) = \sum_{i\in\Omega(t)} \frac{s_i(t)}{\sqrt{\sum_{i\in\Omega(t)} |s_i(t)|^2}} \cdot \sqrt{\rho_0} \cdot e^{j\frac{2\pi}{\lambda L}(x_i\xi + y_i\eta)} \tag{16}$$

for $|\xi| \le \frac{D}{2}$ and $|\eta| \le \frac{D}{2}$. At each time t, $\Omega(t)$ is the set of active signals and $s_i(t)$, which is independent of (ξ,η), represents the desired component of the i^{th} signal.[5]

Now, we assume that multiple close-in users, who are defined to be located within one beamwidth $\frac{\lambda L}{D}$ from each other, should be scheduled at the same timeslot. This can cause significant co-channel interference between close-in users. Some form of mitigating interference is necessary for efficient and reliable communications. We will show that the optimum pattern of each signal depends on the distances between users and their SNRs, and may consist in one of three alternatives regarding user scheduling: (i) complete cancellation of interference, (ii) optimum suppression of interference, and (iii) sequential service of close-in users.

First, we consider the case of complete cancellation with $U_i(x_k, y_k, t) = 0$ for active users i and k ($k \neq i$). Complete cancellation of interference leads to

$$\begin{aligned}\mu(\xi,\eta,t)V_i(\xi,\eta,t) &= s_i(t)e^{j\frac{2\pi}{\lambda L}(x_i\xi+y_i\eta)} \\ &\quad + \sum_{k\neq i} \gamma_{ik}(t) e^{j\frac{2\pi}{\lambda L}(x_k\xi+y_k\eta)}, \end{aligned} \tag{17}$$

where $\mu(\xi,\eta,t)$ is a scaling factor and $\gamma_{ik}(t)$ is a variable for interference cancellation. Unlike the case of sparse users in Eq. (16), $V_i(\xi,\eta,t)$ has more than one terms with a different phase for each. The first term of $s_i(t)e^{j\frac{2\pi}{\lambda L}(x_i\xi+y_i\eta)}$ maximizes the signal power at the desired location as in Eq. (16), whereas the term of $\gamma_{ik}(t)e^{j\frac{2\pi}{\lambda L}(x_k\xi+y_k\eta)}$ cancels interference toward (x_k, y_k) caused by the first term (of user i).

Next, we study optimum suppression, which does not necessarily achieve zero interference, but maximizes the SINR and throughput. Mathematically, the optimum suppression scheme outperforms complete cancellation all the time since complete cancellation has one more constraint of zero interference in the optimization problem. In practice, the interference cancellation terms in Eq. (17) can reduce the desired signal power too much. When active users are very close, the desired signal power of the complete cancellation scheme also approaches zero whereas optimum suppression makes compromise between reducing interference and maintaining the desired signal power level. Nevertheless, in some cases, the performance of optimum suppression is very close to that of complete cancellation, and interference is almost completely suppressed though it still has $U_i(x_k, y_k, t) \neq 0$. Optimum suppression is obtained as follows:

$$\mu(\xi,\eta,t)V_i(\xi,\eta,t) = \frac{\partial\theta}{\partial C_i}\frac{\partial C_i}{\partial V_i^*} + \sum_{k\neq i}\frac{\partial\theta}{\partial C_k}\frac{\partial C_k}{\partial V_i^*}, \tag{18}$$

[5]The detailed derivation of (16) – (20) is referred to [27].

where

$$\frac{\partial C_i}{\partial V_i^*} = \frac{WU_i(x_i,y_i,t)}{WN_0 + \sum_h \alpha_i^2 P_h(x_i,y_i,t)} \cdot \frac{\alpha_i^2}{\lambda L} e^{j\frac{2\pi}{\lambda L}(x_i\xi+y_i\eta)} \tag{19}$$

and

$$\begin{aligned}\frac{\partial C_k}{\partial V_i^*} &= -\frac{WU_i(x_k,y_k,t)}{WN_0 + \sum_h \alpha_k^2 P_h(x_k,y_k,t)} \\ &\quad \cdot \frac{\alpha_k^2 P_k(x_k,y_k,t)}{WN_0 + \sum_{h\neq k} \alpha_k^2 P_h(x_k,y_k,t)} \\ &\quad \cdot \frac{\alpha_k^2}{\lambda L} e^{j\frac{2\pi}{\lambda L}(x_k\xi+y_k\eta)}. \end{aligned} \tag{20}$$

The term of $\frac{\partial\theta}{\partial C_k}\frac{\partial C_k}{\partial V_i^*}$ for $k \neq i$ is adjusted to optimally suppress the interference of the i^{th} signal to the k^{th} user. In most cases, there is no closed-form solution. Instead, numerical answers can be obtained. Unlike the interference-free formulation, the optimization problem here is not convex any more due to the presence of interference in the capacity formula. The solution in (18) may be locally optimal only. Nevertheless, numerical examples will show that this solution is acceptable by outperforming complete cancellation.

Under potentially severe interference (e.g., in a very crowded area), spatial devision multiplexing (SDM) with interference suppression does not perform well because the desired signal is suppressed too much. The exact capacity in this situation, named a "Gaussian interference channel [28]," is still an open problem. So far, orthogonal schemes such as time or frequency division multiplexing (TDM or FDM) are known to give the best performance when interference is less than the desired signal power level but larger than some threshold [29]. Thus, when very close-in users suffer interference larger than the threshold, sequential service in a form of TDM outperforms any interference suppression scheme, by providing orthogonal signals. The threshold is decided by comparing interference suppression and sequential service, as shown in the following numerical examples.

For a simple example, suppose that two active users are separated by distance $l < \frac{\lambda L}{D}$ in the identical static conditions of average arrival rate and signal attenuation, which will result in identical antenna gain patterning and power/beam allocation for both the users. Throughout the paper, we define the carrier-to-interference ratio, CIR $\triangleq -20\log_{10}\left[\mathrm{sinc}\left(\frac{Dl}{\lambda L}\right)\right]$ (dB), where the sinc function is obtained by Fourier transforming the uniform field distribution over the whole antenna aperture with unit power, as we assume that the worst-case interference results from the other active user at the same SNR but with no interference suppression. In Fig. 5 we compare capacities of the two schemes of complete cancellation in Eq. (17) and optimum suppression in Eq. (18), by changing the CIR. The scheme without interference suppression in Eq. (16) is also shown as a benchmark. At the high SNR value of $\frac{E_b}{N_0} = 10.2$ dB, where E_b is the average signal energy per bit, the gap between optimum suppression and complete cancellation is very small. On the other hand, in the low SNR region of $\frac{E_b}{N_0} = 1.76$ dB, the gap between two schemes is wider for every CIR. The complete cancellation scheme is even worse than the scheme without interference suppression for a wide range of CIR. If the CIR is smaller than some threshold of CIR* that is decided by comparing interference

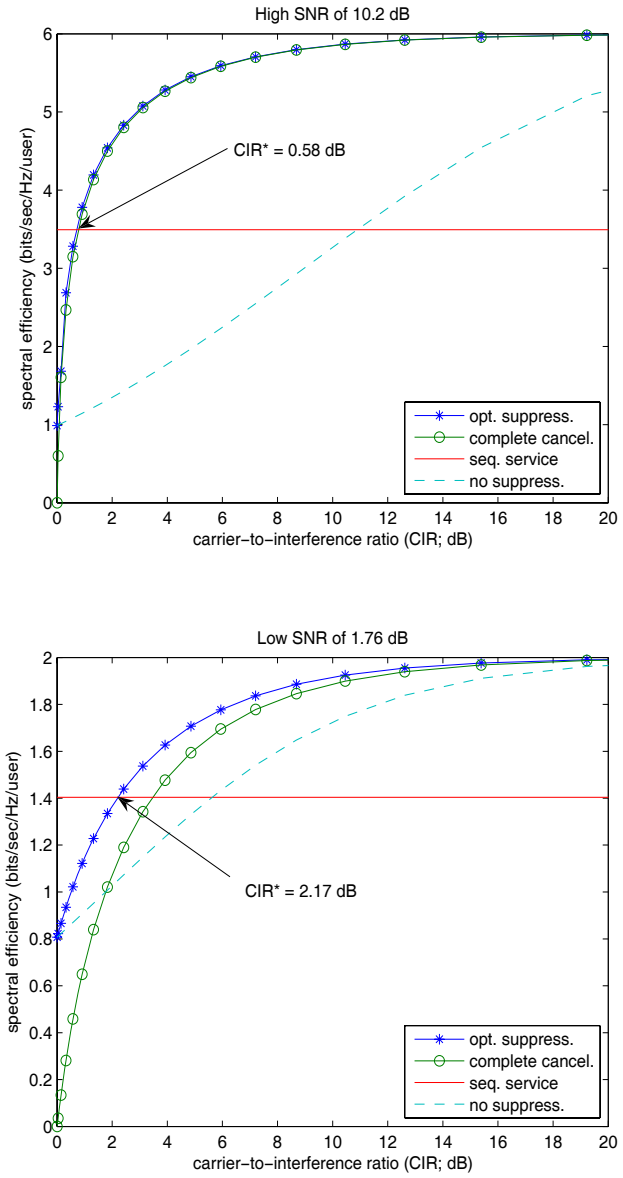


Fig. 5. Capacity of one user as a function of CIR at high and low SNR of $\frac{E_b}{N_0} = 10.2$ and $\frac{E_b}{N_0} = 1.76$ dB, respectively

suppression and sequential service (CIR$^* = 0.58$ dB at the high SNR and CIR$^* = 2.17$ dB at the low SNR), signal degradation is too severe due to co-channel interference and it is better to provide sequential service. If CIR $>$ CIR* and $l < \frac{\lambda L}{D}$, active users share the bandwidth and timeslots, and appropriate optimum antenna gain patterning is deployed with interference suppressed, depending on the operating SNR level.

The two SNR values of $\frac{E_b}{N_0} = 10.2$ dB and 1.76 dB in the example are derived from the Shannon limit

$$\frac{E_b}{N_0} \geq \frac{2^{\frac{R}{W}} - 1}{R/W} \tag{21}$$

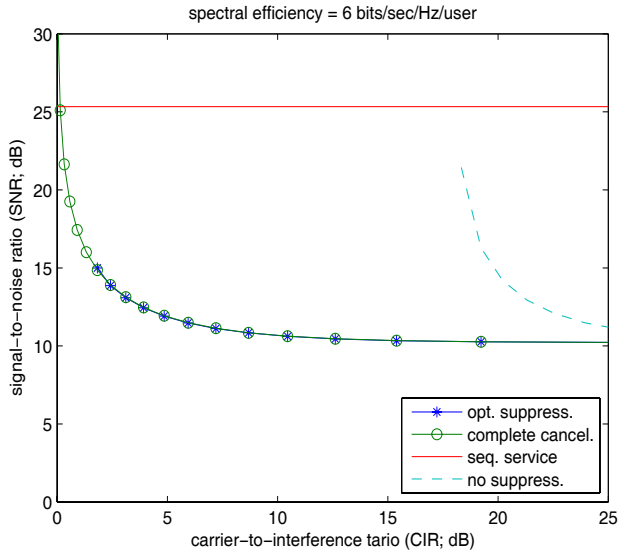
with $\frac{R}{W} = 6$ and 2 bits/sec/Hz, respectively, where R is the bit rate (no larger than the channel capacity) for reliable transmission. One can support up to 6 bits/sec/Hz by the use of 64 quadrature amplitude modulation (64-QAM) and up to 2 bits/sec/Hz by the use of quadrature phase shift keying (QPSK), respectively. In practice, current satellite communication systems operate at the spectral efficiency of 1$\sim$2 bits/sec/Hz. Since bandwidth is precious in satellite communications, bandwidth efficient modulation (BEM) such as 16 or 64-QAM is considered in advanced satellite systems. Nonlinearity of the satellite channel complicates a use of the multi-layered envelope modulation scheme. Moreover, the linear increase of spectral efficiency with BEM requires the near-exponential increase of SNR, as shown in (21). Even with powerful coding schemes such as Turbo codes and low density parity check (LDPC) codes, huge power consumption makes it almost infeasible to apply BEM for a TWTA-based satellite system. A future satellite system will be designed to provide better spectral efficiency as high as 6 bits/sec/Hz by the use of phased array antenna and SSPAs that benefit from a wide range of linearity and power allocation flexibility.

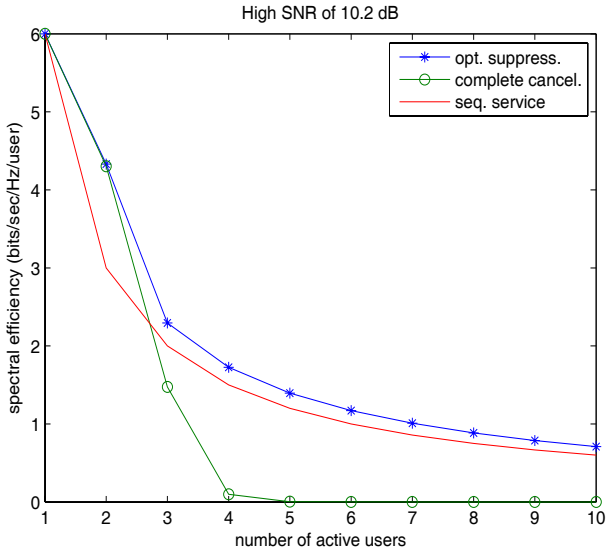
Fig. 6 plots the SNR values of the different schemes according to the CIR for a required spectral efficiency of 6 and 2 bits/sec/Hz/user. As the CIR decreases, the scheme of optimum suppression and the scheme without interference suppression cannot support the required spectral efficiency at all due to the high interference level. Note that optimum suppression behaves as if no interference suppression is applied, shown in Fig. 5, under heavy interference of CIR $= 0$ dB. This suggests the use of adaptive modulation and coding scheme, by which high spectral efficiency can be supported for users with little interference while low spectral efficiency is supported under heavy interference for robust communications. This is in fact the same principle with the terrestrial cellular network, where the users in the middle of cells and those near cell edges are allocated different rates.

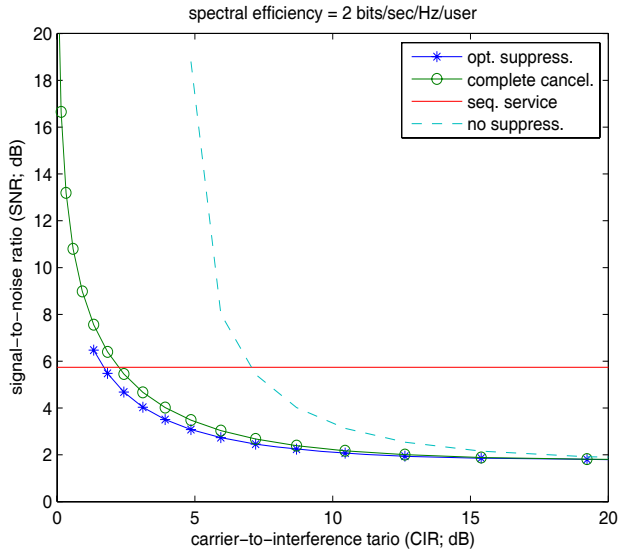
Fig. 7 shows performance comparison of the different schemes in terms of the number of uniformly located active users within $0 < l \leq \frac{\lambda L}{D}$ for high and low SNR levels as in Fig. 5. We again assume identical static conditions of average arrival rate and signal attenuation for all the users. We observe the advantage of multiple signals over a single beam of sequential service for two or three active users, in spite of power loss from interference suppression. The complete cancellation scheme is very vulnerable to more than three active users. As the number of active users increases, the gap between optimum suppression and sequential service decreases and is negligible. Eventually, sequential service outperforms optimum suppression for a large number of active users (not shown in the figures). This is consistent with what is shown in the Gaussian interference channel problem and our two-user example in Fig. 5: sequential service is better than SDM (with or without interference suppression) under severe interference. Thus, the use of simple sequential service can be recommended when an area is crowded with many users.

C. Beam Scheduling

We now address the problem of beam scheduling by considering co-channel interference between close-in active users and the corresponding gain patterning. We assume the use of complete cancellation. As shown in Fig. 5, complete







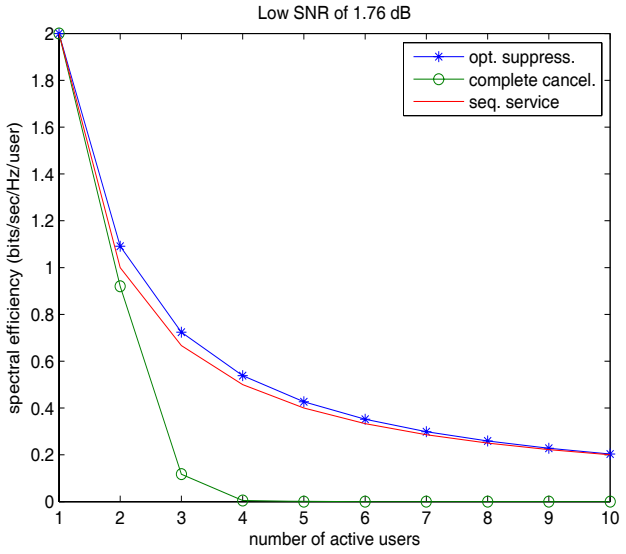


Fig. 6. SNR as a function of CIR at high and low spectral efficiency of 6 and 2 bits/sec/Hz/user, respectively

Fig. 7. Capacity of one user as a function of the number of active users within one beamwidth at high and low SNR of $\frac{E_b}{N_0} = 10.2$ and $\frac{E_b}{N_0} = 1.76$ dB, respectively

cancellation of interference is close approximation to optimum suppression in the high SNR region until active users become too close. Even at low SNR, complete cancellation can achieve more than 90% of optimum suppression and be a reasonable approximation. For power $P_i(t)$ allocated to user i in service at time t, we model channel capacity, given as

$$C_i(t) = W \log\left(1 + \frac{\alpha_i^2 H_i P_i(t)}{W N_0}\right) \quad \text{bits/s}, \tag{22}$$

where the power loss by interference suppression is considered in H_i $(0 \leq H_i \leq 1)$. The value of H_i is given by a function of beamwidth and the distance between active users. If active user i at (x_i, y_i) has another active user k at (x_k, y_k) and the distance l between the two users satisfies $l = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2} < \frac{\lambda L}{D}$, the complete cancellation scheme gives

$$H_i = 1 - \text{sinc}^2\left[\frac{D(x_i - x_k)}{\lambda L}\right] \text{sinc}^2\left[\frac{D(y_i - y_k)}{\lambda L}\right], \tag{23}$$

which is derived in the Appendix. If $l \geq \frac{\lambda L}{D}$, we ignore interference and have $H_i = 1$. Note that H_i depends only on the distance to other active users, not on other parameters such as SNR and traffic demand. Here, the value of H_i only considers the nearest active user. We further assume that $x_i - x_k = l$ and $y_i - y_k = 0$ for mathematical simplicity, and then have

$$H_i = 1 - \text{sinc}^2\left(\frac{l}{\lambda L/D}\right), \tag{24}$$

which is no less than the value of H_i in (23). This approximation is reasonable because we do not provide antenna gain patterning with interference suppression for more than

three active users in a crowded area as shown in the previous antenna gain patterning examples. Advantages of this model are the decoupling of different signals and their capacities with respect to a set of allocated power $\{P_i\}_{i=1}^{M}$, and the derivation of a convex optimization problem. Moreover, since P_i is defined without additional terms for interference suppression, we can obtain a simple and familiar total power constraint:

$$\sum_{i=1}^{M} P_i(t) \le P_{total}. \tag{25}$$

The detailed derivation of P_i and the total power constraint are referred to the Appendix.

The phased array antenna can cycle very rapidly among users. The question is how the active users should be chosen and clustered each time to serve the back-logged and newly arrived packets of all M users in terms of $z_i(t)$ and $P_i(t)$, and maximize the throughput in terms of time-varying $\theta(t)$. We now solve the optimum scheduling problem of (2) $\sim$ (5) with additional constraints of (22), (24), and (25). Since it is hard to solve the binary problem, we relax the binary constraint of $z_i(t)$ and instead observe whether $P_i(t)$ is equal to zero or not.

The corresponding Lagrangian function is $J(P_i) = \theta - \sum_i \kappa_i \cdot (\bar{d}_i - \Delta_i) - \Lambda \cdot (\sum_i P_i - P_{total}) - \sum_i \nu_i \cdot (-P_i)$, where Lagrangian variables κ_i (≥ 0) are for the average delay constraints, Λ (≥ 0) for the total power constraint, and ν_i (≥ 0) for $-P_i \le 0$. The last constraint is added to see which users should be served with non-zero P_i. The Kuhn-Tucker condition [30] yields $\nu_i > 0$ if $P_i = 0$ and $\nu_j = 0$ if $P_j > 0$. Differentiating $J(P_i)$ with respect to P_i gives

$$\frac{\partial J}{\partial P_i} = \frac{\partial \theta}{\partial P_i} + \kappa_i \left| \frac{\partial \bar{d}_i}{\partial P_i} \right| - \Lambda + \nu_i, \tag{26}$$

where we use $\frac{\partial \bar{d}_i}{\partial P_i} = -\left| \frac{\partial \bar{d}_i}{\partial P_i} \right|$ since the average delay is a decreasing function of the allocated power. At the optimum power $P_i^* > 0$, we have $\nu_i = 0$ and $\left. \frac{\partial J}{\partial P_i} \right|_{P_i = P_i^*} = 0$. From Eq. (26), we obtain

$$\left. \frac{\partial \theta}{\partial P_i} \right|_{P_i = P_i^*} + \kappa_i \left. \left| \frac{\partial \bar{d}_i}{\partial P_i} \right| \right|_{P_i = P_i^*} = \Lambda < \left. \frac{\partial \theta}{\partial P_i} \right|_{P_i = 0} + \kappa_i \left. \left| \frac{\partial \bar{d}_i}{\partial P_i} \right| \right|_{P_i = 0}, \tag{27}$$

where the inequality holds when we assume concavity of throughput θ and average delay $\bar{d}_i$ in terms of P_i in the same way as that of capacity. When $P_j = 0$ is optimal for user j, we have $\nu_j \ge 0$ and $\left. \frac{\partial J}{\partial P_j} \right|_{P_j = 0} = 0$. Again from Eq. (26), we obtain

$$\left. \frac{\partial \theta}{\partial P_j} \right|_{P_j = 0} + \kappa_j \left. \left| \frac{\partial \bar{d}_j}{\partial P_j} \right| \right|_{P_j = 0} + \nu_j = \Lambda \ge \left. \frac{\partial \theta}{\partial P_j} \right|_{P_j = 0} + \kappa_j \left. \left| \frac{\partial \bar{d}_j}{\partial P_j} \right| \right|_{P_j = 0}. \tag{28}$$

By comparing the right sides of (27) and (28) with respect to common Λ, we find that the optimum policy serves K users with the highest value of

$$\left. \frac{\partial \theta}{\partial P_i} \right|_{P_i = 0} + \kappa_i \left. \left| \frac{\partial \bar{d}_i}{\partial P_i} \right| \right|_{P_i = 0} = \frac{\alpha_i^2 H_i}{N_0} \left[\left. \frac{\partial \theta}{\partial C_i} \right|_{C_i = 0} + \kappa_i \left. \left| \frac{\partial \bar{d}_i}{\partial C_i} \right| \right|_{C_i = 0} \right], \tag{29}$$

where the chain rule is applied with $\left. \frac{\partial C_i}{\partial P_i} \right|_{P_i = 0} = \frac{\alpha_i^2 H_i}{N_0}$. The result implies that we have to select

- better channel conditions with higher α_i^2,
- less interference with higher H_i, and
- higher marginal returns of the composite cost function (of the throughput with delay penalties), $f = \theta - \sum \kappa_i \cdot (\bar{d}_i - \Delta_i)$, in terms of allocated capacity with higher $\left. \frac{\partial f}{\partial C_i} \right|_{C_i = 0}$.

Though we solved two separate subproblems of antenna gain patterning and scheduling, the optimum solution suggests that the two designs should be made jointly since the selection of K users and the power loss H_i from interference suppression depend on each other. This is in fact the hardest part to solve in practice. A simple but near-optimum algorithm will be proposed in Section V.

IV. Performance Comparison of Phased Array Antenna and Multiple Beam Antenna

Here we compare the average steady-state performance of the phased array antenna with that of the multiple beam antenna. While the multiple beam antenna has a power constraint for each beam, the phased array antenna can provide any power level (up to the total power) for a signal. We will show that due to its flexibility for power allocation the phased array antenna can give higher throughput (i.e., θ) than the multiple beam antenna when a user requires constantly a beam for her/himself.

Let Q_i denote the average power allocated for user i by a phased array antenna when the signal is on, given as $Q_i = E[P_i | z_i = 1]$. With $\bar{z}_i = \Pr[z_i = 1]$, we have an average power constraint of

$$\sum_i \bar{P}_i = \sum_i \bar{z}_i Q_i \le P_{total}. \tag{30}$$

Furthermore, we assume that the average capacity C_i^{avg} is a simple function of Q_i and $\bar{z}_i$, given as

$$\begin{aligned} C_i^{avg} &= E\left[z_i \cdot W \log\left(1 + \frac{\alpha_i^2 H_i P_i}{W N_0} \right) \right] \\ &\simeq \bar{z}_i \cdot W \log\left(1 + \frac{\alpha_i^2 H_i Q_i}{W N_0} \right) \end{aligned} \tag{31}$$

as if the signal for user i has either $P_i = Q_i$ with the signal on or $P_i = 0$ with the signal off. We note that this approximation is reasonable when $P_i(t)$ does not change much around Q_i when $z_i(t) = 1$. If bandwidth is very broad, we can further approximate $C_i^{avg} \simeq E\left[z_i \cdot \frac{\alpha_i^2 H_i P_i}{N_0} \right] = \bar{z}_i \frac{\alpha_i^2 H_i Q_i}{N_0}$. With the multiple beam antenna, power allocation with $z_i(t) = 1$ is fixed at $P_i(t) = P_0 = \frac{P_{total}}{K}$, where the maximum power P_0 of a TWTA before saturation is assumed to be identical for all K TWTAs, so that we have $Q_i = \frac{P_{total}}{K}$ for every i.

To compare the performances of the two antenna schemes, we provide a simple example, where M users have traffic arrival rates of $A_1 \leq A_2 \leq \cdots \leq A_{M-1} \leq A_M$. Each user is assumed to belong to a different cell in case of the multiple beam antenna. From the average delay constraint with $M/M/1$ queue approximation and the same average delay deadline $\Delta_i = \Delta$ for every i, the optimum θ_ϕ of the phased array antenna is given as

$$\begin{aligned}\theta_\phi &= \frac{1}{A_i}\left(C_i^{avg} - \frac{1}{\Delta}\right) \\ &= \frac{1}{A_i}\left(\bar{z}_i \cdot \frac{\alpha_i^2 H_i Q_i}{N_0} - \frac{1}{\Delta}\right) \quad \text{if } W \to \infty \qquad (32) \\ &\leq \frac{1}{A_i}\left(\bar{z}_i \cdot \frac{Q_i}{N_0} - \frac{1}{\Delta}\right) \quad \text{for every } i \\ &= \frac{1}{\frac{1}{M}\sum^M A_i}\left(\frac{1}{M}\frac{P_{total}}{N_0} - \frac{1}{\Delta}\right), \qquad (33)\end{aligned}$$

where we use the following for the last equality: if $Z = \frac{X_i}{Y_i}$ for every $i = 1, \cdots, M$ then, $Z = \frac{\frac{1}{M}\sum_{i=1}^M X_i}{\frac{1}{M}\sum_{i=1}^M Y_i}$. From the result in Section II, the throughput of the multiple beam antenna depends on whether a user requires constantly a beam for her/himself. We assume the identical channel condition of $\alpha_i^2 = 1$ for every i and the broad bandwidth of $W \to \infty$, and characterize the asymptotic performance based on traffic distribution of $A_1 \leq \cdots \leq A_M$. If the most demanding user M does not require constantly a beam for her/himself with $\bar{z}_M < 1$, we have $\theta_{MBA} = \frac{\frac{K}{M} - \frac{1}{M}\sum_i \frac{1}{C_i \Delta}}{\frac{1}{M}\sum_i \frac{A_i}{C_i}}$ from (10), which is the same as θ_ϕ in (33). Otherwise (i.e., if $\bar{z}_M = 1$), we have $\theta_{MBA} = \frac{1}{A_M}\left(\bar{z}_M \cdot \frac{P_{total}}{KN_0} - \frac{1}{\Delta}\right)$, which is compared with θ_ϕ in (33) as follows:

$$\frac{\theta_\phi}{\theta_{MBA}} = \frac{A_M}{\frac{1}{M}\sum_{i=1}^M A_i} \cdot \frac{\frac{1}{M}\frac{P_{total}}{N_0} - \frac{1}{\Delta}}{\frac{1}{K}\frac{P_{total}}{N_0} - \frac{1}{\Delta}} \to \frac{A_M}{\frac{1}{M}\sum_{i=1}^M A_i} \cdot \frac{K}{M} \qquad (34)$$

if $\Delta \to \infty$. Thus, the asymptotic gain of the phased array antenna over the multiple beam antenna with very broad bandwidth, large delay, no interference, and identical channel condition is given as

$$\frac{\theta_\phi}{\theta_{MBA}} = \max\left\{\frac{A_{max}}{\bar{A}} \cdot \frac{K}{M}, 1\right\}, \qquad (35)$$

where A_{max} is the maximum arrival rate (equal to A_M in this case) and $\bar{A} = \frac{1}{M}\sum_{i=1}^M A_i$ is the average of all the arrival rates.

For a numerical example, suppose that 100 users ($M = 100$) can receive up to 20 signals ($K = 20$) at each timeslot. We assume exponentially distributed incoming traffic of $A_i = A_0 \exp(i \cdot \beta_{exp})$, where the parameters of A_0 and β_{exp} shape the traffic distribution among users. Fig. 8 compares θ_ϕ and θ_{MBA} with respect to the traffic distribution (in terms of $A_{max}/\bar{A}$) and the CIR that is a function of the distance between users. The parameter values used are $A_0 = 0.0496$, $\beta_{exp} = 0.0535$, $P_{total} = 2$ kW, and $\Delta = \infty$. We assume perfect scheduling and give an upper bound steady-state result if the total coverage area of K orthogonal beams is smaller than the total area occupied by M users. To mitigate interference as the CIR decreases, the multiple beam antenna reduces the number of active beams since interference suppression cannot be supported. On the other hand, the phased array antenna considers possible interference between close-in users, and then selects the better of interference suppression and sequential service, based on our derivation in (24) and (32). The result shows that the phased array antenna always performs better than the multiple beam antenna.[6] The advantage of the phased array antenna over the multiple beam antenna can be observed especially when traffic distribution is very unbalanced ($A_{max}/\bar{A} > 8$) and/or there is a moderate level of interference between active users, so that interference suppression can be used (0.58 dB < CIR < 2.42 dB). We remark that the performance difference between the two antenna schemes can be reduced if each queue has its own back-off parameter (i.e., different θ_i for each i) and low-traffic users are better served.

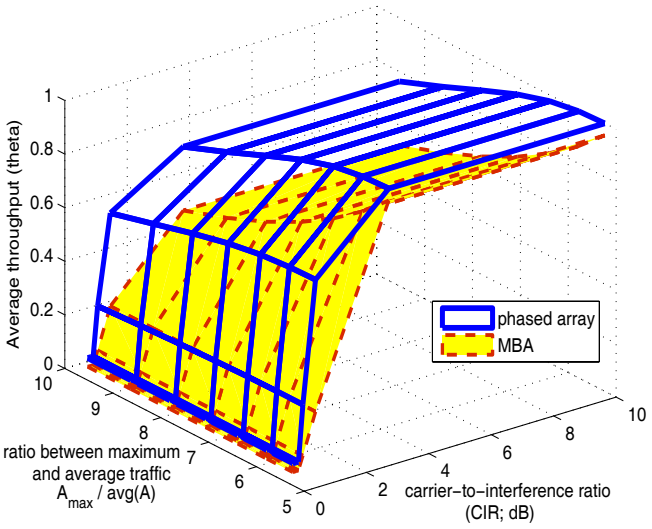


Fig. 8. Steady-state performance comparison of θ of the phased array antenna and the multiple beam antenna as a function of traffic distribution in terms of $A_{max}/\bar{A}$ and CIR

V. Near-Optimum Algorithm for Phased Array Antenna

So far, we have focused on the steady-state performance of the two antenna schemes. Now, we move forward to study real-time operation. In subsection III-C, we derived an optimum scheduling policy, which is, however, too complicated to apply in practice. Here, we try to decouple the joint decisions of optimum scheduling, and develop a near-optimum, low-complexity, and real-time algorithm of performing active user selection, antenna gain patterning, power allocation, and admission control.

We first observe that the average delay constraints $\bar{d}_i \leq \Delta_i$ alone cannot guarantee the system stability, and the number of packets accumulated in the satellite queue should be well-controlled. Here we propose a constraint to the total of

[6]When there is little interference and traffic distribution is not extremely unbalanced (which is not shown in the plot), the two antennas give the same performance.

accumulated delays for all the users. Let F_i denote the amount of accumulated traffic in the i^{th} queue, $B_{i,\tau}$ the amount of traffic having waited for τ timeslots in the i^{th} queue (i.e., the packets were admitted τ timeslots ago, thus the current delay is τ and increases every timeslot), and Ψ some constant to upper-bound the total of accumulated delays. The constraint is given as

$$\sum_{i=1}^{M} \bar{d}_i \cdot F_i = \sum_{i=1}^{M} \sum_{\tau=2}^{t} \tau \cdot B_{i,\tau} + 1 \cdot \theta \sum_{i=1}^{M} A_i \leq \Psi, \qquad (36)$$

with $F_i = \sum_{\tau=1}^{t} B_{i,\tau} = \sum_{\tau=2}^{t} B_{i,\tau} + \theta A_i$. We assume for mathematical tractability that newly admitted packets have delay 1 (i.e., $B_{i,1} = \theta A_i$). The value of Ψ can be decided by the system capacity. If we assume that the maximum queue size is Q_{max}, we can have a reasonable range of $\Psi < Q_{max} \cdot \sum_{i=1}^{M} \Delta_i$.

With this constraint we can decide the amount of newly admitted traffic, based on the amount and delays of backlogged traffic in the system. For the composite cost function of $f = \theta - \sum_i \kappa_i (\bar{d}_i - \Delta_i)$ in subsection III-C, the Lagrangian multiplier κ_i is non-zero only if $\bar{d}_i \geq \Delta_i$ by the Kuhn-Tucker condition. The selection algorithm depends on the admission control policy of the system because $\left.\frac{\partial f}{\partial C_i}\right|_{C_i=0}$ reduces to $\left.\frac{\partial \theta}{\partial C_i}\right|_{C_i=0}$ if the average delay constraint of the i^{th} user is met. To calculate $\left.\frac{\partial f}{\partial C_i}\right|_{C_i=0} = \left.\frac{\partial \theta}{\partial C_i}\right|_{C_i=0}$, we use constraint (36), which is equivalent to

$$\theta \leq \frac{1}{\sum_{i=1}^{M} A_i} \left[\Psi - \sum_{i=1}^{M} \sum_{\tau=2}^{t} \tau \cdot B_{i,\tau} \right]. \qquad (37)$$

To maximize θ we minimize $\sum_{i=1}^{M} \sum_{\tau=2}^{t} \tau \cdot B_{i,\tau}$, which is to serve the packets with the largest delay as long as the average delay constraint $\bar{d}_i \leq \Delta_i$ is satisfied. When traffic demand is beyond the system capacity ($\sum A_i > \Psi$), the equality holds in (37) and we have

$$\begin{aligned} \left.\frac{\partial f}{\partial C_i}\right|_{C_i=0} &= \left.\frac{\partial \theta}{\partial C_i}\right|_{C_i=0} \\ &= \frac{-1}{\sum A_i} \frac{\partial}{\partial C_i} \left[\sum_{i=1}^{M} \sum_{\tau=2}^{t} \tau \cdot B_{i,\tau} \right] \Bigg|_{C_i=0} \\ &= \frac{1}{\sum A_i} d_i^{max}, \end{aligned} \qquad (38)$$

where $d_i^{max} = \max_{\{B_{i,\tau}>0\}} \tau$ is the largest packet delay in the i^{th} queue. By allocating one unit of C_i, we can clear one unit of $B_{i,\tau}$, and it is the optimum to serve the packet with the largest delay. Fig. 9 plots the change of the accumulated delay, which is piecewise linear and convex, with respect to the allocated capacity. If demand is within the system capacity, we can still claim that serving the packet with the largest delay is a sensible solution, considering the future throughput and delay. Thus far, we have derived a closed-form solution of $\left.\frac{\partial f}{\partial C_i}\right|_{C_i=0}$ in terms of the maximum delay in the queue. It is reminded that the optimum scheduling policy should also consider factors of α_i^2 and H_i from $\left.\frac{\partial C_i}{\partial P_i}\right|_{P_i=0}$.

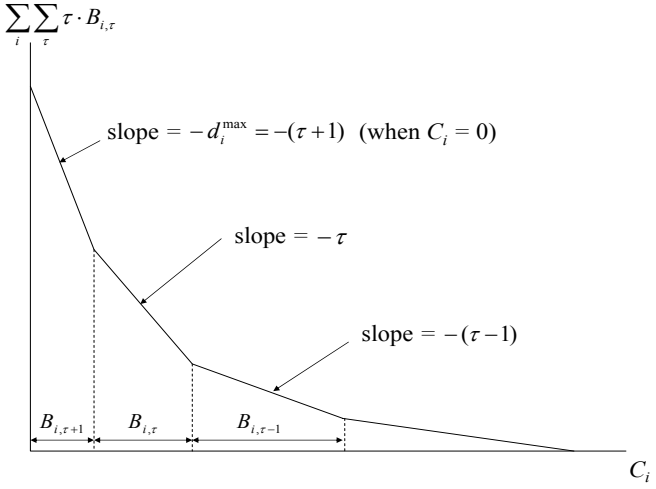


Fig. 9. A plot of the accumulated delay with respect to the allocated capacity

The optimum selection algorithm is to choose K active users with i) $\bar{d}_i = \Delta_i$ and then ii) largest $\alpha_i^2 H_i d_i^{max}$. However, the selection of active users affects the value of H_i and vice versa, which makes the selection process very complicated and time-consuming. Instead, we suggest a sub-optimum selection algorithm that chooses the most demanding user based on average delay constraints and $\alpha_i^2 d_i^{max}$, and then next active users by considering the interference level decided by the users already selected. The detail of how to select active users is as follows.

1) Select users with $\bar{d}_i = \Delta_i$ regardless of their distances from each other, and update H_i that results from the selected users to remaining users, using Eq. (24).
2) For remaining users (or for all M users if no one is selected in Step 1), select the user with the biggest value of $\alpha_i^2 H_i d_i^{max}$ if the user satisfies $l > l^*$, where l is the minimum distance to other active users already selected, and l^* is the distance threshold whether sequential service or interference suppression is implemented (in general $0.1\frac{\lambda L}{D} \leq l^* \leq 0.5\frac{\lambda L}{D}$ and). If $l \leq l^*$, reject the user, proceed to the next user and repeat Step 2.[7]
3) After adding an active user, update interference level H_i that results from the user selected in Step 2 to remaining users, using Eq. (24).
4) Repeat Step 2 and 3 until either
   a) selecting K active users or
   b) scanning all M users. (In this case, serve only less than K users.)

We allocate the optimum power $P_i^\star$ according to the equation in (27): every selected user has an identical marginal return, given as

$$\frac{\alpha_i^2 H_i}{1 + \frac{\alpha_i^2 H_i P_i^\star}{W N_0}} \cdot d_i^{max} = \Lambda', \qquad (39)$$

where Λ' is a Lagrangian multiplier and decided by jointly solving with the total power constraint (25). After power

[7]Since CIR is a function of l only as in subsection III-B, we can compare CIR with $\text{CIR}^* = -20\log_{10}\left[\text{sinc}\left(\frac{Dl^*}{\lambda L}\right)\right]$ instead of l with l^*.

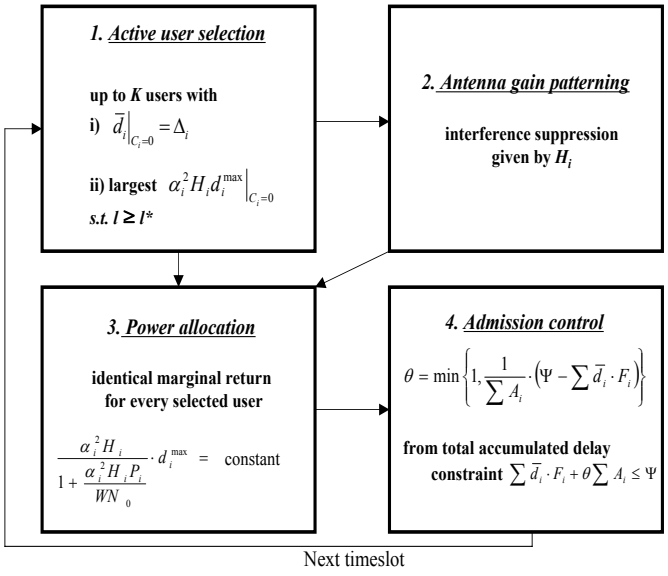


Fig. 10. A diagram for the resource scheduling algorithm

allocation and the corresponding packet service, θ is decided by the remaining traffic, given in constraint (37). The value of θ is fed back to the sources to throttle the incoming traffic rates. The required inputs to the algorithm are user locations, traffic arrival rates, link conditions, and average delay requirements. Under ideal operations such as fast feedback of θ and instantaneous control of traffic rate, there is no packet dropping once packets are admitted to the satellite onboard. In reality, for rapidly time-varying parameters, if any, compared to long feedback latency, parameter mismatch can result in some performance loss. Fig. 10 shows a diagram for the algorithm. For each timeslot, active user selection, antenna gain patterning, power allocation, and admission control are performed in order. The newly admitted and remaining backlogged traffic is the input for the active user selection at the next timeslot.

VI. Simulation Results

We now compare the simulation performance of the algorithm with the steady state analysis developed from the optimum policy. We consider 49 $(= M)$ users uniformly located in a 7×7 planar grid, and change the deterministic distances l between the nearest neighbor users from $0.01\frac{\lambda L}{D}$ to $\frac{\lambda L}{D}$ (i.e., the two-dimensional plane that covers users expands gradually). The location (x_i, y_i) of user i is defined as $x_i/l = \{(i-1) \mod 7\} + 1$ and $y_i/l = \lceil i/7 \rceil$ with $l \leq x_i, y_i \leq 7l$, where mod is a modular function and $\lceil \cdot \rceil$ is a ceiling function. The satellite is assumed to be able to provide up to 20 $(= K)$ signals simultaneously. User traffic is assumed to be linearly increasing with an arrival rate $A_i = i \cdot \beta_{lin}$ for user $i = 1, \cdots, M$, where β_{lin} is a constant slope. The Matlab code simulates two forms of real-time traffic: Poisson arrival random traffic and constant streaming deterministic traffic. With the Poisson arrival traffic, we perform 5 simulations and average the results for every distance. Each of the simulations runs for 500 timeslots. The streaming traffic gives an identical result for every simulation due to its deterministic traffic pattern. Inputs to the simulation are assumed to be known a priori perfectly: static users in fixed location, and quasi-static channel conditions and average traffic rates. In reality, the required inputs should be fed back from users or estimated/predicted using past data, channel reciprocity (for channel conditions), etc, as described in subsection II-A. If propagation delay is too large and/or the inputs are time-varying too rapidly, some conservative estimation may be used to give performance bounds with some margin to the accurate operating point.

Fig. 11 plots the time average of $\theta(t)$ from simulation results for the two types of traffic and compares them with steady-state analytic results. For this example, the algorithm achieves 97.5% of the steady-state performance with the stream traffic and 94.1% with the Poisson traffic, in case of no interference, for $l = \frac{\lambda L}{D}$. The algorithm is simulated in discrete timeslots while the steady-state analysis is assumed to use idealistic continuous-time scheduling. Thus, the simulation may not use all K signals even with no interference, which occurs for the Poisson traffic of this example. When a small number of demanding users consume a large or whole amount of onboard power, other users cannot be served at the same time. The number of active signals is reduced and the throughput also decreases. In addition, the random and bursty Poisson arrival traffic deviates from the steady-state traffic pattern. As a result, the number of active signals used and the throughput are reduced further, compared to the constant stream traffic that is a good approximation to the steady-state. This example shows that efficient resource scheduling for random bursty data traffic is more difficult than scheduling for steady circuit traffic.

In the middle range of distance, $0.25\frac{\lambda L}{D} < l < 0.6\frac{\lambda L}{D}$, where the better choice between interference suppression and the sequential service can change frequently, the simulation performance of the Poisson traffic is not so close to the steady-state solution and the performance of the stream traffic. This is due to the sub-optimum selection process, which is simplified to update the interference level only after user selection. Nevertheless, the algorithm still achieves more than 85% of the steady-state solution except for a small range of distances.

The steep decrease of the average throughput of simulations at $l = 0.25\frac{\lambda L}{D}$ is due to the distance threshold value $l^* = 0.25\frac{\lambda L}{D}$ in this simulation, which is the optimum value chosen in a heuristic manner. The users at the distance of $l < l^*$ from already selected active users are not selected though they have higher values of $\alpha_i^2 H_i d_i^{max}$ than others because their selection will create significant interference to some of already selected users.[8] The simulation performance depends heavily on the distance threshold l^*, which decides how many active signals are used simultaneously under interference. If the value is too big, the algorithm under-performs for small interference because it averts interference too much and loses the throughput gain from having multiple parallel channels. If the value is too small, the algorithm may send too many signals under severe interference and lose efficiency as well. The key point to the optimum scheduling is to increase the number of active signals unless interference becomes too

[8]The worse performance of interference suppression than that of the sequential service in Subsection III-B is reminded at the very small distance between active users.

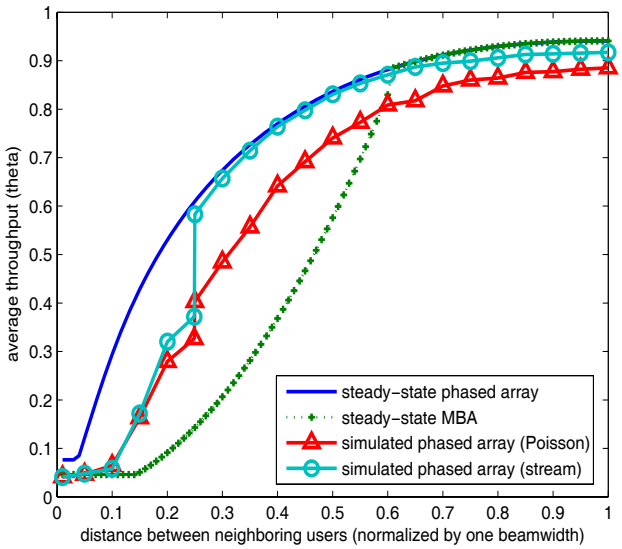


Fig. 11. Comparison of average throughputs between algorithm simulation (with Poisson arrival random traffic and constant stream) and steady-state analytic results (for the phased array antenna and the multiple beam antenna) with constant traffic slope $\beta_{lin} = 0.115$

severe, by choosing the optimum value of l^*.

In summary, our algorithm achieves very good near-optimality in case of small interference. Under potentially severe interference, the algorithm shows sub-optimality but avoids time-consuming exhaustive search and thus, can provide timely service to latency-sensitive users.

VII. Conclusion

In the practical system with finite length queues, if there is no congestion control mechanism, excessive packet arrival will result in packet loss after the queue becomes full, and can initiate unnecessary automatic repeat request (ARQ) functions and retransmissions for dropped packets, inducing possibly more congestion. To meet average delay constraints and to stabilize the system, we should consider some form of congestion control of incoming traffic and couple it with resource allocation. This is very important for data communications with bursty and unscheduled computer traffic.

In this paper, for the multiple beam antenna and the phased array antenna respectively, we have found the optimal solution for joint resource allocation and congestion control over satellite downlinks based on incoming traffic, link qualities, and average delay constraints. We have modeled a throughput maximization problem by assuming quasi-static channel conditions and very fast beam switching techniques. Our congestion control scheme is rate-based and implemented at satellite on-board. We have analytically found the optimal congestion control parameter and the corresponding beam allocation/antenna gain patterning method. We have then developed a simple, near-optimum, and real-time algorithm for active user selection, antenna gain patterning, power allocation and admission control of the phased array antenna system. The algorithm serves users with better channel conditions, less interference and larger queuing delays. Though we have made some simple assumptions for our analysis, such as broad bandwidth, $M/M/1$ queuing model, and single-variable abstraction for admission control, our study shows that the phased array antenna when combined with a specially developed algorithm of resource scheduling and congestion control can provide efficient communications for bursty traffic users and/or under heavy interference.

Appendix

Here we derive power loss H_i of the complete interference cancellation scheme in Eq. (23) and the total power constraint (25). Assuming $\mu(\xi,\eta,t) = \mu(t)$ in Eq. (17) to be constant over every (ξ,η) for simplicity of calculation and only two active users of i and k within one beamwidth, we combine $V_i(\xi,\eta,t)$ of Eq. (17) with the zero interference constraint of $U_i(x_k,y_k,t) = \frac{1}{\lambda L}\int\int V_i(\xi,\eta,t)e^{-j\frac{2\pi}{\lambda L}(x_k\xi+y_k\eta)}d\xi d\eta = 0$, where $U_i(x,y,t)$ is given as the two-dimensional Fourier transform of the field distribution $V_i(\xi,\eta,t)$. Then, we obtain

$$\begin{aligned}\gamma_{ik}(t) &\approx -\frac{1}{D^2}\cdot s_i(t)\int_{-\frac{D}{2}}^{\frac{D}{2}}\int_{-\frac{D}{2}}^{\frac{D}{2}} e^{j\frac{2\pi}{\lambda L}\{(x_i-x_k)\xi+(y_i-y_k)\eta\}}d\xi d\eta \\ &= -s_i(t)\cdot \mathrm{sinc}\left[\frac{D(x_i-x_k)}{\lambda L}\right]\mathrm{sinc}\left[\frac{D(y_i-y_k)}{\lambda L}\right]. \quad (40)\end{aligned}$$

The maximum power density constraint of $|V_i(\xi,\eta,t)|^2 + |V_k(\xi,\eta,t)|^2 \le \rho_0$ gives

$$\mu^2(t) = \frac{|s_i(t)|^2+|s_k(t)|^2}{\rho_0}\cdot G_i \quad (41)$$

where $s_k(t)$ is the desired component of the k^{th} signal and we define $G_i \triangleq 1 - \mathrm{sinc}^2\left[\frac{D(x_i-x_k)}{\lambda L}\right]\mathrm{sinc}^2\left[\frac{D(y_i-y_k)}{\lambda L}\right]$.

We calculate the desired signal power $P_i(x_i,y_i,t) = |U_i(x_i,y_i,t)|^2$, given as

$$\begin{aligned}P_i(x_i,y_i,t) &= \frac{|s_i(t)|^2}{\mu^2(t)}\cdot\frac{D^4}{\lambda^2L^2}\cdot G_i \quad (42)\\ &= P_i^{no-int}(t)\cdot G_i, \quad (43)\end{aligned}$$

with

$$P_i^{no-int}(t) = \frac{|s_i|^2}{|s_i|^2+|s_k|^2}\cdot\frac{\rho_0 D^4}{\lambda^2L^2}, \quad (44)$$

which is the received power when user i has no other active user within one beamwidth, and simply denoted by $P_i(t)$ in subsection III-C. Thus, we conclude $H_i \equiv G_i$ in Eq. (23) from $P_i(x_i,y_i,t)/P_i^{no-int}(t)$. We also see

$$P_i^{no-int}(t) + P_k^{no-int}(t) = P_i(t) + P_k(t) = \frac{\rho_0 D^4}{\lambda^2L^2}, \quad (45)$$

where $\frac{\rho_0 D^4}{\lambda^2L^2} \triangleq P_{total}$ is defined as the total power. By extending to general M users, we obtain the total power constraint (25).

References

[1] R. C. Johnson, *Antenna Engineering Handbook*, 3rd ed. McGraw-Hill, 1993.

[2] J. Goodman, *Introduction to Fourier Optics*, 3rd ed. Roberts & Company, 2005.

[3] J. P. Choi and V. W. S. Chan, "Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2983-2993, Nov. 2005.
[4] S. Egami, "A power-sharing multiple-beam mobile satellite in Ka band," *IEEE J. Select. Areas Commun.*, vol. 17, no. 2, pp. 145-152, Feb. 1999.
[5] S. Egami and M. Kawai, "An adaptive multiple beam system concept," *IEEE J. Select. Areas Commun.*, vol. SAC-5, no. 4, pp. 630-636, May 1987.
[6] J. Romero-Garcia and R. De Gaudenzi, "On antenna design and capacity analysis for the forward link of a multiple power controlled satellite CDMA network," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1230-1244, July 2000.
[7] K.-T. Ko and B. R. Davis, "A space-division multiple-access protocol for spot-beam antenna and satellite-switched communication network," *IEEE J. Select. Areas Commun.*, vol. SAC-1, no. 1, pp. 126-132, Jan. 1983.
[8] T. R. Henderson and R. H. Katz, "Transport protocols for internet-compatible satellite networks," *IEEE J. Select. Areas Commun.*, vol. 17, no. 2, pp. 326-344, Feb. 1999.
[9] I. F. Akyildiz, G. Morabito, and S. Palazzo, "TCP-Peach: a new congestion control scheme for satellite IP networks," *IEEE/ACM Trans. Networking*, vol. 9, no. 3, pp. 307-321, June 2001.
[10] D. Katabi, M. Handley, and C. Rohrs, "Internet congestion control for high bandwidth-delay product networks," in *Proc. ACM Sigcomm 2002*, Pittsburgh, PA, Aug. 2002.
[11] U. K. Mishra, P. Parikh, and Y.-F. Wu, "AlGaN/GaN HEMTs - an overview of device operation and applications," *Proc. IEEE*, vol. 90, no. 6, pp. 1022-1031, June 2002.
[12] P. Khan, L. Epp, and A. Silva, "A Ka-band wideband-gap solid-state power amplifier: architecture identification," IPN Progress Report 42-162, Jet Propulsion Laboratory, Pasadena, CA, Aug. 15, 2005.
[13] J. P. Choi, "Channel prediction and adaptation over satellite channels with weather-induced impairments," Master's Thesis, EECS, MIT, May 2000.
[14] J. P. Choi and V. W. S. Chan, "Predicting and adapting satellite channels with weather-induced impairments," *IEEE Trans. Aerospace Electronic Syst.*, vol. 38, no. 3, pp. 779-790, July 2002.
[15] C. Loo and J. S. Butterworth, "Land mobile satellite channel measurements and modeling," *Proc. IEEE*, vol. 86, no. 7, pp. 1442-1463, July 1998.
[16] W. Li, C. L. Law, V. K. Dubey, and J. T. Ong, "Ka-band land mobile satellite channel model incorporating weather effects," *IEEE Commun. Lett.*, vol. 5, no. 5, pp. 194-196, May 2001.
[17] G. E. Corraza and F. Vatalaro, "A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 738-742, Aug. 1994.
[18] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, pp. 237-252, 1998.
[19] J. Y. Le Boudec, "Rate adaptation, congestion control and fairness: a tutorial," Mar. 2008. [Online] Available: http://ica1www.epfl.ch/PS_files/LEB3132.pdf.
[20] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.
[21] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 4th ed. Pearson Education Inc., 2008.
[22] A. Conti, M. Z. Win, and M. Chiani, "On the inverse symbol-error probability for diversity reception," *IEEE Trans. Commun.*, vol. 51, no. 5, pp. 753-756, May 2003.
[23] A. Conti, M. Z. Win, and M. Chiani, "Slow adaptive M-QAM with diversity in fast fading and shadowing," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 895-905, May 2007.
[24] W. M. Gifford, M. Z. Win, and M. Chiani, "Antenna subset diversity with non-ideal channel estimation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, part 1, pp. 1527-1539, May 2008.
[25] E. Lutz, M. Werner, and A. Jahn, *Satellite Systems for Personal and Broadband Communications*. Springer, 2000.
[26] A. Conti, D. Dardari, and V. Tralli, "An analytical framework for CDMA systems with a nonlinear amplifier and AWGN," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1110-1120, July 2002.
[27] J. P. Choi, "Resource allocation and scheduling for communication satellites with advanced transmission antennas," Ph.D. Thesis, EECS, MIT, Sept. 2006. [Online] Available: http://dspace.mit.edu/handle/1721.1/38298.
[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
[29] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 5, pp. 607-615, Sept. 1985.
[30] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.

Jihwan P. Choi [S'00, M'06] received his B.S. degree from Seoul National University, Seoul, Korea in 1998 and both S.M. and Ph.D. degrees from Massachusetts Institute of Technology in 2000 and 2006, respectively, all in Electrical Engineering and Computer Science. From 1999 to 2006, he participated in the next-generation-Internet (NGI) satellite network project at MIT. Since 2006, he has been researching and developing system design and standardization of 3G/4G commercial wireless communications at Marvell Semiconductor Inc., Santa Clara, California, where he is currently a Staff Systems Engineer. His research interests are in wireless/space communications and cross-layer network optimization. Dr. Choi is a recipient of KFAS (the Korea Foundation for Advanced Studies) fellowships.

Vincent W. S. Chan [S'69, M'88, SM'92, F'94], the Joan and Irwin Jacobs Professor of EECS, MIT, received his BS(71), MS(71), EE(72), and Ph.D.(74) degrees in EE from MIT. From 1974 to 1977, he was an assistant professor, EE, at Cornell University. He joined MIT Lincoln Laboratory in 1977 and had been Division Head of the Communications and Information Technology Division until becoming the Director of the Laboratory for Information and Decision Systems (1999-2007). In July 1983, he initiated the Laser Intersatellite Transmission Experiment Program and in 1997, the follow-on GeoLITE Program. In 1989, he formed the All-Optical-Network Consortium among MIT, AT&T and DEC. He also formed and served as PI the Next Generation Internet Consortium, ONRAMP among AT&T, Cabletron, MIT, Nortel and JDS, and a Satellite Networking Research Consortium formed between MIT, Motorola, Teledesic and Globalstar. This year he helped formed and is currently a member of the Claude E. Shannon Communication and Network Group at the Research Laboratory of Electronics of MIT. He is on the Board of Vitesse Corporation and a Member of the Corporation of Draper Laboratory, member of Eta-Kappa-Nu, Tau-Beta-Pi and Sigma-Xi, a Fellow of the IEEE and the Optical Society of America.