Power and Server Allocation in a Multi-Beam Satellite with Time Varying Channels

Michael J. Neely Eytan Modiano Charles Rohrs
MIT--LIDS: {mjneely@mit.edu, modiano@mit.edu, crohrs@mit.edu}

***Abstract* -- We consider power and server allocation in a multi-beam satellite downlink which transmits data to *N* different ground locations over *N* time-varying channels. Packets destined for each ground location are stored in separate queues, and the server rate for each queue *i* depends on the power $p_i(t)$ allocated to that server and the channel state $c_i(t)$ according to a concave rate-power curve $\mu_i(p_i, c_i)$. We establish the capacity region of all arrival rate vectors $(\lambda_1,...,\lambda_N)$ which admit a stabilizable system. For the case when channel states and arrivals are *iid* from timeslot to timeslot, we develop a particular power allocation policy which stabilizes the system whenever the rate vector lies within the capacity region. Such stability is guaranteed even if the channel model and the specific arrival rates are unknown. As a special case, this analysis verifies stability of the "Choose-the-*K*-Largest-Connected-Queues" policy when channels can be in one of two states (ON or OFF) and *K* servers are allocated at every timestep ($K<N$). These results are extended to treat a joint problem of routing and power allocation, and a throughput maximizing algorithm for this joint problem is constructed. Finally, we address the issue of inter-channel interference, and develop a modified policy when power vectors are constrained to feasible activation sets. Our analysis and problem formulation is also applicable to power control for wireless applications.**

I. Introduction

In this paper we consider power allocation in a satellite which transmits data to *N* ground locations over *N* different downlink channels. Each channel is assumed to be time varying (e.g., due to changing weather conditions) and the overall channel state is described by the ergodic vector process $\vec{C}(t) = (c_1(t), \ldots, c_N(t))$. Packets destined for ground location *i* arrive from an input stream X_i and are placed in an output queue to await processing (Fig. 1). The servers of each of the *N* output queues may be activated simultaneously at any time *t* by assigning to each a power level $p_i(t)$, subject to the total power constraint $\sum p_i(t) \leq P_{tot}$. The transmission rate of each server *i* depends on the allocated power $p_i(t)$ and on the current channel state $c_i(t)$ according to a general concave rate-power curve $\mu_i(p_i, c_i)$. A controller allocates power to each of the *N* queues at every instant of time in reaction to channel state and queue backlog information. The goal of the controller is to stabilize the system and thereby achieve maximum throughput and maintain acceptably low levels of unfinished work in all of the queues.

We establish the capacity region of the system: the multi-dimensional region of all arrival rate vectors $(\lambda_1,...,\lambda_N)$ which admit a stabilizable system under some power allocation policy. Stability in this region holds for general ergodic channel and packet arrival processes. It is shown that if the channel model and arrival rates are known, any power allocation policy which stabilizes the system--possibly by making use of special knowledge of future events--can be transformed into a stabilizing policy which considers only the current channel state. We next consider the case of a slotted time system when arrivals and channel state vectors vary *iid* from one timeslot to the next, but the channel model and the exact values of arrival rates $(\lambda_1,...,\lambda_N)$ are unknown. A particular power allocation policy is developed which stabilizes the system whenever the rates $(\lambda_1,...,\lambda_N)$ are within the capacity region. This result is extended to treat a joint routing and power allocation problem, and a simple policy is developed which maximizes throughput and ensures stability whenever the system is stabilizable. Finally, we address the issue of interchannel interference due to bandwidth limitations, and develop a modified policy when power vectors are constrained to activation sets. This analysis makes use of a Lyapunov function defined over the state of the queues.

Previous work on queue control problems for satellite and wireless applications is found in [1-6]. In [1] a parallel queue system with a single server is examined, where every timeslot the transmit channels of the queues vary between ON and OFF states and the server selects a queue to service from those that are ON. The capacity region of the system is developed when packet arrivals and channel states are *iid* Bernoulli processes, and sto-

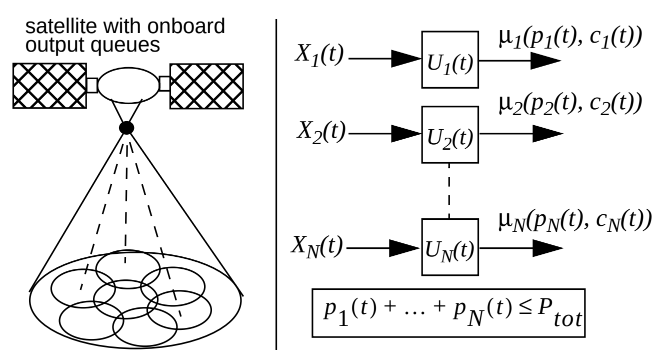

Figure 1: A multi-beam satellite with *N* time-varying downlink channels and *N* onboard output queues.

chastic coupling is used to show optimality of the "*Serve-the-Longest-Connected-Queue*" policy in the symmetric situation that arrival and channel processes are identical for all queues (i.e., $\lambda_1 = \ldots = \lambda_N$, $p_1^{on} = \ldots = p_N^{on}$). Such a server allocation problem can be viewed as a special case of our power allocation formulation, and in Section IV we verify stability of the "*Serve-the-K-Longest-Queues*" policy for symmetric and asymmetric systems with multiple servers.

In [2] a wireless network of queues is analyzed when input packets arrive according to memoryless processes and have exponentially distributed length. A Lyapunov function is used to establish a stabilizing routing and scheduling policy under network connectivity constraints. Such a technique for proving stability has also been used in the switching literature [7-10]. In [7] an $N \times N$ packet switch with blocking is treated and input/output matching strategies are developed to ensure 100% throughput whenever arrival rates are within the capacity region. In [9],[11] the method of Lyapunov stability analysis is improved upon and used to prove queues are not only stable but have finite backlog moments.

The main contribution in this paper is the formulation of a general power control problem for multi-beam satellites and the development of throughput maximizing power and server allocation algorithms for the system. The method extends to other wireless networking problems where power allocation and energy efficiency is a major issue. Recent work in [12] treats a problem of minimizing the total energy expended to transmit blocks of data arriving to a single queue, and it is shown that power control can be effectively used to extend longevity of network elements. In [13] power allocation for wireless networks is addressed. The authors consider ON/OFF type power allocation policies and observe that for random networks, capacity regions are not extended much by including more power quantization levels. Our capacity results in Section III illustrate that the capacity region is often considerably extended if multiple power levels are utilized for the satellite downlink problem.

In the next section, we introduce the power and server allocation problems. In Section III we develop several stability results for single queue systems with ergodic and non-ergodic processing rates $\mu(t)$, and establish the capacity region of the satellite downlink with power control. In Section IV a stabilizing power allocation policy is developed for systems with *iid* inputs and channel states. In Section V a joint routing and power allocation policy is treated using similar analysis, and in Section VI we extend the problem to treat channel interference issues.

II. Power and Server Allocation

Consider the N queue system of Fig. 1. Each time varying channel i can be in one of a finite set of states S_i. We represent the channel process by the channel vector $\vec{C}(t) = (c_1(t), \ldots, c_N(t))$, where $\vec{C}(t) \in S_1 \times \ldots \times S_N$. Channels hold their state for timeslots of length T, with transitions occurring on slot boundaries $t=kT$. The channel process is assumed to be ergodic and yields time average probabilities $\pi_{\vec{C}}$ for each state $\vec{C}$. At every timeslot, the server transmission rates can be controlled by adjusting the power allocation vector $\vec{P}(t) = (p_1(t), \ldots, p_N(t))$ subject to the total power constraint $\sum p_i(t) \le P_{tot}$.

For any given state c_i of downlink channel i, there is a corresponding rate-power curve $\mu_i(p_i, c_i)$ which is increasing, concave, and continuous in the power parameter (Fig. 2). This power curve could represent the logarithmic Shannon capacity curve of a Gaussian channel, or could represent a rate curve for a specific set of coding schemes designed to achieve a sufficiently low probability of error in the given channel state. In general, any practical set of power curves will have the concavity property, reflecting diminishing returns in transmission rate with each incremental increase in signal power.

The continuity property is less practical. A real system will rely on a finite databank of coding schemes, and hence actual rate/power curves restrict operation to a finite set of points. For such a system, we create a new, *virtual power curve* $\tilde{\mu}_i(p_i, c_i)$ by a piecewise linear interpolation of the operating points (see Fig. 3a). Such virtual curves have the desired continuity and concavity properties, and are used as the true curves in our power allocation algorithms. Clearly a virtual system which allocates power according to the virtual curves has a capacity region which contains that of a system restricted to allocate power on the vertex points. However, the capacity regions are in fact the same, as any point on a virtual curve can effectively be achieved by time-averaging two or more feasible rate-power points over many timeslots. Indeed, in Section IV we design a stabilizing

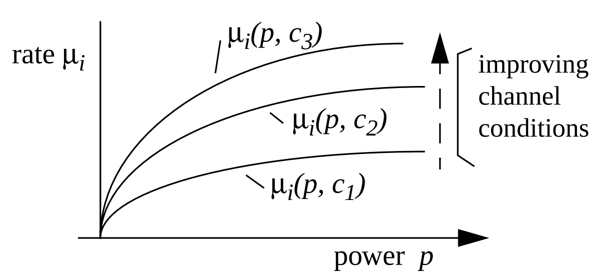

Figure 2: A set of concave power curves $\mu_i(p_i, c_i)$ for channel states c_1, c_2, c_3.

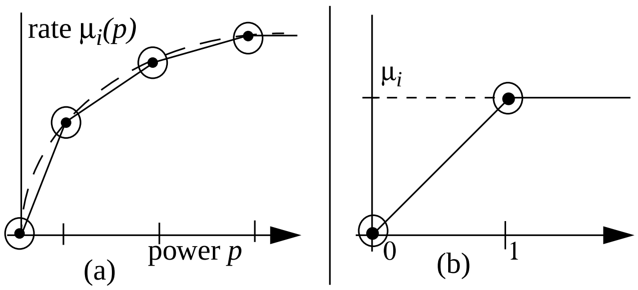

Figure 3: Virtual power curves for systems with a finite set of operating points.

policy for any set of concave power curves which naturally selects vertex points at every timeslot if power curves are piecewise linear.

This power allocation formulation generalizes a simpler problem of server allocation. Assume that there are K servers, and every timeslot the servers are scheduled to serve K of the N queues ($K<N$). A given queue i transmits data at a fixed rate μ_i whenever a server is allocated to it, and transmits nothing when no server is allocated. This problem can be transformed into a power allocation problem by defining the virtual power constraint $\sum p_i(t) \leq K$ and the virtual power curves:

$$\tilde{\mu}_i(p) = \begin{cases} \mu_i p \ , & p \in [0, 1] \\ \mu_i \ \ , & p > 1 \ . \end{cases} \quad (1)$$

Such a virtual curve contains the feasible points $(p = 0, \tilde{\mu}_i = \mu_i)$, $(p = 1, \tilde{\mu}_i = \mu_i)$ (see Fig. 3b).

Example Server Allocation Algorithm: One might suspect the policy of serving the K fastest, non-empty queues would maximize data output and achieve stability. However, we provide the following counterexample which illustrates this is not the case. Consider a 3-queue, 2-server system with constant processing rates (μ_1, μ_2, μ_3) = (1, 1, 1/2). All arriving packets have length L=1 and arrive according *iid* Bernoulli processes with the packet arrival probabilities $(p_1, p_2, p_3) = (p, p, (1- p^2)/2 + \varepsilon)$, where p<1/2 and ε>0.

Note that the policy of serving the two fastest queues removes a server from queue 3 whenever there are simultaneous arrivals at queues 1 and 2. This happens with probability p^2, and hence the time average processing rate at queue 3 is no more than $(1- p^2)/2$ (where the factor 1/2 is due to the rate of server 3). This effective service rate cannot support the input rate, and hence queue 3 is unstable under this server allocation policy. However, the system is clearly stabilizable: The policy of always allocating a server to queue 3 and using the remaining server to process packets in queues 1 and 2 stabilizes all queues.

III. Stability and the Downlink Capacity Region

To understand the capacity region of the downlink system, we first develop a simple criterion for stability of a single queue with an input stream $X(t)$ and a time varying processing rate $\mu(t)$ (Fig. 4). We assume the input stream is ergodic with rate λ. However, because an arbitrary power control scheme could potentially yield a non-ergodic processing rate, we must consider general processes $\mu(t)$ which may or may not have well defined time averages. We make the following definitions:

$X(t)$ = Total amount of bits that arrived during [0, t].
$U(t)$ = Unprocessed bits in the queue at time t.
$\mu(t)$ = Instantaneous bit processing rate in the server.

$$\lambda = \lim_{t \to \infty} \frac{X(t)}{t} \quad , \quad \underline{\mu} = \liminf_{t \to \infty} \frac{1}{t}\int_0^t \mu(\tau)d\tau . \quad (2)$$

The above limits[1] exist with probability 1. We assume the processing rate is always bounded above by some maximum value ($\mu(t) \leq \mu_{max}$ for all t) and hence $0 \leq \underline{\mu} \leq \mu_{max}$. As a measure of the fraction of time the unfinished work in a queue is above a certain value M, we define the following "overflow" function $g(M)$:

$$g(M) = \limsup_{t \to \infty} \frac{1}{t}\int_0^t 1_{[U(\tau) > M]} d\tau \quad (3)$$

where the indicator function 1_E used above takes the value 1 whenever event E is satisfied, and 0 otherwise.

Definition: A single server queueing system is *stable* if $g(M) \to 0$ as $M \to \infty$.

Notice that if steady state behavior exists and if sample paths of unfinished work in the queue are ergodic, the overflow function $g(M)$ is simply the steady state probability that the unfinished work in the queue exceeds the value M. Stability in this case is identical to the usual notion of stability defined in terms of a vanishing complementary occupancy distribution (see [2,7,11]).

Lemma 1: For the single queue system (Fig. 4) with general input and server rate processes $X(t)$ and $\mu(t)$, a necessary condition for stability is $\lambda \leq \underline{\mu}$. If the arrival process $X(t)$ and the rate process $\mu(t)$ evolve according to an ergodic, finite state Markov chain, then a sufficient

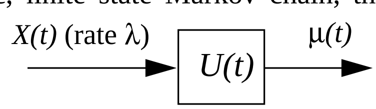

Figure 4: A single queue system with input stream $X(t)$ and time varying processing rate $\mu(t)$.

1. Where the *lim inf* of a function $f(t)$ is defined:
$$\liminf_{t \to \infty} f(t) = \lim_{t \to \infty} \left[\inf_{\tau \geq t} f(\tau)\right].$$

condition for stability is $\lambda < \underline{\mu}$.

Proof: The sufficient condition for Markovian arrivals and linespeeds is well known (see large deviations results in [12]). The necessary condition is proven in Appendix A1 by showing that if $\lambda > \underline{\mu}$, there exist arbitrarily large times t_i such that the average fraction of time the unfinished work is above M during $[0, t_i]$ is greater than a fixed constant for any value of M. ❑

We use this single queue result to establish the capacity region of the power constrained, multi-channel system of Fig. 1. We define the capacity region as the compact set of points $\Omega \subset [0, \infty)^N$ such that all queues of the system can be stabilized (with some power allocation policy) whenever the vector of input bit rates $\vec{\lambda} = (\lambda_1, \ldots, \lambda_N)$ is strictly in the interior of Ω, and, conversely, no stabilizing policy exists whenever $\vec{\lambda} \notin \Omega$. (The system may or may not be stable if $\vec{\lambda}$ lies on the boundary of the capacity region).

Assume arrivals and channel states are modulated by an ergodic, finite state Markov chain, and transitions occur on timeslots of duration T. Let $\pi_{\vec{C}}$ represent the steady state probability that the channel vector is in state $\vec{C} = (c_1, \ldots, c_N)$.

<u>Theorem 1</u>: The capacity region of the downlink channel of Fig. 1 with power constraint P_{tot} and rate-power curves $\mu_i(p, c_i)$ is the set of all input rate vectors $\vec{\lambda}$ such that there exist power levels $p_i^{\vec{C}}$ satisfying $\sum_{i=1}^{N} p_i^{\vec{C}} \le P_{tot}$ for all channel states $\vec{C}$, and such that:

$$\lambda_i \le \sum_{\vec{C}} \pi_{\vec{C}} \mu_i(p_i^{\vec{C}}, c_i)\,. \qquad (4)$$

Proof: Using the stationary policy of allocating a power vector $\vec{P}^{\vec{C}} = (p_1^{\vec{C}}, \ldots, p_N^{\vec{C}})$ whenever the system is in channel state $\vec{C}$ creates a Markov modulated processing rate $\mu_i(t)$ for all queues i, with an average rate given by the right hand side of inequality (4). Lemma 1 thus ensures stability whenever the vector $\vec{\lambda}$ satisfies (4) with strict inequality in all entries.

In Appendix A2 we show that restricting power control to such stationary policies (which use only the current channel state $\vec{C}(t)$ when making power allocation decisions) does not restrict the capacity region, and hence the region in (4) captures all input rates which yield stable systems. ❑

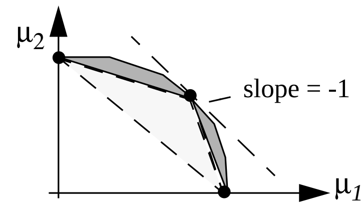

Figure 5: A capacity region for 2 static channels.

In the case when the channel does not vary but stays fixed, we have one power curve $\mu_i(p)$ for each queue i, and the expression for the capacity region above can be greatly simplified:

<u>Corollary 1</u>: For static channels, the capacity region is the set of all $\vec{\lambda}$ vectors such that:

$$\sum_{i=1}^{N} \mu_i^{-1}(\lambda_i) \le P_{tot}$$

where

$$\mu_i^{-1}(\lambda_i) = \begin{cases} \text{The smallest } p \text{ such that } \mu_i(p)=\lambda \\ \infty \quad \text{if no such } p \text{ exists.} \end{cases}$$ ❑

In Fig. 5 we illustrate a general capacity region for N=2 channels with fixed channel states and concave power curves $\mu_1(p)$ and $\mu_2(p)$. In this case of fixed channel states, one might suspect the optimal solution to be the one which maximizes the instantaneous output rate at every instant of time: allocate full power to one queue whenever the other is empty, and allocate power to maximize the sum output rate $\mu_1(p_1)+\mu_2(p_2)$ subject to $p_1 + p_2 \le P_{tot}$ whenever both queues are full. Doing this restricts the capacity region to linear combinations of the three operating points, as illustrated in Fig. 5. The shaded regions in the figure represent the capacity gains obtained by power allocation. Note that the region is restricted further if only ON/OFF allocations are considered.

<u>Corollary 2</u>: For the K-server allocation problem where the channel rate of queue i is μ_i when it is allocated a server (and 0 otherwise), the capacity region is the polytope set of all $\vec{\lambda}$ vectors such that:

$$\sum_{i=1}^{N} \frac{\lambda_i}{\mu_i} \le K \qquad (5)$$

$$\lambda_i \in [0, \mu_i] \quad \text{for all } i. \qquad (6)$$

Proof: Using the virtual power curves and constraints given in Section II, we find that the region described by (5) and (6) contains the true capacity region. However, the K-server problem is constrained to allocate rates only on the vertex points of the capacity region (see Fig. 6).

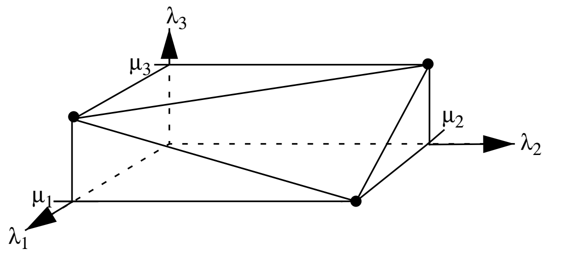

Figure 6: An example illustration of the capacity region for the *K*-server allocation problem with *N*=3, *K*=2.

Timesharing amongst vertex points, however, yields any desired point within the polytope. ❑

IV. A Stabilizing Power Allocation Algorithm

Here we assume that channel state vectors $\vec{C}$ vary *iid* from timeslot to timeslot with probability distribution $\pi_{\vec{C}}$. Likewise, assume that packets bring a new batch of unfinished work *iid* from timeslot to timeslot in the form of an arrival vector $\vec{A} = (a_1, \ldots, a_N)$, with distribution $f(a_1,...,a_N)$ and expectation $E[\vec{A}] = \vec{\lambda}T$. Note that entries of the channel state vector and the arrival vector may be correlated within the same timeslot. We assume that new arrivals have bounded second moments: $E[a_i^2] < \infty$.

Let $\vec{U}(t) = (U_1(t), \ldots, U_N(t))$ represent the vector of unfinished work in each queue at time t (where $t=kT$). We assume channel and queue state vectors $\vec{C}(t)$ and $\vec{U}(t)$ are known at the beginning of each timeslot, and seek a control policy which allocates power based on this information. Assuming this power allocation $\vec{P}(t)$ is held constant during the full timeslot $[t, t+T]$, the unfinished work dynamics proceed according to the one-step equation:

$$U_i(t+T) = max(U(t) + a_i(t) - \mu_i(p_i(t), c_i(t))T, 0). \quad (7)$$

Notice that for a given stationary power allocation policy, the unfinished work vector at timestep $t+T$ is independent of the past given the current value of unfinished work. Hence, the system can be viewed as evolving according to a Markov chain on an N-dimensional, uncountably infinite state space $\vec{U}$. For stability analysis, we define the Lyapunov function $L(\vec{U}) = \sum \theta_i U_i^2$ (for arbitrary positive weights $\{\theta_i\}$) and make use of a well developed theory of stability in Markov chains using negative Lyapunov drift (see [13], [2], [6], [9]). Below we state a sufficient condition for the system to be stable and have a well defined steady state distribution of unfinished work $\vec{U}$. The statement below is new in that it involves a Markov chain on an uncountably infinite state space, although the proof differs only cosmetically to that given in [15] and is omitted for brevity.

Theorem 2: For the given Lyapunov function $L(\vec{U})$, if there exists a compact region Ω of $\Re^N$ and a number $\alpha>0$ such that:

1. $E[L(\vec{U}(t+T)) | \vec{U}(t)] < \infty$ for all $\vec{U} \in \Re^N$
2. $E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \le -\alpha$ whenever $\vec{U} \notin \Omega$

then a steady state distribution on the vector $\vec{U}$ exists (clearly with the property that $Pr[U_i > u] \to 0$ as $u \to \infty$ for all i) and hence the system is stable.[2] ❑

Now consider the following power allocation policy: At the beginning of each timeslot, observe $\vec{U}(t)$ and $\vec{C}(t)$ and allocate a power vector $\vec{P}(t) = (p_1(t), \ldots, p_N(t))$ such that:

$$\vec{P}(t) = \underset{\sum p_i \le P_{tot}}{arg\ max} \sum \theta_i U_i(t) \mu_i(p_i, c_i) \quad (8)$$

where $\{\theta_i\}$ is any set of positive weights. This policy chooses $\vec{P}$ to maximize $\sum \theta_i U_i \mu_i$ subject to $\sum p_i \le P_{tot}$. Notice that the policy acts only through the current value of $\vec{U}(t)$ and $\vec{C}(t)$ without specific knowledge of the arrival rate vector $\vec{\lambda}$ or the channel state probabilities. Intuitively, we desire a policy that gives more power to queues with currently high data rates (to achieve maximum throughput) as well as gives more power to queues with large backlog (to ensure these queues are stabilized). The above policy does both by considering as a metric the product of backlog and data rate for each queue.

Theorem 3: The power allocation policy of choosing the power vector $\vec{P}(t) = \underset{\sum p_i \le P_{tot}}{arg\ max} \sum \theta_i U_i(t) \mu_i(p_i, c_i)$ stabilizes the system whenever the arrival rate vector $\vec{\lambda}$ is interior to the capacity region given in (4).

Proof: Consider the one-step drift in the Lyapunov function from Theorem 2. For ease of notation, let $U_i = U_i(t)$, $a_i = a_i(t)$, and let $\mu_i = \mu_i(p_i(t), c_i(t))$. From (7), we have:

$$U_i^2(t+T) \le (U_i + a_i - \mu_i T)^2$$
$$\le U_i^2 - 2TU_i\left(\mu_i - \frac{a_i}{T}\right) + \mu_i^2 T^2 + a_i^2. \quad (9)$$

[2] In [10,11] it is shown that if stronger conditions on the Lyapunov function are satisfied (such that the negative drift gets larger in magnitude as $|\vec{U}|$ increases) the moments of unfinished work are finite and can be bounded.

From (9) it is clear that property 1 of Theorem 2 holds. Now define the following constants:

$$\beta = \max_{\vec{C},\ \sum_i p_i = P_{tot}} \left[T^2 \sum \theta_i \mu_i^2(p_i, c_i)\right] \qquad (10)$$

$$B = \beta + \sum \theta_i E[a_i^2]\,. \qquad (11)$$

Taking conditional expectations of (9), scaling by weights θ_i and summing over all i, we have:

$$E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \le$$

$$B - 2T\sum_{i=1}^{N} \theta_i U_i (E[\mu_i | \vec{U}(t)] - \lambda_i) \quad (12)$$

where the *iid* nature of the packet arrivals has been used in the identity $E\left[\frac{a_i}{T}\middle|\vec{U}(t)\right] = \lambda_i$. Now notice that the term $\sum \theta_i U_i E[\mu_i | \vec{U}(t)]$ maximizes the value of $\sum \theta_i U_i \gamma_i$ over all vectors $\vec{\gamma} = (\gamma_1, \ldots, \gamma_N)$ in the capacity region (4). To see this, note that for any $\vec{\gamma}$ in the capacity region, there is a set of $\{p_i^{\vec{C}}\}$ values such that:

$$\sum_{i=1}^{N} \theta_i U_i \gamma_i \le \sum_{i=1}^{N} \theta_i U_i \sum_{\vec{C}} \pi_{\vec{C}} \mu_i(p_i^{\vec{C}}, c_i) \qquad (13)$$

$$= \sum_{\vec{C}} \pi_{\vec{C}} \sum_{i=1}^{N} \theta_i U_i \mu_i(p_i^{\vec{C}}, c_i) \qquad (14)$$

$$\le \sum_{\vec{C}} \pi_{\vec{C}} \max_{\sum p_i \le P_{tot}} \left[\sum_{i=1}^{N} \theta_i U_i \mu_i(p_i, c_i)\right] \qquad (15)$$

$$= \sum_{i=1}^{N} \theta_i U_i E[\mu_i | \vec{U}(t)]\,. \qquad (16)$$

Now, because the arrival rate vector $\vec{\lambda}$ is assumed to be strictly in the interior of the capacity region, we can add a positive vector $\vec{\varepsilon} = (\varepsilon, \ldots, \varepsilon)$ to produce another vector $(\vec{\lambda} + \vec{\varepsilon})$ which is in the capacity region. Hence, $\sum \theta_i U_i E[\mu_i | \vec{U}(t)] \ge \sum U_i(\lambda_i + \varepsilon)$, and we have:

$$\sum \theta_i U_i (E[\mu_i | \vec{U}] - \lambda_i) = \sum \theta_i U_i (E[\mu_i | \vec{U}] - (\lambda_i + \varepsilon) + \varepsilon)$$

$$\ge \varepsilon \sum \theta_i U_i\,. \qquad (17)$$

Using (17) in (12), we find that

$$E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \le B - 2T\varepsilon \sum_{i=1}^{N} \theta_i U_i\,. \quad (18)$$

Choose any number $\alpha > 0$ and define the compact region: $\Omega = \left\{\vec{U} \in \Re^N \middle| \sum \theta_i U_i \le \left(\frac{B+\alpha}{2T\varepsilon}\right)\right\}$.

We find from (18) that the Lyapunov drift is less than $-\alpha$ whenever $\vec{U} \notin \Omega$. ❑

Using the results of [10], it can be shown from the strong negative drift condition in (18) that the steady state unfinished work in all queues has a finite first moment and satisfies $\sum \theta_i \overline{U}_i \le B/(2T\varepsilon)$.

Note that the positive weights $\{\theta_i\}$ can be chosen arbitrarily. Choosing weights θ_i=1 for all i yields a policy which chooses a power vector that maximizes $\sum U_i \mu_i$ at every timestep. The following corollary makes use of a different set of weights.

Consider again the *K*-server allocation problem where each queue has only 2 channel states, ON or OFF, and these states vary *iid* over each timeslot as an *N*-dimensional vector. When a server is allocated to queue i while it is in the ON state, the server transmits data from the queue at a rate μ_i (the transmission rate is zero when in the OFF state or when no server is allocated). Defining the virtual rate-power curves as in Section II, we have the following corollary:

<u>*Corollary*</u>*:* For the *K*-server allocation problem with ON/OFF channel states, the policy of allocating the *K* servers to the *K* longest ON queues stabilizes the system whenever the system is stabilizable.

Proof: Assume the system operates according to virtual power curves as in Section II (eq. (1)), and define the Lyapunov function $L(\vec{U}) = \sum (U_i^2)/\mu_i$. With this Lyapunov function, we know that allocating power to maximize $\sum ((U_i(t))/\mu_i)\mu_i(p_i, c_i)$ (where $c_i \in \{ON, OFF\}$) stabilizes the system. Clearly the optimization needs not place any power on queues in the OFF state, so the summation can be restricted to queues that are ON:

$$\text{Maximize} \sum_{i | c_i = ON} U_i(t) \frac{\mu_i(p_i, c_i)}{\mu_i} \text{ subject to } \sum p_i \le K\,. \quad (19)$$

However, notice that the above maximization effectively chooses a rate vector $\vec{\mu}$ within the polytope capacity region specified in (5) and (6). The optimal solution for maximizing a linear function over a polytope will always be a vertex point. Fortunately, such a vertex point corresponds to the feasible allocation of *K* servers (with full power p_i=1) to *K* queues. Considering (19), the optimal way to do this is to choose the *K* queues with the largest value of $U_i(t)$. ❑

Using the same reasoning as in the proof above, it follows that the power allocation policy (8) naturally chooses a vertex point for any set of piecewise linear power curves, such as the virtual curves described in Section II. It follows that optimization can be restricted to searches over the vertex points without loss of optimality.

The above theorem uses the *iid* assumptions on packet arrivals and channel states to establish the negative drift condition for the Lyapunov function. We conjecture that the same policy stabilizes the system for general Markovian arrival and channel processes whenever the arrival rate vector $\vec{\lambda}$ is in the capacity region.

V. Joint Routing and Power Allocation

Consider now the following joint routing and power allocation problem: A stream of packets enters the satellite and the goal is to simply transmit all data to the ground as soon as possible, without regard to the specific ground location. Such a problem arises when the ground units are connected together via a reliable ground network, and the wireless paths from satellite to ground form the rate bottleneck (see Fig. 7).

As expected, treating all input streams $X_1(t), X_2(t),..., X_N(t)$ as an aggregate stream $X(t)$ and exploiting the routing options considerably expands the capacity region of the system. This capacity gain is achieved by utilizing the extra bandwidth offered by the ground network.

Specifically, let the input stream $X(t)$ (with rate λ) be composed of *iid* packet arrivals every timeslot T. Channel states vary according to an *iid* state vector $\vec{C}(t)$ as before. Every timeslot, we choose a power allocation $\vec{P}(t)$. Additionally, for every packet that enters the system, we make a routing decision and route the packet to one of the N queues. We assume that all queues have segregated buffers and routing decisions must be made immediately upon packet arrival.

In general, both the queue and channel state vectors $\vec{U}(t)$ and $\vec{C}(t)$ are important in both the routing and power allocation decisions. For example, clearly any power allocated to an empty queue is wasted and should be re-allocated to improve processing rates amongst the non-empty queues. Likewise, a router is inclined to place packets in faster queues, especially if the rates of those queues are guaranteed to operate at high levels for one or more timeslots. However, below we show that the routing and power allocation problem can be *decoupled* into two policies: a routing policy which considers only $\vec{U}(t)$, and a power allocation policy which considers only $\vec{C}(t)$. The resulting strategy maximizes total system throughput.

<u>*Theorem 4*</u>*:* The satellite downlink with joint routing and power allocation can stably support any arrival rate λ

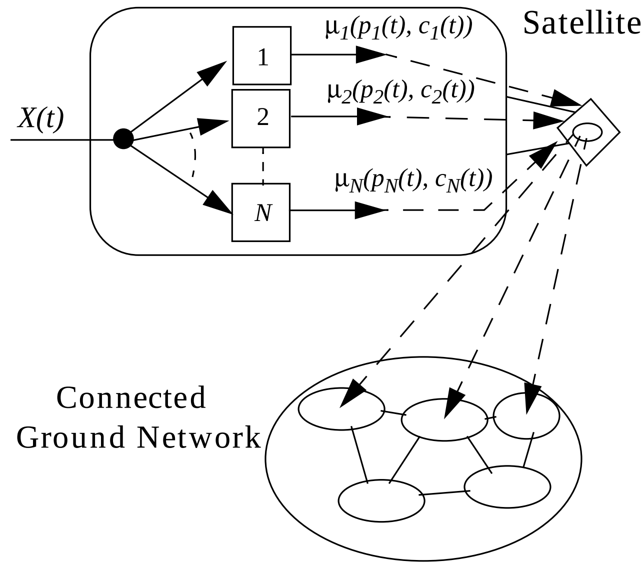

Figure 7: A joint routing/power allocation problem where the goal is to transmit the data to any node of the reliable ground network.

such that $\lambda < \bar{\mu}$, where we define the constant $\bar{\mu}$:

$$\bar{\mu} \triangleq \sum_{\vec{C}} \pi_{\vec{C}} \max_{\sum p_i \le P_{tot}} \left[\sum_{i=1}^{N} \mu_i(p_i, c_i) \right].$$

Furthermore, any arrival rate $\lambda > \bar{\mu}$ creates an unstable system under any routing/power allocation policy.

We prove the theorem by developing a throughput maximizing strategy. The strategy is decoupled into a routing policy and a power allocation policy:

Power Allocation: For each timeslot, allocate a power vector $\vec{P}(t) = (p_1(t), \ldots, p_N(t))$ that maximizes the sum output rate $\sum \mu_i(p_i(t), c_i(t))$ subject to $\sum p_i(t) \le P_{tot}$.

Routing: Route every packet that arrives in a given timeslot to the queue i with the least unfinished work $U_i(t)$ at the beginning of the timeslot.

Proof of Instability when $\lambda > \bar{\mu}$*:* Notice that the power allocation maximizes the sum rate $\mu(t) = \mu_1(t) + \ldots + \mu_N(t)$ at every instant of time. We make the simple sample path observation that the unfinished work in a single-server queue with input stream $X(t)$ and time varying processing rate $\mu(t)$ is always less than or equal to the total unfinished work in a system of N parallel queues with transmission rates $\mu_1(t),..., \mu_N(t)$ such that $\sum \mu_i(t) \le \mu(t)$ for all t (see [16]). From Lemma 1, we know that a single queue system with $\lambda > \underline{\mu}$ is unstable. Hence, the multi-queue system will also be unstable under any power allocation and routing policy whenever $\lambda > \bar{\mu}$.

Proof of stability when $\lambda < \bar{\mu}$ *:* Again define the Lyapunov function $L(\vec{U}) = \sum U_i^2$. Let *A(t)* represent the total amount of bits from packets arriving in timeslot [t, $t+T$], and let ($a_1(t),..., a_N(t)$) represent the bit length of packets routed to queues $i \in \{1, \ldots, N\}$ (where $A(t) = \sum a_i(t)$, and $E[A] = \lambda T$). Let μ_i represent the transmission rate $\mu_i(p_i(t), c_i(t))$ of queue i during timeslot [t, $t+T$] under the specified power allocation policy. As in the proof of Theorem 4, we have:

$$U_i^2(t+T) \le U_i^2 - 2TU_i\left(\mu_i - \frac{a_i}{T}\right) + \mu_i^2 T^2 + a_i^2 \quad \text{for all } i. \tag{20}$$

For a fixed power allocation policy, the a_i values are the only variables dependent upon our routing decisions. Define the constant:

$$C = T^2 \sum E[\mu_i^2] + E[A^2]. \tag{21}$$

Summing (20) over all $i \in \{1, \ldots, N\}$, taking conditional expectations, and noting that $A^2 \ge \sum a_i^2$ yields:

$$E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \le C - 2T\sum_{i=1}^{N} U_i\left(\bar{\mu}_i - E\left[\frac{a_i}{T}\middle|\vec{U}(t)\right]\right) \tag{22}$$

where $\bar{\mu}_i = E[\mu_i]$ is the expected processing rate of server i under the given power allocation policy and channel state probabilities. Now notice that the strategy of routing all bits $\sum a_i$ to the queue i with the smallest value of U_i leads to a term $\sum U_i E\left[\frac{a_i}{T}\middle|\vec{U}(t)\right]$ in the above inequality which minimizes the function $\Phi(\vec{\gamma}) = \sum U_i \gamma_i$ for all positive vectors $\vec{\gamma}$ subject to $\sum \gamma_i \ge \lambda$. If $\lambda < \bar{\mu}$, then there exists an $\varepsilon > 0$ such that $\sum(\bar{\mu}_i - \varepsilon) \ge \lambda$. Hence, adding and subtracting the ε values in the summation of (22), we find:

$$E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \le C - 2T\varepsilon \sum_{i=1}^{N} U_i$$

Defining any α>0 and choosing the compact set Ω to be:

$$\Omega = \left\{\vec{U} \in \Re^N \middle| \sum U_i \le \left(\frac{C+\alpha}{2T\varepsilon}\right)\right\}$$

ensures the negative drift condition of Theorem 3 whenever $\vec{U} \in \Omega$. ❑

This separate buffer scenario is useful in cases when it is economical to segregate buffers among queues, or when the parallel queues are physically aboard different satellites within a space constellation. In cases where a shared buffer can be used and packets can be routed immediately when the next server becomes empty, a stronger stability result can be obtained: Allocating power as before and using the shared buffer to employ this work conserving routing strategy ensures that the unfinished work in the system is no more than $(N\text{-}1)L_{max}$ bits in excess of any other routing and power allocation strategy, where L_{max} is the bit length of the maximum size packet (see [16]).

A variation of this joint routing and power allocation scenario restricts the routing options for each data stream. Traffic intended for a certain ground location can be routed to a subset of neighboring locations, but cannot be routed to ground nodes outside of this subset (Fig. 8). Inputs are divided into M traffic classes $X_1,...,X_M$ corresponding to M disjoint queue clusters $Q_1,...,Q_M$. Packets from stream X_j can be routed to any queue in cluster Q_j.

Such a problem can be treated using the analysis and of this section as well as Section IV. Indeed, maximum throughput can be achieved using the algorithms from these sections in a hierarchical manner. First, each queue cluster Q_j is treated as a single virtual queue with a rate-power curve defined as:

$$\tilde{\mu}_j(p, \vec{C}_j) = \max_{\sum_{i \in Q_j} p_i \le p} \left[\sum_{i \in Q_j} \mu_i(p_i, c_i)\right]$$

where $\vec{C}_j$ represents the vector of channel states for all queues in cluster Q_j. It can be shown that the functions $\tilde{\mu}_j(p, \vec{C}_j)$ are concave in the power variable p. The first level of control decisions uses the power allocation policy of Section IV to allocate power ($p_1,...,p_M$) (subject to $\sum p_j \le P_{tot}$) to the M queue clusters to maximize $\sum_j \tilde{\mu}(p_j, \vec{C}_j)\tilde{U}_j(t)$, where $\tilde{U}_j(t)$ represents the total unfinished work in queues from cluster Q_j. For the second level of the hierarchy, each queue cluster carries out the joint power and routing algorithm specified in this section, using the Join-the-Shortest-Queue strategy for

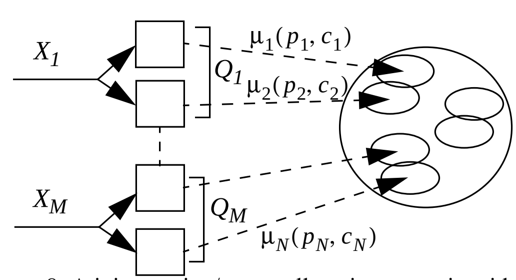

Figure 8: A joint routing/power allocation scenario with M different routing clusters $\{Q_j\}$.

each traffic stream X_j and queue cluster Q_j. Such a hierarchical strategy stabilizes the system whenever it is stabilizable.

VI. Connectivity Constraints

It has been assumed throughout that all transmit channels can be activated simultaneously, subject only to the total power constraint $\sum p_i(t) \le P_{tot}$ for all time t. Hence, it is implicitly assumed that there is no interchannel interference. Such an assumption is valid when there is sufficient bandwidth to ensure potentially interfering channels can transmit using different frequency bands. However, in bandwidth limited scenarios, power allocation vectors $\vec{P}(t)$ may be additionally restricted to channel *activation sets*: finite sets $P_1,...,P_R$, where each set P_i is a convex set of points $(p_1,...,p_N)$ representing power vectors which, when allocated, ensure interchannel interference is at an acceptable level. This use of activation sets is similar to the treatment in [2], where activation link sets for scheduling ON/OFF links in a wireless network are considered. Here, the definition has been extended from sets of links to sets of power vectors to treat power control.

As an example of an activation set, consider the system of Fig. 1 and suppose that downlink channels 1, 2, and 3 can be activated simultaneously if all other transmitters are silent. Such an activation set can be represented:

$$P_j = \left\{(p_1, p_2, p_3, 0, \ldots, 0) \in \Re^N \middle| p_i \ge 0, \sum_{i=1}^{3} p_i \le P_{tot}\right\}$$

Another type of system constraint is when power allocation is further restricted so that no more than K transmitters are active at any given time. Such a constraint corresponds to $\binom{N}{K}$ convex activation sets.

In the following, we assume that each activation set incorporates the power constraint $\sum p_i \le P_{tot}$. Consider the downlink system of Fig. 1. Packets arrive according to a random arrival vector (*iid* on each timeslot) with rates $(\lambda_1, \ldots, \lambda_N)$, and channel states $\vec{C}(t)$ vary *iid* every timeslot with steady state probabilities $\pi_{\vec{C}}$. Each timeslot a power allocation vector $\vec{P}(t)$ is chosen such that it lies within one of the activation sets $P=\{P_1,...,P_R\}$.

Theorem 5: (a) The capacity region of the system is the set of all arrival rate vectors $\vec{\lambda}$ such that:

$$\vec{\lambda} \in \Omega \triangleq \sum_{\vec{C}} \pi_{\vec{C}}\ \textit{Convex Hull}\left(\{\vec{\mu}(\vec{P}, \vec{C}) | (\vec{P} \in P_j)\}\Big|_{j=1}^{R}\right)$$

where addition and scalar multiplication of sets has been used above.[3]

[3] For sets A, B and scalars α, β, the set $\alpha A+\beta B$ is defined as $\{\gamma \mid \gamma = \alpha a+\beta b$ for some $a \in A$, $b \in B\}$.

(b) The policy of allocating a power vector $\vec{P}=(p_1,...,p_N)$ at each timestep to maximize the quantity $\sum_{i=1}^{N} U_i(t)\mu_i(p_i, c_i(t))$ (subject to $\vec{P} \in P = \{P_1, \ldots, P_R\}$) stabilizes the system whenever the $\vec{\lambda}$ vector is in the interior of the capacity region.

We note that the allocation policy specified in part (b) of the theorem involves the non-convex constraint $\vec{P} \in P$. Optimizing the given metric over individual activation sets P_j is a convex optimization problem, although a complete implementation of the given policy is non-trivial if the number of activation sets is large.

However, the proof of parts (a) and (b) are simple extensions of the analysis presented in Sections III and IV.

Proof of (a): To establish that $\vec{\lambda} \in \Omega$ is a necessary condition for stability, suppose the system is stable using some power allocation function $\vec{P}(t)$ which satisfies $\vec{P}(t) \in P$ for all time. We thus know that $\lambda_i \le \underline{\mu}_i$ for all i (Lemma 1), and the proof proceeds as the proof of Theorem 1 (Appendix A2), where for any fixed $\varepsilon > 0$ we can find a large time $\tilde{t}$ such that the following entrywise vector inequality is satisfied:

$$\vec{\lambda} \le \underline{\vec{\mu}} \le \frac{1}{\tilde{t}}\int_0^{\tilde{t}} \vec{\mu}(\vec{P}(\tau), \vec{C}(\tau))d\tau + \vec{\varepsilon}$$

The main difference from Theorem 1 is that the above integral is broken into a double summation over intervals when the channel is in state $\vec{C}$ and when the power vector is in set P_j. Let $T_{\vec{C}}(\tilde{t})$ represent the intervals of time during $[0, \tilde{t}]$ when the channel is in state $\vec{C}$, and let $T_{\vec{C}, Pj}(\tilde{t})$ represent the subintervals of $T_{\vec{C}}(\tilde{t})$ when the power function $\vec{P}(t)$ is in activation set P_j. We have:

$$\vec{\lambda} \le \underline{\vec{\mu}} \le \sum_{\vec{C}} \sum_{Pj} \frac{\|T_{\vec{C}, Pj}(\tilde{t})\|}{\tilde{t}} \frac{1}{\|T_{\vec{C}, Pj}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}, Pj}(\tilde{t})} \vec{\mu}(\vec{P}(\tau), \vec{C})d\tau + \vec{\varepsilon}$$

$$\le \sum_{\vec{C}} \frac{\|T_{\vec{C}}(\tilde{t})\|}{\tilde{t}} \sum_{Pj} \frac{\|T_{\vec{C}, Pj}(\tilde{t})\|}{\|T_{\vec{C}}(\tilde{t})\|} \vec{\mu}\left(\frac{1}{\|T_{\vec{C}, Pj}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}, Pj}(\tilde{t})} \vec{P}(\tau)d\tau, \vec{C}\right) + \vec{\varepsilon}$$

$$\le \sum_{\vec{C}} \pi_{\vec{C}} \left[\sum_{Pj} \frac{\|T_{\vec{C}, Pj}(\tilde{t})\|}{\|T_{\vec{C}}(\tilde{t})\|} \vec{\mu}(\vec{P}_{\vec{C}, Pj}, \vec{C})\right] + O(\varepsilon)$$

where $\vec{P}_{\vec{C}, Pj} \triangleq \frac{1}{\|T_{\vec{C}, Pj}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}, Pj}(\tilde{t})} \vec{P}(\tau)d\tau$.

Note that any point $\vec{\gamma}$ in the convex hull of a collection of convex sets can be written as a linear combination of points $\vec{\gamma}_1,...,\vec{\gamma}_N$ in the sets: $\vec{\gamma} = \alpha_1\vec{\gamma}_1 + \ldots + \alpha_N\vec{\gamma}_N$ where $\alpha_j \geq 0$ and $\sum \alpha_j \leq 1$. Letting $\alpha_j = \|T_{\vec{C}, Pj}(\tilde{t})\| / \|T_{\vec{C}}(\tilde{t})\|$, we see the inequality above indicates that $\vec{\lambda}$ is arbitrarily close to a point in the capacity region Ω, and hence $\vec{\lambda} \in \Omega$.

The sufficiency condition is implied by part (b). ❑

Proof of (b): Define the Lyapunov function $L(\vec{U}) = \sum U_i^2$. The proof of Theorem 3 can literally be repeated up to eq. (12):

$$E[L(\vec{U}(t+T)) - L(\vec{U}(t)) | \vec{U}(t)] \leq B - 2T\sum_{i=1}^{N}\theta_i U_i(E[\mu_i | \vec{U}(t)] - \lambda_i)$$

From this point, negative drift of the Lyapunov function can be established by again noting that the value of $E[\mu_i | \vec{U}(t)]$ maximizes $\sum U_i\gamma_i$ over all vectors $\vec{\gamma}$ within the capacity region. To see this, note that any $\vec{\gamma}$ in the capacity region can be written:

$$\vec{\gamma} = \sum_{\vec{C}}\pi_{\vec{C}}\sum_{Pj}\alpha_{\vec{C}, Pj}\vec{\mu}(\vec{P}_{\vec{C}, Pj}, \vec{C})$$

for some vectors $\vec{P}_{\vec{C}, Pj} \in P_j$, and some scalar values $\alpha_{\vec{C}, Pj} \geq 0$ such that $\sum_{P_j}\alpha_{\vec{C}, Pj} \leq 1$ for all channel states $\vec{C}$. The result follows by an argument similar to (13)-(16). ❑

VII. Conclusions

We have treated data transmission over multiple time-varying channels in a satellite downlink using power control. Processing rates for each channel *i* were assumed to be determined by concave rate-power curves $\mu_i(p_i,c_i)$, and the capacity region of all stabilizable arrival rate vectors $\vec{\lambda}$ was established. This capacity region is valid for general Markovian input streams, and inputs with arrival rates $\vec{\lambda}$ in the interior of the capacity region can be stabilized with a power allocation policy which only considers the current channel state $\vec{C}(t)$. In the case when arrival rates and channel probabilities $\vec{\lambda}$ and $\pi_{\vec{C}}$ are unknown but packet arrivals and channel state transitions are *iid* every timeslot, a stabilizing policy which considers both current state and current queue backlog was developed. Intuitively, the policy favors queues with large backlogs and better channels by allocating power to maximize $\sum U_i\mu_i$ at every timeslot.

The power control formulation was shown to contain the special case of a server allocation problem, and analysis verified stability of the "Serve-the-*K*-Largest-Connected-Queue" policy. In the case of interchannel interference, modified power allocation policies were developed when power vectors are constrained to a finite collection of activation sets. A stabilizing policy was developed, although the policy is difficult to implement if the number of activation sets is large.

A joint routing and power allocation scenario was also considered, and a throughput maximizing algorithm was developed. Stability properties of these systems were established by demonstrating negative drift of a Lyapunov function defined over the current state of unfinished work in the queues. The *iid* assumptions for packet arrivals and channel transitions were needed to establish the negative drift condition. We conjecture that the same policies also stabilize the system for general Markovian arrival and channel processes whenever $\vec{\lambda}$ is in the capacity region.

Our focus was power control for a satellite downlink, although the results extend to other wireless communication scenarios where power allocation and energy efficiency is a major issue. The use of power control can considerably extend the throughput and performance properties of such systems.

Appendix A:

A1. Lemma 1b: If an input stream *X(t)* to a single queue system is rate-ergodic of input rate λ, a necessary condition for queue stability is $\lambda \leq \underline{\mu}$.

Proof: Suppose $\lambda > \underline{\mu}$ and choose $\varepsilon > 0$ such that $\lambda - \underline{\mu} - 2\varepsilon > 0$. The limits in (2) ensure that, with probability 1, we can find a set of times $\{t_i\}$ ($i \in \{1, 2, \ldots\}$) where $t_i \rightarrow \infty$ with increasing *i*, and such that for all t_i:

$$\frac{X(t_i)}{t_i} \geq \lambda - \varepsilon \quad , \quad \frac{1}{t_i}\int_0^{t_i}\mu(\tau)d\tau \leq \underline{\mu} + \varepsilon \qquad (23)$$

However, it is clear that

$$U(t_i) \geq X(t_i) - \int_0^{t_i}\mu(\tau)d\tau \qquad (24)$$

From (23) and (24), it follows that $U(t_i) \geq (\lambda - \underline{\mu} - 2\varepsilon)t_i$ for all t_i. Define $\alpha = \lambda - \underline{\mu} - 2\varepsilon$, and let T_i represent the extra time it takes the unfinished work in the queue to empty below a threshold value *M*, starting at value $U(t_i)$ at time t_i. Clearly $T_i \geq (\alpha t_i - M)/\mu_{max}$, and hence at any time t_i+T_i, the empirical fraction of time the unfinished work in the queue exceeded the value *M* is greater than or equal to $T_i/(t_i+T_i)$, which is greater than or equal to $(\alpha t_i - M)/(\mu_{max}t_i + \alpha t_i - M)$. Taking limits as $t_i \to \infty$ reveals that $g(M) \geq \alpha/(\alpha + \mu_{max})$ for all M, and hence the system is unstable. ❑

A2. Theorem 1b: A necessary condition for stability of the downlink channel of Fig. 1 is $\lambda_i \leq \sum_{\vec{C}} \pi_{\vec{C}} \mu_i(p_i^{\vec{C}}, c_i)$ for all i.

Proof: Suppose all queues of the downlink channel can be stabilized with some power control function $\vec{P}(t)$ which meets the power constraints--perhaps a function derived from a policy which knows future events. From the necessary condition of Lemma 1 we know that the *lim inf* of the resulting rate process satisfies $\lambda_i \leq \underline{\mu}_i$ for all queues $i \in \{1, \ldots, N\}$.

We upper bound $\underline{\mu}_i$ as follows. Let $T_{\vec{C}}(t)$ represent the subintervals of [0, t] during which the channel is in state $\vec{C}$, and let $\|T_{\vec{C}}(t)\|$ denote the total length of these subintervals. Fix ε>0, and let $|\vec{C}|$ represent the total number of channel states of the system. Because the channel process is ergodic, and because there are a finite number of queues and channel states, there exists a time $\tilde{t}$ such that the time average fraction of time in each channel state and the time average processing rate of all queues are simultaneously within ε of their limiting values:

$$\frac{\|T_{\vec{C}}(\tilde{t})\|}{\tilde{t}} \leq \pi_{\vec{C}} + \varepsilon \quad \text{for all channel states } \vec{C} \tag{25}$$

$$\underline{\mu}_i \leq \frac{1}{\tilde{t}} \int_0^{\tilde{t}} \mu_i(p_i(\tau), c_i(\tau)) d\tau + \varepsilon \quad \text{for all } i \in \{1, \ldots, N\}. \tag{26}$$

Thus, under power decisions $\vec{P}(t)$, we have for all i:

$$\lambda_i \leq \underline{\mu}_i \leq \sum_{\vec{C}} \frac{\|T_{\vec{C}}(\tilde{t})\|}{\tilde{t}} \frac{1}{\|T_{\vec{C}}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}}(\tilde{t})} \mu_i(p_i(\tau), c_i) d\tau + \varepsilon \tag{27}$$

$$\leq \sum_{\vec{C}} \frac{\|T_{\vec{C}}(\tilde{t})\|}{\tilde{t}} \mu_i\left(\frac{1}{\|T_{\vec{C}}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}}(\tilde{t})} p_i(\tau) d\tau, c_i \right) + \varepsilon \tag{28}$$

$$\leq \sum_{\vec{C}} (\pi_{\vec{C}} + \varepsilon) \mu_i\left(\frac{1}{\|T_{\vec{C}}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}}(\tilde{t})} p_i(\tau) d\tau, c_i \right) + \varepsilon \tag{29}$$

where (28) follows from concavity of the $\mu_i(p, c_i)$ functions with respect to the power variable p, and (29) follows from (25). We define for all states $\vec{C}$ and queues i:

$$\tilde{p}_i^{\vec{C}} = \frac{1}{\|T_{\vec{C}}(\tilde{t})\|} \int_{\tau \in T_{\vec{C}}(\tilde{t})} p_i(\tau) d\tau \tag{30}$$

Hence, from (29) and (30):

$$\lambda_i \leq \sum_{\vec{C}} \pi_{\vec{C}} \mu_i(\tilde{p}_i^{\vec{C}}, c_i) + \varepsilon(1 + |\vec{C}| \mu_{max}) \tag{31}$$

where μ_{max} is defined as the maximum processing rate of a queue (maximized over all queues and channel states) when it is allocated the full power P_{tot}.

Because the original power function satisfies the power constraint $\sum p_i(t) \leq P_{tot}$ for all times t, from (30) it is clear that the $\tilde{p}_i^{\vec{C}}$ values satisfy the constraint $\sum_i \tilde{p}_i^{\vec{C}} \leq P_{tot}$ for all channel states $\vec{C}$. Thus, (31) indicates that the arrival vector $\vec{\lambda}$ is arbitrarily close to a point in the region specified by (4). Because the region (4) is closed, it must contain $\vec{\lambda}$, and hence (4) represents the capacity region of the system. ❑

REFERENCES:

[1] L. Tassiulas and A. Ephremides, "Dynamic Server Allocation to Parallel Queues with Randomly Varying Connectivity," *IEEE Trans. on Information Theory*, vol.39, no.2, March 1993.

[2] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks," *IEEE Transactions on Automatic Control*, Vol. 37, no. 12, Dec. 1992.

[3] Bruce Hajek, "Optimal Control of Two Interacting Service Stations," *IEEE Trans. on Autom. Control*, vol.29,no.6, June 1984.

[4] I. Viniotis and A. Ephremides, "Extension of the Optimality of the Threshold Policy in Heterogeneous Multiserver Queueing Systems,"*IEEE Trans. on Autom. Contr.*, vol.33,no.1, Jan.1988.

[5] M. Carr and Bruce Hajek, "Scheduling with Asynchronous Service Opportunities with Applications to Multiple Satellite Systems," *IEEE Trans. on Autom. Control*, vol.38,no.12,Dec. 1993.

[6] G. Bongiovanni, D.T. Tang, C.K. Wong, "A general Multibeam satellite switching algorithms," *IEEE Trans. on Communications*, July, 1981.

[7] N. McKeown, A.Mekkittikul, "A Practical Scheduling Algorithm to achieve 100% Throughput in Input-Queued Switches," *IEEE INFOCOM Proceedings* 1998, pp. 792-799.

[8] N. McKeown, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar. 1996, pp. 296-302.

[9] M. Karol, M. Hluchyj, and S. Morgan, "Input versus output queueing on a space division switch," *IEEE Trans. Commun.*, vol. 35, pp. 1347-1356, 1987.

[10] E. Leonardi, M. Mellia, F. Neri, and M. Ajmone Marson, "Bounds on Average Delays and Queue Size Averages and Variances in Input-Queued Cell-Based Switches," *IEEE INFOCOM* Proceedings 2001, vol 2.

[11] P.R.Kumar, S.P.Meyn, "Stability of Queueing Networks and Scheduling Policies," IEEE Transactions on Automatic Control, vol.40,.n.2, Feb. 1995, pp.251-260.

[12] B. Prabhakar, Elif Uysal Biyikoglu, and A. El Gamal, "Energy-Efficient Transmission Over a Wireless Link via Lazy Packet Scheduling," *Proceedings of IEEE INFOCOM* 2001.

[13] S. Toumpis and A. Goldsmith, "Some Capacity Results for Ad Hoc Networks," *38th Annual Allerton Conf. Proceedings*, 2000.

[14] Ioannis Ch. Paschalidis, *Large Deviations in High Speed Communication Networks*. Ph.D. Thesis, MIT LIDS May 1996.

[15] Soren Asmussen, *Applied Probability and Queues.* New York: John Wiley, 1987.

[16] M.J.Neely, "Packet Routing over Parallel Time-Varying Queues with Application to Satellite and Wireless Networks," submitted to UIUC Allerton Conference 2001.

[17] M.J.Neely and E.Modiano, "Convexity and Optimal Load Distributions in Work Conserving */*/1 Queues," *Proceedings of IEEE INFOCOM,* 2001.