

Packet-Based Power Allocation for Forward Link Data Traffic

Peijuan Liu Randall Berry Michael L. Honig Scott Jordan

Abstract

We consider the allocation of power across forward-link packets in a wireless data network. The packets arrive according to a random (Poisson) process, and have fixed length so that the data rate for a given packet is determined by the assigned power and the channel gain to the designated user. Each user's service preferences are specified by a utility function that depends on the received data rate. The objective is to determine a power assignment policy that maximizes the time-averaged utility rate, subject to a constraint on the probability that the total power exceeds a limit (corresponding to an outage). For a large, heavily loaded network, we introduce a Gaussian approximation for the total transmitted power, which is used to decompose the power constraint into three more tractable constraints. We present a solution to the modified optimization problem that is a combination of admission control and pricing. The optimal trade-off between these approaches is characterized. Numerical examples illustrate the achievable utility rate and power allocation as a function of the packet arrival rate.

Submitted to *IEEE Transactions on Wireless Communications* September, 2005. This work was supported by the Motorola-Northwestern Center for Telecommunications, and by NSF under grant CCR-9903055. This work was presented in part at *Globecom 2002*, Taipei, Taiwan.

P. Liu was with the Dept. of Electrical & Computer Engineering, Northwestern University. She is now with Motorola Labs, Schaumburg, IL 60196; Tel: (847) 538-0699, e-mail: Peijuan.Liu@motorola.com.

R. Berry is with the Dept. of Electrical & Computer Engineering, Northwestern University, Evanston, IL 60208; Tel/Fax: (847) 491-7074/467-3550, e-mail: rberry@ece.northwestern.edu.

M. L. Honig is with the Dept. of Electrical & Computer Engineering, Northwestern University, Evanston, IL 60208; Tel/Fax: (847) 491-7803/467-3550, e-mail: mh@ece.northwestern.edu.

S. Jordan is with the ECE Department, University of California, Irvine, 544D Engg. Tower, Irvine, CA 92697-2625, e-mail: sjordan@uci.edu.

I. INTRODUCTION

Efficient allocation of radio resources, such as transmission power, is essential for supporting diverse applications over wireless networks. Here we investigate power allocation for the forward link in a wireless network with rate adaptive data traffic. We consider a code division multiple access (CDMA) system that simultaneously transmits to all active flows; the available transmission power must be allocated among these flows. A utility-based approach is adopted, in which the service preferences of each packet are specified by a utility function. The network objective is to maximize a time average utility. It is well-known that such utility functions can capture many common definitions of fairness within a network [1], [2] and can provide for different priorities among users.

Power control in cellular CDMA systems based on utility maximization has been studied for both the reverse link [3]–[8] and the forward link [9]–[15]. In the forward link, the typical problem is to maximize the aggregate utility subject to constraints on the total available resources. For example, [11] considers a constraint on the transmitted power, while [13] considers constraints on both the available spreading codes and power. The solution to these problems can often be interpreted in a pricing framework, where prices are announced for the constrained resources and users maximize their net benefit (utility minus cost). The optimal allocation of resources can be found by choosing the appropriate resource prices. In most of this work, a static situation is assumed, where the set of active users is fixed. In this paper, the set of active users is dynamically varying over the time period during which resources are allocated. Random traffic variations must therefore be taken into account when allocating resources.

We consider a model in which packets arrive to the base station according to a Poisson process. The packets are designated for different users with random channel gains, and the time to transmit a packet depends on the power allocation and the associated channel gain. Here a “packet” could also represent fixed length flow or session for a particular user as in [19]. An orthogonal signaling scheme is assumed, in which multiple packets are simultaneously transmitted to different users,

and the packets do not interfere with each other. Each transmitted packet contributes a utility to the designated user, which depends only on the transmission time (equivalently, the data rate). Our problem is to determine a policy for allocating power to each packet, which maximizes the time-average utility rate (i.e., total accumulated utility per unit time), subject to a constraint on the total power transmitted by the base station.

Since the number of active users is randomly varying, the total power transmitted by the base-station is a random process. We consider an outage constraint on this process, in which the total power can exceed a given value with some small probability.¹ We characterize the solution to this problem for a system with a large number of users, so that the transmitted power can be approximated as a Gaussian random process. In that case, the outage constraint can be decomposed into three simpler constraints. The solution to this simplified problem can be viewed in a pricing framework as in [13]; however, there are several fundamental differences. First, in addition to pricing, explicit admission control is also needed. Second, the price is not the conventional (linear) price for the constrained resource, which is the transmit power. Instead, the total charge is the price times the product of the transmission power times energy. This can be viewed as a price for power times energy, or as a non-linear price for the required power.

Our focus is on the situation where traffic variations occur on a much faster time-scale than that over which resource allocation is performed. Specifically, we assume power is allocated to each packet based on the user's channel gain and utility, and this assignment is fixed for the duration of the packet. The power allocation does not depend on the instantaneous system state (e.g., the number of active requests), but only on long-term statistics (e.g., packet arrival rates). An alternate approach may take into account the current system state and reallocate resources at every arrival and departure (e.g., see [16], [17]). Clearly, allocating resources on a faster time scale may improve the resulting utility rate. However, such an approach may not be feasible, due to various system constraints, and leads to a more complicated allocation policy. Also, since

¹Here we assume that power is the limiting resource, and that there is sufficient bandwidth to support non-interfering (orthogonal) transmissions to all active users.

the allocation considered here is not state dependent, each designated user derives a fixed utility rate upon admission. In contrast, with state dependent reallocations the utility associated with a packet can vary depending on future events.

We also assume that the channel varies on a slower time-scale than the traffic requirements. Specifically, the channel gain does not change during the time required to serve a packet. If this were not the case, the performance could be improved by utilizing an opportunistic scheduling algorithm, such as the proportional fair rule for the CDMA 1xEVDO system [18], [19]. We note that many opportunistic scheduling algorithms can also be viewed in terms of maximizing an aggregate utility rate [20]–[22].

The rest of the paper is organized as follows. In Section II, we introduce a model for the forward link of a single cell. In Section III, we formulate a constrained optimization problem where the objective is to maximize the time-averaged utility rate subject to a stochastic constraint on the total power. In Section IV, a solution to the simplified problem is presented in which the power constraint is decomposed into three more tractable constraints. We then characterize the optimized system behavior. Numerical results, which illustrate the accuracy of the Gaussian approximation for the power distribution, and optimized power allocations, are presented in Section V.

II. SYSTEM MODEL

We consider a model for the forward link within a single cell, where the base station transmits simultaneously to all active users, and transmissions to different users are assumed to be orthogonal. For example, this models a CDMA system with orthogonal spreading codes.² Suppose that a user with channel gain h is allocated transmission power $P(h)$. The received Signal-to-Interference Plus Noise Ratio (SINR) for this user is given by $SINR = hP/\sigma^2$, where σ^2 is the total noise plus interference power. We assume that the received data rate for a user is

²We assume that the number of orthogonal codes, or equivalently, the available bandwidth is not a limiting resource. A bandwidth constraint could be introduced in what follows by adding a constraint on the number of simultaneous transmissions.

a function of the received power, or equivalently received SINR; this relationship is given by $R(h) = \mathcal{C}(hP)$, where $\mathcal{C}(\cdot)$ is an increasing function.

Packets arrive to the base station according to a Poisson process with overall rate λ . Each packet has a fixed length of L (bits).³ We consider a system with a large number of users, and assume that each packet corresponds to a new user. The channel gain for each user is assumed to be distributed on the interval $\mathcal{H} = [h_{min}, h_{max}]$, where $h_{min} \geq 0$ and $h_{max} < \infty$, with continuous density function $f_H(h)$. This density can be used to model the users' geographic distribution within the cell, and also various propagation effects such as random shadowing. The channel gain corresponding to each arrival is chosen independently according to this distribution and stays fixed during the entire transmission of the packet.

A utility function is associated with each packet, which reflects the designated user's desired quality of service. We assume that the utility depends only on the transmission rate R . Since each packet has a fixed length, this is equivalent to defining utility as a function of the transmission time for a packet. In this paper, we assume that all users have the same utility function, $U(R)$; however, this formulation can be extended to scenarios with multiple utility classes. We assume that $U(0) = 0$ and that $U(R)$ is increasing, concave and continuously differentiable with respect to R , for $R \geq 0$. These are common assumptions for so-called "elastic" traffic, which describes many data applications [1]. An example utility function, $U(R)$, with these characteristics is depicted in Figure 1.

The power allocated to a user depends only on the utility function $U(\cdot)$ and the associated channel gain h . For each $h \in \mathcal{H}$, it will be useful to define the function $\tilde{U}_h(P)$, which relates the utility received by a user with channel gain h to the transmitted power P . This function is given by

$$\tilde{U}_h(P) = U(\mathcal{C}(hP)). \quad (1)$$

Notice that $\tilde{U}_h(P)$ is different for users with different channel gains even though $U(R)$ is the

³The following can be extended to the case where the length of each request is random, but we will not address this extension here.

same for those users.

III. PROBLEM FORMULATION

Our objective is to allocate transmission power to maximize the utility rate given a constraint on the total transmission power. A power allocation is specified by a function $P : \mathcal{H} \mapsto \mathbb{R}^+$ that indicates the power used to transmit a packet to a user with channel gain $h \in \mathcal{H}$. If $P(h) = 0$, the corresponding packet is considered blocked and not transmitted. If $P(h) > 0$, the corresponding packets are transmitted with a transmission time given by

$$T(h) = \frac{L}{\mathcal{C}(hP(h))}.$$

Let $\{H_i\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed random variables representing the channel gain of the i th arrival, and let $K(t)$ denote the number of arrivals in the interval $[0, t)$. For a given power allocation, the time average utility rate is given by

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^{K(t)} \tilde{U}_{H_i}(P(H_i)) = \lim_{t \rightarrow \infty} \frac{K(t)}{t} \left(\frac{1}{K(t)} \sum_{i=1}^{K(t)} \tilde{U}_{H_i}(P(H_i)) \right) \quad (2)$$

$$= \lambda \mathbb{E}_H \left\{ \tilde{U}_H(P(H)) \right\}, \quad (3)$$

assuming the system is ergodic, where the expectation is an average over f_H .

Let $\mathcal{A}(t)$ denote the set of active transmissions at time t . The cardinality of $\mathcal{A}(t)$ is $N(t)$, which is the number of the current active packet transmissions. The total power transmitted at time t can then be written as

$$P_{sum}(t) = \sum_{i \in \mathcal{A}(t)} P(H_i). \quad (4)$$

This is a stochastic process with statistics that depend on the power allocation and the channel distribution. We assume that under any power allocation, $P_{sum}(t) \rightarrow P_{sum}$ in distribution as $t \rightarrow \infty$, where P_{sum} is a random variable with the steady-state distribution. For any power allocation, we constrain the steady-state total power, given that the system is not empty ($N(t) > 0$), to be no greater than some value, \bar{P} , with probability $1 - q_0$, i.e., $\Pr(P_{sum} > \bar{P} | P_{sum} > 0) \leq q_0$ where $q_0 > 0$ is a small constant. Assuming that the system is ergodic, this constraint implies that the

fraction of time the total power is greater than \bar{P} , when the system is not empty, is no greater than q_0 , which can be viewed as a target outage probability.

The resource allocation problem can be formally stated as

Problem MAXU:

$$\underset{P: \mathcal{H} \rightarrow \mathbb{R}^+}{\text{maximize}} \quad \lambda \mathbb{E}_H(\tilde{U}_H(P(H))) \quad (5)$$

$$\text{subject to} \quad \Pr(P_{sum} > \bar{P} | P_{sum} > 0) \leq q_0 \quad (6)$$

Note that conditioning on $P_{sum} > 0$ (the system not being empty) is needed to avoid the impulsive solution in which each packet is transmitted with infinite power and has infinitesimal duration.

Solving Problem MAXU directly appears to be difficult in general. In the next section we simplify the problem by approximating P_{sum} as a Gaussian random variable. This can be justified when the number of active users contributing to P_{sum} is large. To see when this is likely to be true, consider the special case in which all users within the system have the same channel gain h . With this assumption the solution to MAXU is given by a single value P . For the sake of this example, we further assume that the utility function is given by $U(R) = 1 - e^{-\mu R}$.

Since each active user is assigned the same power, the average utility is given by

$$U_{avg} = \lambda \left(1 - e^{-\mu \frac{L}{T}} \right)$$

where T is the transmission time for a packet of length L , and λ is the arrival rate for transmission requests. Clearly, we wish to choose P to minimize T . Hence we select the largest value of P , which satisfies the constraint (6). Assuming Poisson arrivals, and noting that T is the same for all packets, the number of active (transmitting) users, N , is the occupancy of an $M/D/\infty$ queue, which is Poisson with parameter λT . Hence the probability that the system is not empty

is $\Pr[N > 0] = \Pr[P_{sum} > 0] = 1 - e^{-\lambda T}$, and

$$\begin{aligned} \Pr[P_{sum} > \bar{P} | P_{sum} > 0] &= \frac{1}{1 - e^{-\lambda T}} \sum_{n=0}^{\infty} \Pr[P_{sum} > \bar{P} | N = n] \Pr[N = n] \\ &= \frac{1}{1 - e^{-\lambda T}} \sum_{n=0}^{\infty} \mathcal{I}_{n > \bar{P}/P} \Pr[N = n] \\ &= \frac{1}{1 - e^{-\lambda T}} \sum_{n=n_0}^{\infty} \frac{e^{-\lambda T} (\lambda T)^n}{n!} \\ &\leq q_0 \end{aligned}$$

where \mathcal{I}_V is the indicator function for the event V , and $n_0 = \lceil \frac{\bar{P}}{P} \rceil$ is the minimum number of active users, which causes the total power to exceed \bar{P} . Given any $q_0 > 0$, we can choose P small enough (equivalently, n_0 large enough) to satisfy the preceding outage constraint. The objective is then to find the smallest n_0 , and the corresponding largest P , such that the constraint is satisfied.

If λT is large enough, then $\Pr[P_{sum} > 0] = 1 - e^{-\lambda T} \approx 1$ and the Poisson random variable N can be accurately approximated as Gaussian. In that case, the constraint (6) can be replaced by the constraint $\Pr[P_{sum} > 0] < q_0$, where P_{sum} is Gaussian. To see how λT depends on the target outage probability q_0 , Fig. 2 shows $\Pr[P_{sum} > \bar{P} | P_{sum} > 0]$ vs. λT for different ratios $n_0/\lambda T$ (i.e., n_0 is normalized by the average number of active users $\lambda T = \mathbb{E}[N]$). The discontinuities in the plots are due to the ceiling function used to define n_0 . As $n_0/(\lambda T)$ increases, $\Pr[P_{sum} > \bar{P} | P_{sum} > 0]$ must increase, as shown in the figure. Furthermore, the figure shows that given a target q_0 (e.g., $< 5\%$), we must have $n_0/(\lambda T) > 1$ and $\lambda T > 5$. The Gaussian approximation for P_{sum} is therefore accurate in this scenario.

Verifying the accuracy of the Gaussian approximation by solving MAXU directly becomes significantly more difficult with more general channel distributions. However, the preceding analysis indicates that when $P(h)$ is optimized, the number of active users should be relatively large when P_{sum} is close to \bar{P} . Hence for the analytical results, which follow, we will assume that the distribution of P_{sum} has a Gaussian tail. We remark that the accuracy of the Gaussian assumption also depends on the choice of utility function. In particular, it is less accurate for a

logarithmic utility function, as discussed in Section V.

IV. UTILITY BASED POWER ALLOCATION

A. Decomposition of Power Constraint

Let δh be a small constant such that $h_{max} - h_{min} = K\delta h$, for some integer K . For $i = 0, \dots, K$, define $h_i = h_{min} + i\delta h$. For $i = 0, \dots, K - 1$, let $N(i)$ be a random variable representing the number of active users in steady-state with channel gain in $[h_i, h_{i+1})$. The steady-state total power, P_{sum} , can then be approximated as:

$$P_{sum} \approx \sum_{i=0}^{K-1} P(h_i)N(i).$$

Taking expected values, we have

$$\mathbb{E}(P_{sum}) \approx \sum_{i=0}^{K-1} P(h_i)\bar{N}(i),$$

where $\bar{N}(i)$ is the expected number of active users with channel gains in $[h_i, h_{i+1})$. Since arrivals are Poisson with overall rate λ , $N(i)$ is the occupancy of a $M/G/\infty$ queue with arrival rate $\approx \lambda f_H(h_i)\delta h$ and service time $\approx T(h_i)$. Therefore $N(i)$ is approximately Poisson distributed, and

$$\bar{N}(i) \approx \lambda f_H(h_i)\delta h T(h_i) = \bar{N}(h_i)\delta h, \quad (7)$$

where $\bar{N}(h_i) = \lambda f_H(h_i)T(h_i)$. Assuming that $P(h)\bar{N}(h)$ is Riemann integrable, then letting $\delta h \rightarrow 0$, we have

$$\mathbb{E}(P_{sum}) = \int_{\mathcal{H}} P(h)\bar{N}(h) dh. \quad (8)$$

Likewise, since $N(i)$, $i = 0, \dots, K - 1$ are independent, the second moment of P_{sum} is given by

$$\mathbb{E}(P_{sum}^2) = \int_{\mathcal{H}} P^2(h)\bar{N}(h) dh. \quad (9)$$

For a large number of active users, P_{sum} can be approximated as a Gaussian random variable. As discussed in the preceding section, we therefore rewrite the constraint (6) as

$$\Pr [P_{sum} > \bar{P} | P_{sum} > 0] \approx Q \left(\frac{\bar{P} - \mathbb{E}(P_{sum})}{\sqrt{\text{Var}(P_{sum})}} \right) \leq q_0$$

where $Q(x) = \int_x^\infty \frac{1}{2\pi} e^{-t^2/2} dt$ is the complementary Gaussian cumulative distribution function (*c.d.f.*).

This constraint reduces to

$$\int_{\mathcal{H}} P(h) \bar{N}(h) dh + k_1 \sqrt{\int_{\mathcal{H}} P^2(h) \bar{N}(h) dh} \leq \bar{P}, \quad (10)$$

where $k_1 = Q^{-1}(q_0)$.

Since $\bar{N}(h) = \lambda f_H(h) T(h)$, we have

$$\int_{\mathcal{H}} P(h) \bar{N}(h) dh = \lambda \mathbb{E}_H[E(H)], \quad (11)$$

and

$$\int_{\mathcal{H}} P^2(h) \bar{N}(h) dh = \lambda \mathbb{E}_H[P(H)E(H)], \quad (12)$$

where $E(h) = P(h)T(h)$ is the energy consumed by user with channel gain h . An inactive user is allocated zero energy.

Substituting (11) and (12) into (10), constraint (6) can be approximated by

$$\lambda \mathbb{E}_H(E(H)) + k_1 \sqrt{\lambda \mathbb{E}_H(P(H)E(H))} \leq \bar{P}. \quad (13)$$

Finally, this can be further decomposed into the three constraints

$$\begin{cases} \lambda \mathbb{E}_H(E(H)) \leq \mathcal{E} & \text{average energy} \\ \mathbb{E}_H(P(H)E(H)) \leq G & \text{average power} \times \text{energy} \\ \mathcal{E} + k_1 \sqrt{\lambda G} \leq \bar{P} & \text{tradeoff of } \mathcal{E} \text{ vs. } G \end{cases} \quad (14)$$

We will refer to Problem MAXU when (6) is replaced with (14) as *Problem MAXUA*. A solution to Problem MAXUA is provided next. This is accomplished in two steps. First, we find the utility maximizing power assignment subject to the first two constraints in (14) for given values of \mathcal{E} and G . Next, the combination of \mathcal{E} and G that yields the highest utility rate is derived.

B. Solution with Fixed \mathcal{E} and G

Given values for \mathcal{E} and G , consider the following problem:

Problem P1:

$$\underset{P:\mathcal{H}\rightarrow\mathbb{R}^+}{\text{maximize}} \quad \lambda\mathbb{E}_H(\tilde{U}_H(P(H))) \quad (15)$$

$$\text{subject to} \quad \lambda\mathbb{E}_H(E(H)) \leq \mathcal{E} \quad (16)$$

$$\mathbb{E}_H(P(H)E(H)) \leq G \quad (17)$$

To gain insight into this problem, we consider each of the constraints separately. First, we examine the problem with only the energy constraint, i.e.,

Problem P2:

$$\underset{P:\mathcal{H}\rightarrow\mathbb{R}^+}{\text{maximize}} \quad \lambda\mathbb{E}_H(\tilde{U}_H(P(H)))$$

$$\text{subject to} \quad \lambda\mathbb{E}_H(E(H)) \leq \mathcal{E}.$$

From (8) and (11), $\mathbb{E}(P_{sum}) = \lambda\mathbb{E}_H(E(H))$, so that Problem P2 is equivalent to constraining the average sum power.

To continue, we assume that the transmission rate is proportional to the received power, i.e.,

$$\mathcal{C}(hP(h)) = k_0hP(h) \quad (18)$$

where k_0 is a constant.⁴ It follows directly from (18) that the energy consumed by a user depends only on whether a user's transmission power is nonzero, and not on the specific power level, i.e.,

$$E(h) = \begin{cases} P(h)T(h) = L/k_0h, & \text{for } P(h) > 0, \\ 0, & \text{for } P(h) = 0. \end{cases} \quad (19)$$

Since utility is strictly increasing in received power, it follows from (19) that the solution to Problem P2 is for each packet to be either denied transmission (blocked) or transmitted with

⁴This linear relationship between rate and power is a reasonable approximation for many practical systems, e.g., with low SNR and/or high bandwidth. For large enough rates, capacity considerations imply that this is optimistic.

infinite power. If no users are blocked and the energy constraint (16) is violated, then admission control is required to block some users. This solution is stated as the following lemma.

Lemma 1: A power allocation, which achieves the maximum average steady-state utility in Problem P2, satisfies

$$P(h) = \begin{cases} \infty, & \text{for } h > h_e, \\ 0, & \text{for } h < h_e, \end{cases} \quad (20)$$

where h_e is the minimum value in \mathcal{H} such that (16) is satisfied.

The lemma follows from the preceding discussion and by noting that the energy required by a user decreases with the channel gain. Hence, blocking those users with smallest channel gains minimizes the number of blocked users and maximizes (5). We note that if $U(R)$ is unbounded, then the solution to Problem P2 is also unbounded, so that an *arbitrary* set of users can be blocked. The Lemma implies that $P_{sum}(t) = 0$ with probability one (i.e., for almost all t), and $P_{sum}(t) = \infty$ whenever a new request arrives.⁵ Of course, this power assignment is not realistic. This type of behavior is eliminated by adding the constraint (17).

Next we consider Problem P1 with *only* constraint (17).

Problem P3:

$$\begin{aligned} & \underset{P: \mathcal{H} \rightarrow \mathbb{R}^+}{\text{maximize}} && \lambda \mathbb{E}_H(\tilde{U}_H(P(H))) \\ & \text{subject to} && \mathbb{E}_H(P(H)E(H)) \leq G. \end{aligned}$$

This is a standard optimization problem with a concave objective and linear constraints,⁶ and is mathematically equivalent to the problem studied in [13]. As in [13], the solution can be attained via a pricing scheme.

Lemma 2: Consider the following pricing scheme: a channel-dependent price per unit transmit power of the form $\alpha_p(h) = \alpha E(h)$ is announced; users respond by requesting power to maximize

⁵This type of *flash* signaling also arises in the context of ultra-wideband communications [23], in which case the assumed linear rate-power relation is valid.

⁶Note, we are still assuming the linear relationship between rate and power, in which case $E(H)$ can be viewed as a constant independent of $P(H)$.

their surplus (utility minus cost), i.e.,

$$P^*(h) = \arg \max_{P(h)} \{\tilde{U}_h[P(h)] - \alpha E(h)P(h)\}. \quad (21)$$

If α is set such that (17) is satisfied with equality, then this pricing scheme provides a power allocation that is the solution to Problem P3.

This lemma follows directly from the Kuhn-Tucker optimality conditions, where α corresponds to a Lagrange multiplier for constraint (17). The set of active users and the assigned power levels are determined by α , which can be interpreted as a fixed unit price on the product of power times energy. For each active user, the marginal utility with respect to power equals the price per unit power, $\frac{d\tilde{U}_h(P(h))}{dP(h)} = \alpha_p(h)$. Inactive users have lower marginal utility than the price at zero power, i.e., $\frac{d\tilde{U}_h(P(h))}{dP(h)} \Big|_{P(h)=0} < \alpha_p(h)$. Since $\tilde{U}_h(P(h))$ is concave, $\frac{d\tilde{U}_h(P(h))}{dP(h)}$ is decreasing with $P(h)$. Hence for inactive users, a positive power assignment provides less utility than the cost (negative surplus). We call those inactive users *intimidated* due to a combination of high price and small initial slope of $\tilde{U}_h(P)$.

Assuming all users have the same $U(\cdot)$ and that (18) holds, the set of users that are intimidated can be characterized as follows:

Theorem 1: There exists a threshold $h_i \in \mathcal{H}$ such that the optimal power allocation to Problem P3 satisfies $P(h) > 0$ if and only if $h > h_i$. The threshold h_i satisfies:

$$\left. \frac{dU(R)}{dR} \right|_{R=0} = \frac{\alpha(h_i)}{k_0 h_i}.$$

The theorem follows directly from the fact that $\frac{dU(R)}{dR} = \frac{d\tilde{U}_h(P)}{dP} \frac{dP}{dR}$ and that $\alpha_p(h)$ is decreasing in h . This theorem implies that given two users with different channel gains, the user with the smaller channel gain is penalized twice. First, that user requires more power to achieve a target SINR, and second, the user is charged a higher unit price per power. Notice that as G increases, α decreases and $P(h)$ increases for all active users. This in turn increases the utility for each active user, and hence results in a higher utility rate. Also notice that the constraint in Problem P3 does not depend on the traffic intensity λ , but only on the channel distribution, $f_H(h)$. It

follows that changes in the arrival rate, for a fixed $f_H(h)$, do not effect the optimal price in Theorem 1.

Now we return to Problem P1. The solution to this problem is a combination of admission control, as in Lemma 1, and the pricing procedure stated in Lemma 2. The resource allocation can be accomplished in two steps. First, the admission control step specifies an active channel set $\mathcal{H}_a = \{h : P(h) > 0\}$. That is, users with $h \notin \mathcal{H}_a$ are blocked. Second, the pricing step determines the power assignments across users in the active set. Note that some users not blocked in the first step still may be intimidated in the second step.

Suppose that the average energy is $\lambda \mathbb{E}_H(E(H)) = \hat{\mathcal{E}}$ for some $\hat{\mathcal{E}} \in [0, \mathcal{E}]$. Conditioned on this, the solution to Problem P1 is given as follows:

- 1) Assume $P(h) > 0$ for any h . Given $f_H(h)$ and $E(h)$ in (19), check if $\lambda \mathbb{E}_H(E(h)) \leq \hat{\mathcal{E}}$. If so, admit all users. Otherwise, block users with channel gains $h \leq h_e(\hat{\mathcal{E}})$ where $h_e(\hat{\mathcal{E}})$ is selected to satisfy

$$\lambda \mathbb{E}_H(E(H) | H > h_e(\hat{\mathcal{E}})) \Pr(H > h_e(\hat{\mathcal{E}})) = \hat{\mathcal{E}}.$$

- 2) Find α so that (17) is binding for the set of active users taking into account that users blocked in the previous step are assigned zero power. The optimal power allocation across active users is given by (21).

Finally, the solution to Problem P1 can be found by searching for the value $\hat{\mathcal{E}} \in [0, \mathcal{E}]$ that maximizes the total utility rate $\lambda \mathbb{E}\{\tilde{U}_h[P(h)]\}$.

For a fixed $\hat{\mathcal{E}}$, users with the lowest channel gains are blocked because they derive the lowest utility for any given α . Therefore, there exists an energy induced threshold $h_e(\hat{\mathcal{E}})$ such that users with $h \leq h_e(\hat{\mathcal{E}})$ are blocked via admission control. Recall that following Lemma 1, we concluded that blocked users should have the worst channels. This conclusion assumed only an energy constraint and bounded utility functions. Here we have shown that this conclusion is valid with both energy and power-times-energy (G) constraints, and any increasing concave utility function. We note that at the optimum, (17) is always binding, whereas (16) might not

be binding.

Note that lowering $\hat{\mathcal{E}}$ may decrease the size of the active set, but also increases the average utility derived per active packet.⁷ A complete solution to Problem P1 requires finding the optimal $\hat{\mathcal{E}} \in [0, \mathcal{E}]$ to balance this tradeoff. Next we show that this search is simplified when we include the last constraint in Problem MAXUA.

C. Optimal Admission Control/Pricing Trade-off

Given \mathcal{E} and G , we have shown that the optimal solution to P1 consists of a combination of admission control and pricing. Returning to problem MAXU, notice that any pair of values, \mathcal{E} and G , that satisfy

$$\mathcal{E} + k_1 \sqrt{\lambda G} \leq \bar{P} \quad (22)$$

results in a solution to Problem P1 that is also a feasible power allocation for Problem MAXUA. The solution to Problem MAXUA is given by the combination that maximizes the utility rate.

Theorem 2: The power allocation which solves Problem MAXUA satisfies both (16) and (17) with equality.

Proof: As noted previously, the constraint (17) is tight under an optimal power allocation. From this it can be seen that the utility rate in Problem P1 increases monotonically with G . Suppose the energy constraint (16) is loose. Then \mathcal{E} can be decreased to the point where the energy constraint is tight, resulting in a larger G in (22), which in turn gives a higher utility rate. ■

To solve Problem MAXUA, we can therefore use the following procedure. First, for each pair $(\mathcal{E}, G = \frac{1}{\lambda} \left(\frac{\bar{P} - \mathcal{E}}{k_1} \right)^2)$, a feasible solution to Problem P1 can be found via the previous steps 1 and 2 with $\hat{\mathcal{E}} = \mathcal{E}$. Letting $\mathcal{U}(\mathcal{E})$ be the resulting utility rate, the solution to Problem MAXUA is then given by the solution to Problem P1, where \mathcal{E} is replaced by $\mathcal{E}^* = \arg \max\{\mathcal{U}(\mathcal{E}), 0 \leq \mathcal{E} \leq \bar{P}\}$. That is, the solution to MAXUA is achieved with $\mathcal{E} = \mathcal{E}^*$. This is because for each pair (\mathcal{E}, G) ,

⁷Given an arrival rate λ , $\hat{\mathcal{E}}$ may change over a range in which all users remain active.

for which the utility is evaluated, Theorem 2 implies that there is no need to search for the optimal $\hat{\mathcal{E}} \in [0, \mathcal{E}]$.

The set of users blocked through admission control and intimidation is determined by the channel gain thresholds $h_e(\mathcal{E})$ and $h_i(\mathcal{E})$, respectively. We distinguish the following 3 cases:

C1: $h_e \geq h_{min}$ and $h_e \geq h_i$. (Active users are determined by h_e .)

C2: $h_{min} > h_e$ and $h_{min} \geq h_i$. (All users are active.)

C3: $h_i > h_e$ and $h_i > h_{min}$. (Active users are determined by h_i .)

The next theorem characterizes the transition between these cases.

Theorem 3: Consider Problem P1 with constraints $(\mathcal{E}, G = \frac{1}{\lambda} \left(\frac{\bar{P}-\mathcal{E}}{k_1} \right)^2)$. As \mathcal{E} increases from 0 to \bar{P} , the optimal power allocation transitions through the cases C1, C2, C3 in one of the following sequences: C1 \rightarrow C2 \rightarrow C3 or C1 \rightarrow C3.

Proof: Let A1 denote the set of values of \mathcal{E} where the optimal solution to P1 is in C1. Define A2 and A3 similarly. At $\mathcal{E} = 0$, $h_e = h_{max}$ and $h_i = 0$; therefore $0 \in A1$. As \mathcal{E} increases, G decreases; this results in h_e decreasing with \mathcal{E} and h_i increasing. This implies that if $\mathcal{E} \in A1$ then $\mathcal{E}' \in A1$ for all $\mathcal{E}' \leq \mathcal{E}$ and likewise, if $\mathcal{E} \in A3$ then $\mathcal{E}' \in A3$ for all $\mathcal{E}' \geq \mathcal{E}$. When $\mathcal{E} = \bar{P}$, $G = 0$, in which case $h_i = \infty$, thus, $\bar{P} \in A3$. Therefore the only possible sequences are C1 \rightarrow C2 \rightarrow C3 or C1 \rightarrow C3. Which of these occurs depends on whether or not $h_{min} \leq h_e(\tilde{\mathcal{E}})$, where $\tilde{\mathcal{E}}$ satisfies $h_e(\tilde{\mathcal{E}}) = h_i(\tilde{\mathcal{E}})$. (See Figure 6.) ■

Corollary: The optimal $\mathcal{E}^* \in A1$.

This follows from the observation that A1 is the only region where both constraints are tight. In A2 or A3, the energy constraint is always loose.

V. NUMERICAL RESULTS

In this section, we present numerical results to illustrate the optimization described in the preceding section. The results that follow assume the exponential utility function $U(R) = 1 - \exp(-\mu R)$ for $\mu > 0$, which is concave, increasing, and has $U(0) = 0$. The channel density is given by $f_H(h) = \frac{1}{4}h^{-\frac{5}{4}}$ for $h \in (1, \infty)$. This corresponds to a channel gain $h(r) = r^{-4}$, where

r is the distance of a user from the base station, and each user's location is chosen uniformly in the interval $(0, 1)$. We assume the file length is normalized so that $L = k_0$, the scale factor in (18), which relates transmission rate to received power. That is, one unit of received power results in a completion time of one unit. From (21), the surplus maximizing power assignment for active packets with $h > \max(1, h_e, h_i)$ and price α is given by $P(h) = \frac{1}{\mu k_0 h} \ln \left(\frac{k_0^2 h^2 \mu}{\alpha L} \right)$.

A. Accuracy of Gaussian Approximation

We first illustrate the accuracy of the Gaussian approximation, which is used to estimate the outage probability.⁸ Figure 3(a) compares the simulated outage probability with the Gaussian approximation as a function of the utility parameter μ , assuming a target outage probability $q_0 = 0.05$ and average power $\bar{P} = 10$. For the same parameters, Figure 3(b) shows the average number of active packets from simulation and using the Gaussian approximation, as a function of μ . In both cases, simulated curves are shown for different packet arrival rates λ . In the simulation model, packets arrive according to a Poisson process, and are either blocked ($h < h_e$), intimidated ($h < h_i$), or served at the rate $k_0 h P(h)$. The simulated outage probability is then the fraction of time for which $P_{sum}(t) > \bar{P}$. The analytical curves are obtained using the optimal (\mathcal{E}^*, G^*) for each parameter setting.

In Figure 3(a), the simulated outage probability approaches the target q_0 as either λ or μ increases. The gap is more sensitive to the arrival rate λ than the utility parameter μ . Figure 3(b) shows that the simulated and analytical values of the average number of active packets are nearly identical. As μ and λ increase, the average number of active users increases. Furthermore, the results show that the average number of active packets in the system increases more rapidly with λ than with μ .

⁸The comparisons with simulation results shown here illustrate how close the simulated outage probability is to that obtained with the Gaussian approximation given specific model parameters. These comparisons do not illustrate how well the solution to MAXUA approximates the solution to the problem MAXU, since as discussed in Section III, solving MAXUA directly appears to be difficult.

Figure 3(b) shows that the average number of active packets \bar{N} varies approximately linearly with μ . To see why, from the analysis in Section IV-A, we can write

$$\begin{aligned}\bar{N} &= \int_{\max\{1, h_e, h_i\}}^{\infty} \bar{N}(h) dh \\ &= \int_{\max\{1, h_e, h_i\}}^{\infty} \lambda f_H(h) T(h) dh.\end{aligned}\quad (23)$$

With the exponential utility function, we have $T(h) = L/k_0 h P(h) = \frac{\mu L}{\log(\frac{\mu k_0^2 h^2}{\alpha L})}$, and combining this with (23) gives

$$\bar{N} = \lambda \mu L \int_{\max\{1, h_e, h_i\}}^{\infty} f_H(h) \frac{1}{\log(\frac{\mu k_0^2 h^2}{\alpha L})} dh.\quad (24)$$

The integral varies slowly with respect to λ and μ , so that \bar{N} is nearly proportional to $\lambda \mu$.

From these results we conclude that the Gaussian approximation is accurate when the average number of active users is relatively large, i.e., greater than ten. Furthermore, the approximate Gaussian mean $m_G = \mathbb{E}(P_{sum}) = \mathcal{E}^*$ and standard deviation $\sigma_G = \sqrt{Var(P_{sum})} = \sqrt{\lambda G^*}$ must satisfy $m_G = c \sigma_G$ with $c \gtrsim 3$. This is due to the fact that $P_{sum} \geq 0$, hence a Gaussian distribution with significant mass in the negative region cannot be a good approximation for the distribution of P_{sum} . (We have observed that the ratio m_G/σ_G tends to increase with μ and λ .) To illustrate this point, Figure 4 shows the empirical density function for the total power P_{sum} with two different arrival rates.⁹ The Gaussian distribution with mean $m_G = \mathcal{E}^*$ and variance $\sigma_G^2 = \lambda G^*$ is also shown. When $\lambda = 100$, the Gaussian and empirical densities are nearly identical. The approximation is not as accurate with $\lambda = 25$, although it is still reasonable.

The accuracy of the Gaussian approximation also depends on the assigned utility functions. Additional results in [24] show similar trends to those shown here for a piecewise linear utility function with one breakpoint. In contrast, the Gaussian approximation is typically not accurate for the logarithmic utility function $U(R) = \log(1 + \mu R)$. Namely, the empirical *p.d.f.* has a much heavier tail than the Gaussian *p.d.f.*, due to the fact that the assigned power $P(h) = \frac{k_0 h}{\alpha L} - \frac{1}{\mu k_0 h}$ approaches infinity as the user gets close to the base station.

⁹This is computed from a histogram of the total powers seen at packet arrivals.

B. Utility and Optimized Power Allocations

In this section we show numerical results for utility rate and power allocations based on the Gaussian approximation. Figure 5 shows how the average utility per user, $\mathbb{E}_H \tilde{U}_H[P(H)]$, varies with λ and \mathcal{E} (and therefore G) when $\bar{P} = 10$ and $q_0 = 0.01$. The classification of the resulting allocation, as in Theorem 3, is also indicated on the figure. The maximum point is always in A1, as stated in the Corollary. As \mathcal{E} increases, the solution transitions from C1 \rightarrow C2 \rightarrow C3 when λ is small (i.e., $\lambda = 1, 10, 20$). For larger λ , the allocation transitions directly from C1 to C3.

Figure 6 shows how $h_e(\mathcal{E})$ and $h_i(\mathcal{E})$ vary with \mathcal{E} given different arrival rates λ . The minimum channel gain $h_{min} = 1$ is also shown. For $h_e(\mathcal{E}) < h_{min}$, we choose h_e to satisfy $\lambda \int_{h_e(\mathcal{E})}^{h_{max}} \frac{1}{4} h^{-\frac{5}{4}} \frac{L}{k_0 h} dh = \mathcal{E}$ so that the curve is extended continuously from where $h_e(\mathcal{E}) \geq h_{min}$. As expected, $h_e(\mathcal{E})$ decreases with \mathcal{E} , whereas $h_i(\mathcal{E})$ increases with \mathcal{E} . When $\lambda = 10$, the system transitions from C1 to C2 when h_e falls below h_{min} , and from C2 to C3 when h_i increases above h_{min} . When $\lambda = 40$, the intersection point of h_e and h_i is larger than h_{min} ; in this case the solution transitions directly from C1 to C3.

The resulting optimal power allocation is shown in Figure 7(a) as a function of the designated user's distance from the base station with $\lambda = 20, 30, 50$. The closer a user is to the base station, the better the channel. The received power, which is proportional to the data rate, is also shown in Figure 7(b). The active radius shrinks slightly as λ increases. As the traffic intensity increases, the optimal allocation blocks the users with the worst channels while trying to maintain the received power level for the remaining active users (see Figure 7(b)). Apparently, this blocking strategy yields a higher utility rate than offering a relatively degraded service to all users.

Figure 8 shows the maximum average utility per user $\mathbb{E}_H \{\tilde{U}_h[P(h)]\}$ versus the target outage probability q_0 with different arrival rates. As λ increases, the average utility per user decreases; however, the overall utility rate $\lambda \mathbb{E}_H \{\tilde{U}_h[P(h)]\}$ increases. A smaller q_0 , or equivalently, a tighter power constraint, results in a lower utility per user. Notice that the average utility per user is

insensitive to q_0 when λ is small (< 10). This is because the exponential utility $U(R) = 1 - e^{-R}$ is relatively flat (close to one) when R becomes large. When λ is small, the optimal power allocation to active users is quite large, so that the corresponding transmission rates (R) are also large. Therefore, the average utility per user is close to one.

VI. CONCLUSIONS

We have studied forward link power allocation for stochastically varying data traffic. Each power assignment remains constant for the duration of the packet, and along with the channel gain and associated utility function, determines the utility for transmitting the packet. The objective is to determine a power assignment policy that maximizes the time average utility rate. We introduced an outage constraint on the total power in order to derive a simple power assignment policy, which depends only on steady-state system properties. Specifically, this policy depends only on the distribution of channel gains, packet arrival rate, and utility functions. Each power assignment then depends only on the designated user's channel state and associated utility function.

By approximating the steady-state total power as a Gaussian random variable, the outage constraint was decomposed into simpler constraints on energy and power times energy. This approximation is accurate provided that the average number of packets being simultaneously transmitted in steady-state is large enough, corresponding to a heavily loaded system. A procedure for maximizing the time-averaged utility rate was presented, which enforces those constraints, respectively through a combination of admission control and pricing of power times energy. The optimal combination depends on the system characteristics, namely, the packet arrival rate, assigned utility functions, and distribution of channel gains. Numerical results were presented to illustrate the effect of the constraints, the optimal power allocation, and the corresponding utility rate. The results show that it can be beneficial to block users, rather than use pricing only to enforce the constraints.

REFERENCES

- [1] S. Shenker, "Fundamental design issues for the future Internet," *IEEE Journal on Selected Areas in Communications*, vol.13, pp. 1176-1188, 1995.
- [2] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *Proc. SPIE '98 International Symposium on Voice, Video and Data Communication*, Boston, MA, Oct. 1998.
- [3] H. Ji and C.-Y. Huang, "Non-cooperative uplink power control in cellular radio systems," *Wireless Networks*, vol. 4, pp. 233-240, Apr. 1998.
- [4] N. Feng, N. Mandayam, and D. J. Goodman, "Joint power and rate optimization for wireless data services based on utility functions", *Proc. CISS 1999*, Johns Hopkins University, March 1999.
- [5] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1883-1892, Oct. 2001.
- [6] T. Alpcan, T. Basar, R. Srikant, E. Altman, "CDMA Uplink Power Control as a Noncooperative Game," *Wireless Networks*, Vol. 8, No. 6, pp. 659-670, Nov. 2002.
- [7] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks", *IEEE Transactions on Communications*, Vol. 50, No. 2, pp. 291-303, Feb. 2002.
- [8] C. W. Sung and W. S. Wong, "A noncooperative power control game for multirate CDMA data networks," *IEEE Trans. on Wireless Communications*, vol. 2, pp. 186-194, Jan. 2003.
- [9] P. Liu, M. Honig and S. Jordan, "Forward-Link CDMA Resource Allocation Based on Pricing", *Proceedings IEEE Wireless Communications and Networking Conference*, Chicago, IL, Oct. 2000.
- [10] L. Song and N. B. Mandayam, "Hierarchical SIR and rate control on the forward link for CDMA data users under delay and error constraints," *IEEE Journal on Selected Areas in Communications*, 19(10):1871-1882, October 2001.
- [11] J-W. Lee, R. Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class CDMA wireless networks," *Proceedings of IEEE INFOCOM 2002*, 2002.
- [12] M. Xiao, N. Shroff, and E. Chong, "A Utility-Based Power-Control Scheme in Wireless Cellular Systems," *IEEE/ACM Trans. on Networking*, Vol. 11, No. 2, Apr. 2003.
- [13] P. Liu, P. Zhang, S. Jordan, and M. L. Honig, "Single-Cell Forward Link Power Allocation Using Pricing in Wireless Networks", *IEEE Trans. on Wireless Communications*, vol. 3, March, 2004.
- [14] C. Zhou, P. Zhang, M. L. Honig, and S. Jordan, "Two-Cell Power Allocation for Downlink CDMA", *IEEE Transactions on Wireless Communications*, Vol. 3, No. 6, Nov. 2004.
- [15] C. Zhou, M. L. Honig, and S. Jordan, "Two-Cell Power Control for Downlink CDMA Data", *IEEE Transactions on Wireless Communications*, to appear.
- [16] J-B Kim, M. L. Honig, and S. Jordan, "Dynamic Resource Allocation for Integrated Voice and Data Traffic in DS-CDMA", *Proc. IEEE Vehicular Tech. Conf.*, Atlantic City, NJ, Sept. 2001.
- [17] S. Kalyanasundaram, M. Needham and R. Agrawal, "Utility function-based optimal resource allocation with a mixture of reallocation-tolerant and reallocation-intolerant users," *Proc. of SPIE*, vol. 4531, Aug. 2001.

- [18] Bender, P., Black, P., Grob, M., Padovani, R., Sindhushayana, N., Viterbi, A. "CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, no. 7, pp. 70-77, July 2000.
- [19] S. Borst, "User-Level Performance of Channel-Aware Scheduling Algorithms in Wireless Data Networks," *Proc. of IEEE INFOCOM 2003*.
- [20] H. Kushner and P. Whiting, "Asymptotic Properties of Proportional-Fair Sharing Algorithms," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.
- [21] R. Agrawal and V. Subramanian, "Optimality of Certain Channel Aware Scheduling Policies," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.
- [22] P. Liu, R. Berry and M. Honig, "A Fluid Analysis of Utility-based Wireless Scheduling Policies," *Proc. of IEEE Conference on Decision and Control*, Paradise Island, the Bahamas, Dec. 2004.
- [23] S. Verdú, "Spectral Efficiency in the Wideband Regime", *IEEE Transactions on Information Theory*, Vol. 48, No. 6, pp. 1319-1343, June 2002.
- [24] P. Liu, "Utility-based resource allocation for wireless networks in static and dynamic environments", Ph.D. Thesis, Northwestern University, Evanston, IL, December 2003.

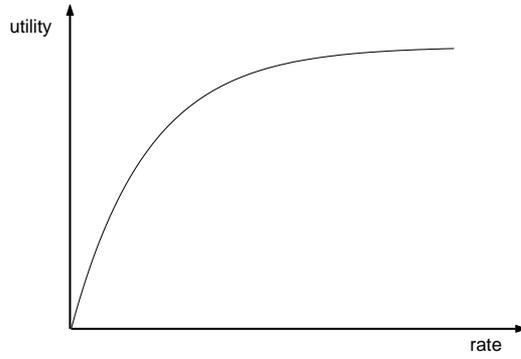


Fig. 1. Example utility function for data traffic.

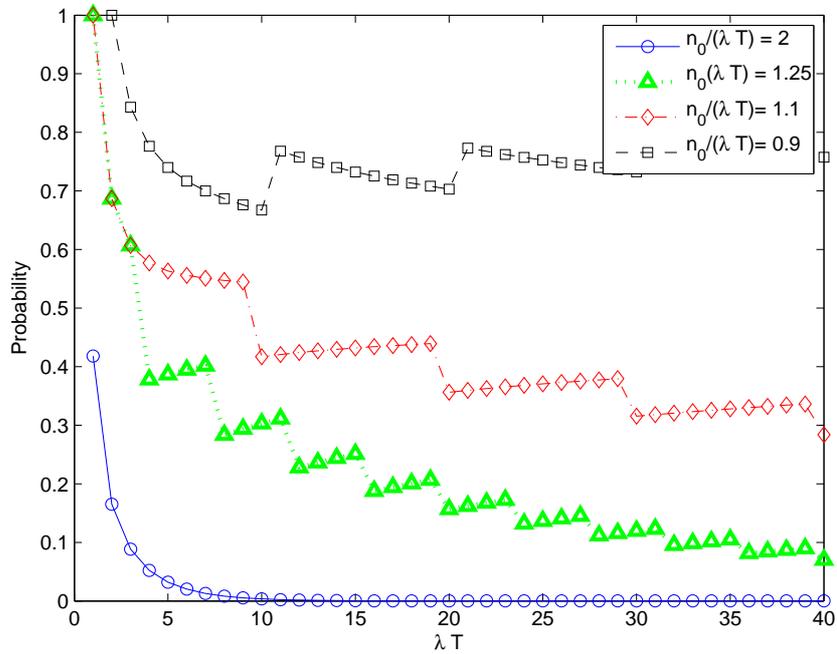


Fig. 2. $\Pr[P_{sum} > \bar{P} | P_{sum} > 0]$ versus λT for different ratios $n_0/\lambda T$.

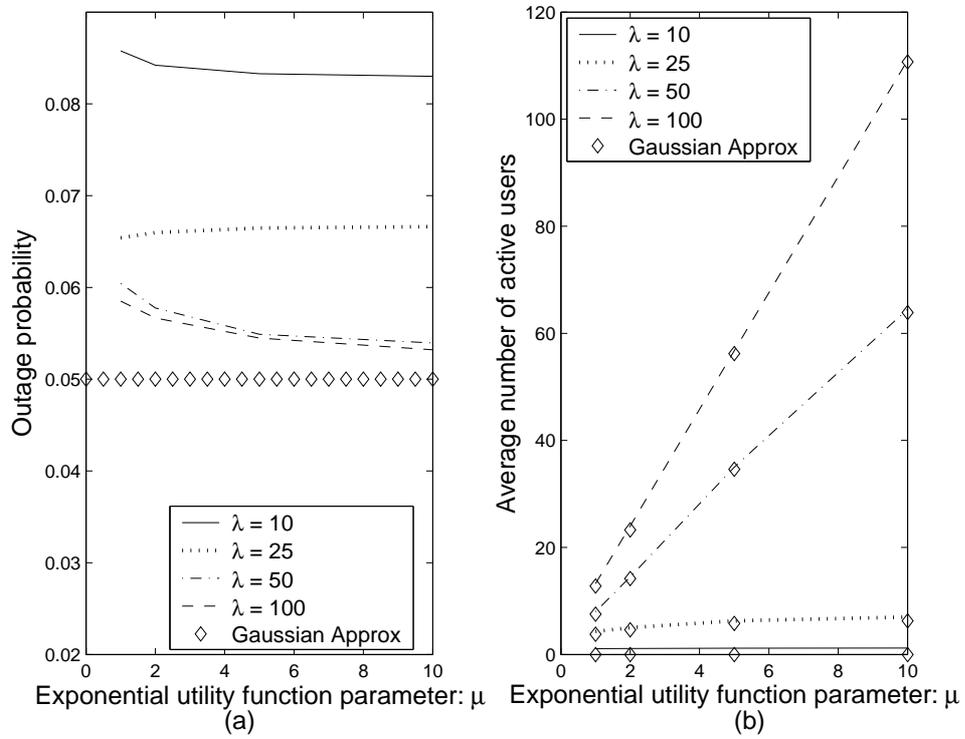


Fig. 3. (a) Outage probability vs. μ with different arrival rates; (b) Average number of active packet vs. μ with different arrival rates.

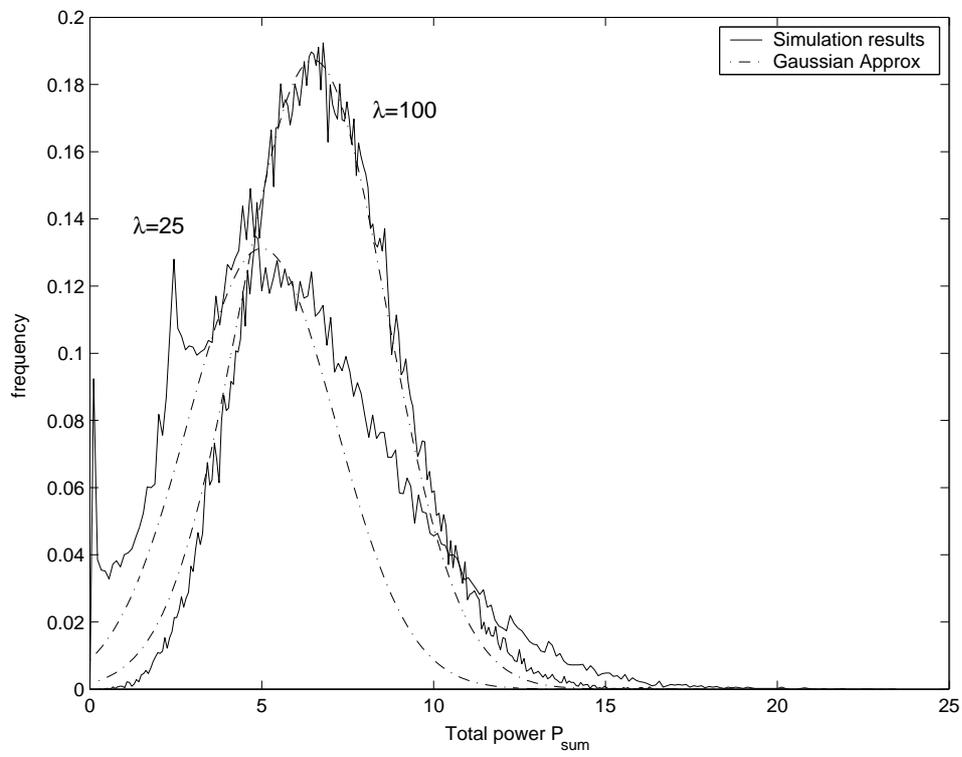


Fig. 4. The empirical p.d.f. of P_{sum} from simulation and the Gaussian approximation ($m_G = \mathcal{E}^*$, $\sigma_G = \sqrt{\lambda G^*}$) for $\lambda = 25$ and 100.

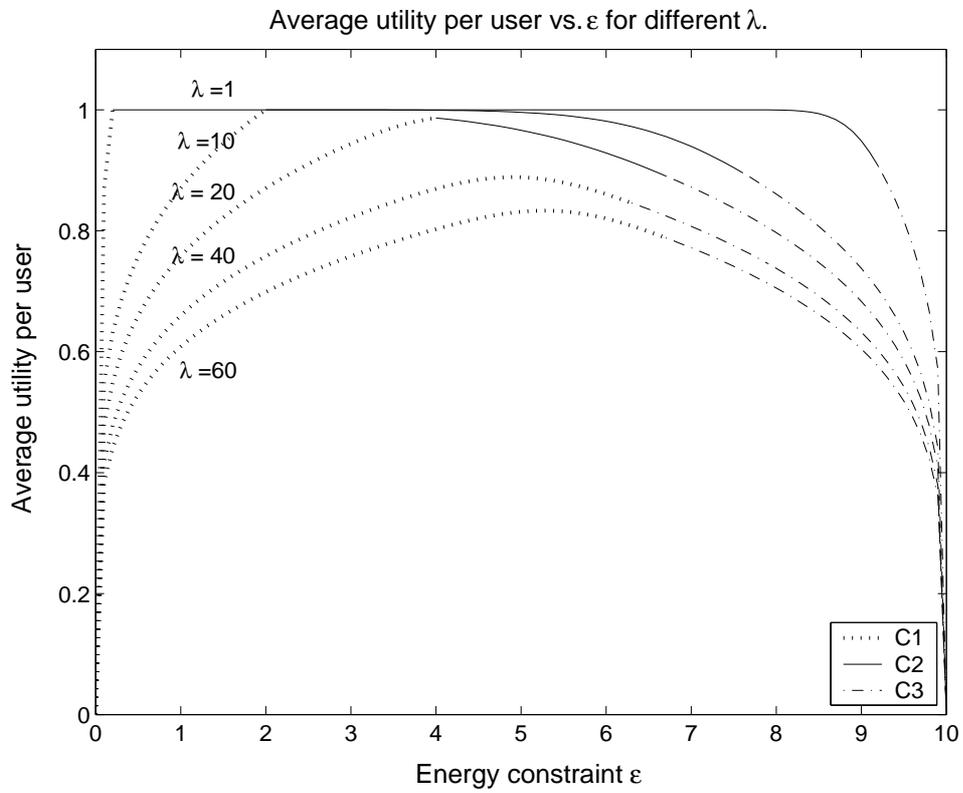


Fig. 5. Average utility per user vs. \mathcal{E} with different arrival rates λ .

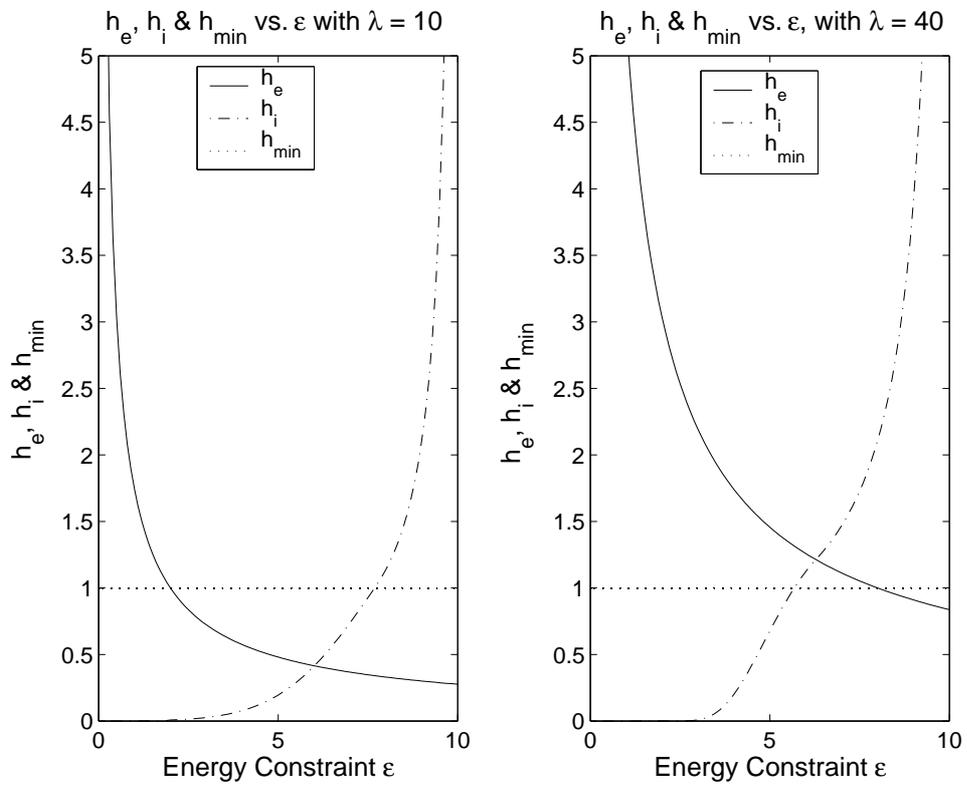


Fig. 6. Channel thresholds h_e , h_i vs. \mathcal{E} for different λ 's.

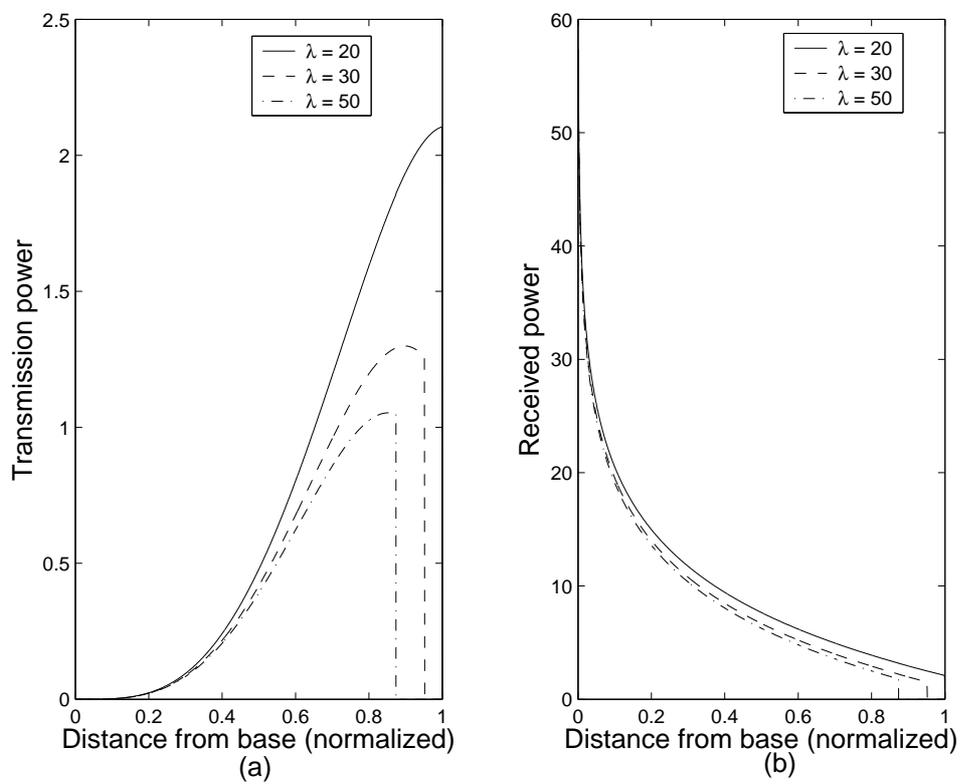


Fig. 7. Optimal power allocation: (a) Transmission power vs. distance from the base station; (b) Received power vs. distance from the base station.

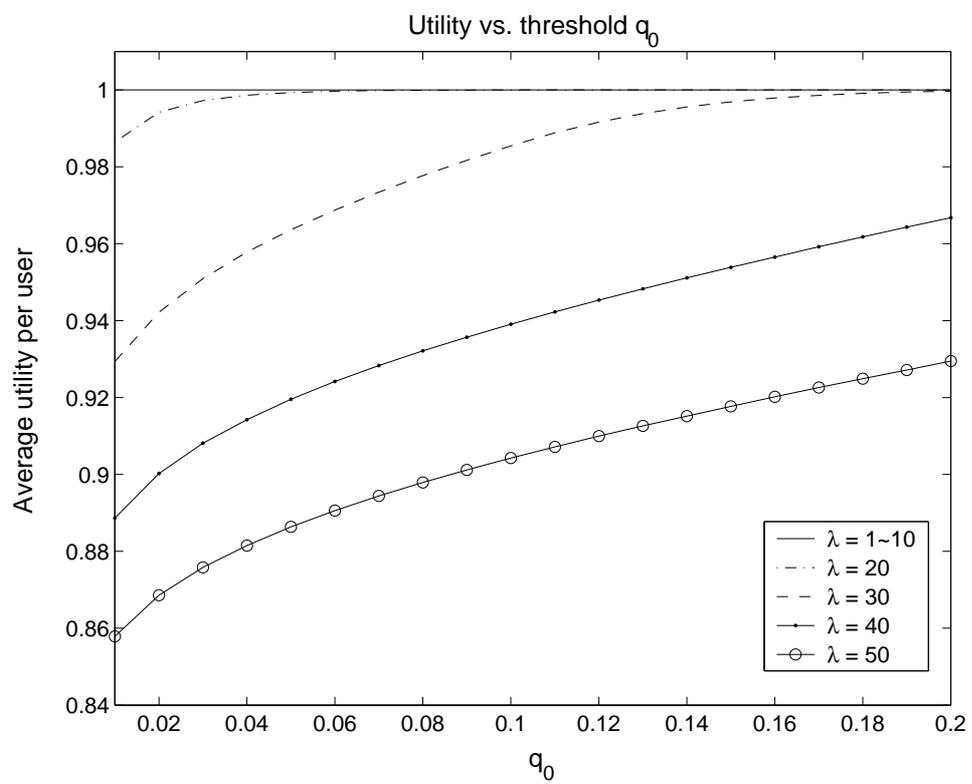


Fig. 8. Maximum average utility per user, $\mathbb{E}_H\{\tilde{U}_H[P(H)]\}$, versus the target outage probability q_0 for different arrival rates, λ .