Forty-Fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
Sept 27-29, 2006

Resource Allocation and Quality of Service Evaluation for Wireless Communication Systems using Fluid Models

Lingjia Liu, Parimal Parag, Jia Tang, Wei-Yu Chen, and Jean-François Chamberland

Abstract— **Wireless systems offer a unique mixture of connectivity, flexibility, and freedom. It is therefore not surprising that wireless technology is being embraced with increasing vigor. For real-time applications, user satisfaction is closely linked to quantities such as queue length, packet loss probability, and delay. System performance is therefore related to, not only Shannon capacity, but also Quality of Service (QoS) requirements. This work studies the problem of resource allocation in the context of stringent QoS constraints. The joint impact of spectral bandwidth, power, and coderate is considered. Analytical expressions for the probability of buffer overflow, its associated exponential decay rate, and the effective capacity are obtained. Fundamental performance limits for Markov wireless channel models are identified. It is found that, even with an unlimited power and spectral bandwidth budget, only a finite arrival rate can be supported for a QoS constraint defined in terms of exponential decay rate.**

I. INTRODUCTION

Motivated by the emergence of wireless technologies and by the constantly increasing demand for connectivity, we study the interplay between resource allocation at the physical layer and quality of service (QoS) in wireless communication systems. Radio resources typical of wireless communication systems include spectral bandwidth and power. Much research in recent years has been devoted to developing techniques and strategies that enhance the spectral efficiency of wireless systems. The prevalent framework used to evaluate these techniques is information theory, with the indissociable concept of Shannon capacity. While this framework is suitable for an analysis of maximum throughput, it overlooks many aspects of user satisfaction. Queue length distribution, packet loss probability, and delay all influence the perceived quality of a communication link. Quality of service has been studied extensively in the context of wired networks [1], [2]. This research led to the unifying concept of effective bandwidth, which has been applied to admission control and pricing. Effective bandwidth and the related notion of effective capacity provide a framework to identify the statistical characteristics of a stochastic process over various time scales. The literature on effective bandwidth is rich. Comprehensive discussions on the subject and its applications are provided by Kelly [3] and Chang [2].

User satisfaction and quality of service also play an important role in the design of wireless systems, especially for mobile terminals that support real-time applications. Unlike its wired counterpart, a wireless connection is subject to attenuation and fading. The time-varying nature of a wireless link will affect the queue length distribution at the terminal and, consequently, it will have a significant impact on performance. The allocation of system resources for real-time applications is therefore critically important and demands a careful analysis. Performance and user satisfaction may not be captured adequately by the sole attribute of Shannon capacity. The goal of this paper is to relate power and spectral bandwidth to quality of service using an evaluation framework akin to effective bandwidth. Two elements are needed to achieve this goal: a meaningful performance metric and a wireless channel model. In [4], [5], Wu and Negi extend the concept of effective bandwidth to effective capacity, which can be viewed as the dual of the effective bandwidth. It characterizes the maximum arrival rate that a wireless system can sustain subject to a given QoS requirement. Much like its precursor, effective capacity is a useful tool to identify system limitations as a function of statistical delay-bound violation probabilities.

To relate radio resources to system performance and statistical QoS requirements, we link the behavior of the system to its physical-layer infrastructure. A mobile terminal and its associated wireless connection can be modeled as a single-server queue, provided that the receiver has the ability to acknowledge reception of the data. For example, a simple physical-layer automatic repeat request (ARQ) mechanism may be incorporated in the communication protocol to insure that erroneous data is retransmitted. We assume that such a mechanism is in place throughout. Drawing intuition from information theory and error-control coding, the service offered to a mobile terminal can be modeled as a Markov-modulated fluid process. Previous results on Markov-modulated fluid processes and large deviations can therefore be leveraged to characterize the interplay between system resources at the physical layer and the statistical behavior of queues. In particular, we derive closed-form expressions for the effective capacity of time-correlated channels in terms of spectral bandwidth, power, and coderate. These results are new and lead to meaningful guidelines for the design of wireless systems.

The remainder of this paper is as follows. In Section II, we describe the generic wireless connection that is used as an abstraction for the physical layer. Based on a Markov assumption, we then construct a mathematical representation for the overall channel behavior. Section III contains a derivation of the equilibrium distribution for a system with a constant arrival rate and a Gilbert-Elliott wireless channel. Buffer overflow probabilities, the corresponding large deviations, and the effective capacity function are given explicit

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA

expressions in terms of physical system parameters. This analysis is subsequently extended to a variable data source. The performance analysis of the Gilbert-Elliott queueing system is presented in Section IV. Finally, we give our conclusions in the last section.

II. WIRELESS CHANNEL

We assume that the wireless channel, denoted by $h(t)$, is well-modeled as a zero-mean, proper complex Gaussian process. The envelope process $|h(t)|$ and the phase process are independent, with $|h(t)|$ having a Rayleigh probability distribution function and the phase being uniform over $[0, 2\pi)$. While it is straightforward to describe the first order statistics of $h(t)$, a complete characterization of this random process requires that joint distributions be specified as well. Under a Gaussian process model, it suffices to describe the correlation between any two sample points of the process. For the sake of mathematical tractability, we adopt a slightly simplified channel model. We retain the first order statistics of the channel and assume that the marginal distribution of the envelope process is Rayleigh. Second, we assume that for a fixed threshold η the probability of $|h(t)|$ being above or below this threshold is accurately modeled as a continuous-time Markov chain. We refer to the channel envelope exceeding η as the "ON" state; otherwise the channel is in its "OFF" state. Such a channel structure is commonly referred to as the Gilbert-Elliott model. It is assumed to provide a sufficiently accurate representation for the autocorrelation function of the quantized Rayleigh channel. This quantized channel model appears in Figure 1. The transition rate from OFF to ON is

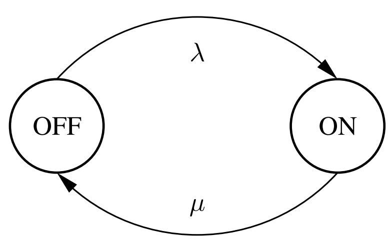

Fig. 1. Continuous-time Gilbert-Elliott Markov representation of a wireless communication link.

denoted by λ; while the transition rate from ON to OFF, by μ. The generator matrix for this Markov chain is given by

$$\begin{aligned} Q_s &= \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \\ &= \frac{1}{\lambda+\mu} \begin{bmatrix} 1 & \lambda \\ 1 & -\mu \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -(\lambda+\mu) \end{bmatrix} \begin{bmatrix} \mu & \lambda \\ 1 & -1 \end{bmatrix}. \end{aligned} \tag{1}$$

For consistency, the stationary distribution of the Markov chain should agree with the marginal distribution of the underlying channel,

$$\begin{aligned} \Pr\{|h(t)| \le \eta\} &= \frac{\mu}{\lambda+\mu} = \int_0^{\eta} 2\xi e^{-\xi^2} d\xi \\ &= 1 - e^{-\eta^2} = 1 - \Pr\{|h(t)| > \eta\} \end{aligned} \tag{2}$$

where $f(\xi) = 2\xi e^{-\xi^2}$ with $\xi \ge 0$ is the marginal distribution of the normalized envelope process. To solve for λ and μ, two equations are needed. The first one is given by condition (2). The second condition comes from the Markov structure of the wireless link. This latter condition captures the exponential decay rate of the correlation between samples. For a time interval t, the transition probability matrix $P_t(t)$ associated with the Gilbert-Elliott channel is given by

$$\begin{aligned} P_t(t) &= e^{Q_s t} \\ &= \frac{1}{\lambda+\mu} \begin{bmatrix} \mu + \lambda e^{-t(\lambda+\mu)} & \lambda - \lambda e^{-t(\lambda+\mu)} \\ \mu - \mu e^{-t(\lambda+\mu)} & \lambda + \mu e^{-t(\lambda+\mu)} \end{bmatrix}. \end{aligned}$$

If the Markov channel has an exponential decay rate κ, this implies that $\lambda + \mu = \kappa$ and

$$\lambda = \kappa e^{-\eta^2}, \quad \mu = \kappa - \kappa e^{-\eta^2}.$$

The quantized channel and its associated Markov structure will prove instrumental in computing the probability of buffer overflow and the effective capacity of the associated wireless connection.

A. Coding and Information Theory

A celebrated result from information theory is the Shannon capacity of the Gaussian channel,

$$C = W \log_2 (1 + P/N_0 W) \quad \text{bits per second.}$$

The variable P represents the power of the signal, $N_0/2$ denotes the power spectral density of the noise process, and W is the channel bandwidth. In theory, error-free communication can be achieved on this channel for any rate below the capacity using asymptotically long codewords. Today, there exists a collection of practical codes that operate close to capacity, with minimal error-rates and small delays. The Shannon capacity can therefore be employed as an optimistic approximation of code performance. If a code is designed to operate at a rate R, the sent information is decoded reliably whenever $R < C$; it is lost otherwise.

A similar performance description can be employed for time-varying channels. Suppose that a wireless channel varies slowly, data is assumed to reach its destination provided that $R < C(t)$, where

$$C(t) = W \log_2 \left(1 + \frac{P|h(t)|^2}{N_0 W} \right) \quad \text{bits per second} \tag{3}$$

is the instantaneous capacity. On the other hand, if $R \ge C(t)$ then information is lost. This simplified characterization is valid provided that the decoding delay is small compared to the coherence time of the wireless channel. The state of the Gilbert-Elliott channel is related to the instantaneous capacity and the coderate as follows. Let coderate R be given. The Gilbert-Elliott channel is ON if $R < C(t)$ or, equivalently,

$$|h(t)| > \eta = \sqrt{\frac{N_0 W}{P} \left(2^{\frac{R}{W}} - 1 \right)}. \tag{4}$$

It is OFF otherwise. The corresponding service rate is zero when the channel is OFF and R when the channel is in its ON state. Under these assumptions, the maximum throughput of this wireless channel is immediately seen to equal $R \Pr\{|h(t)| > \eta\} = Re^{-\eta^2}$.

III. QUEUEING PERFORMANCE OF MARKOV-MODULATED PROCESSES

In a wireless system, much like in a broadband network, the usage of system resources may not be well assessed by the single value of throughput or Shannon capacity. Requirements on the queue length, the packet loss probability, and the delay may force a wireless system to operate much below its theoretical Shannon capacity. Furthermore, the unreliable link-quality intrinsic to wireless communication along with stringent delay and loss constraints may significantly alter the optimal allocation of system resources.

Consider a simple queueing system with a single queue and one server. Let $a(t)$ denote the arrival rate, and let $s(t)$ be the offered service rate. The cumulative arrival function over interval $[0,t]$ is given by

$$A[0,t] = \int_{[0,t]} a(\tau) d\tau.$$

Similarly, the amount of service offered in the interval $[0,t]$ is equal to

$$S[0,t] = \int_{[0,t]} s(\tau) d\tau.$$

Under a work-conserving policy and provided that the queue is initially empty, the state of the queue is governed by the following equation [6],

$$L_t = (A[0,t] - S[0,t]) - \inf_{0<\tau<t} \{A[0,\tau] - S[0,\tau]\}.$$

This generic model provides an appropriate framework for evaluating the performance of a queueing system subject to QoS constraints.

A natural choice for the queueing representation of the communication system introduced in the previous section is a Markov-modulated fluid process. Let Q be the generator matrix of the underlying finite-state Markov process with state space $\{1,\ldots,M\}$. The state m is associated with a rate d_m, which represents the difference between the instantaneous arrival rate and the instantaneous service rate. Hence, the net rate of change in the buffer while in state m is d_m when the buffer is not empty and $\max\{0, d_m\}$ when the buffer is empty. We denote the diagonal matrix $\mathrm{diag}(d_1,\ldots,d_M)$ by D.

If we use L_t to denote the level of fluid in the buffer at time t and we let u_t be the state of the underlying Markov chain at time t, then (L_t, u_t) forms a continuous-state Markov process. Define the event probability

$$F(x,m,t) = \Pr\{u_t = m, L_t \le x\}.$$

Using this notation, the Chapman-Kolmogorov equation can be written as

$$\frac{\partial F}{\partial t} = FQ - \frac{\partial F}{\partial x} D \tag{5}$$

where $F = (F(x,1,t),\ldots,F(x,M,t))$. The equilibrium distribution $F(x,m)$ of the continuous state Markov process (L_t,u_t) is subject to

$$FQ = \frac{\partial F}{\partial x} D. \tag{6}$$

Let the stationary probability distribution of the underlying Markov chain be denoted by w, with $wQ = 0$. Then, $\lim_{x\to\infty} F(x,m) = w_m$. Since the equilibrium distribution is a bounded solution to (6), it has the spectral representation

$$F(x,\cdot) = w - \sum_{i=1}^{k} a_i \phi_i e^{x z_i} \tag{7}$$

where $\{(\phi_i, z_i)\}$ are the stable eigenvector-eigenvalue pairs of the eigenvalue problem

$$\phi Q = z \phi D. \tag{8}$$

If Q is reversible, then all such eigenvalues are real numbers [7]. Moreover, there are $k = |\{m : d_m > 0\}|$ strictly negative eigenvalues (counting multiplicity). These values are the ones included in (7). The coefficients $\{a_i\}$ are found using the boundary conditions

$$F(0,m) = 0 \quad \forall \{m : d_m > 0\}.$$

The unique solution to this boundary value problem is the equilibrium distribution.

A. Equilibrium Distribution of Gilbert-Elliott Systems

Consider a communication system where data arrives in a buffer at a constant rate $a(t) = a$. Suppose that this buffer is serviced through a wireless connection at a rate $s(t)$, where $s(t)$ is the Markov-modulated process described in Section II. We assume that the generator matrix of the underlying finite-state Markov chain is the matrix Q_s obtained in (1). The net arrival rate in the buffer when the buffer is not empty is $D = \mathrm{diag}(a, a-s)$ where s is the service rate when the channel is ON. The eigenvalue problem $\phi Q = z\phi D$ has two solution pairs:

$$(w,0) = \left(\left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu} \right), 0 \right)$$
$$(\phi, z) = \left((s-a, a), \frac{a\lambda + a\mu - s\lambda}{a(s-a)} \right).$$

The queue will be stable provided that

$$a < \frac{\lambda}{\lambda+\mu} s. \tag{9}$$

Under this condition, L_t converges in distribution to a finite random variable L. Using the boundary condition $F(0,1) = 0$, we obtain the equilibrium solution

$$\begin{aligned} F(x,\cdot) &= \left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu} \right) \\ &- \left(\frac{\mu}{\lambda+\mu}, \frac{a}{s-a} \frac{\mu}{\lambda+\mu} \right) \exp\left(\frac{a\lambda + a\mu - s\lambda}{a(s-a)} x \right). \end{aligned} \tag{10}$$

Based on this equilibrium distribution, we can compute a number of performance metrics including the probability of buffer overflow, its exponential rate of convergence to zero, and the effective capacity of the system.

The probability of buffer overflow is an important performance metric. For the Gilbert-Elliott system at hand, the probability of the buffer exceeding a threshold x is given by

$$\begin{aligned}\Pr\{L > x\} &= 1 - \langle F(x, \cdot), (1, 1)\rangle \\ &= \frac{s}{s-a}\frac{\mu}{\lambda+\mu}\exp\left(\frac{a\lambda + a\mu - s\lambda}{a(s-a)}x\right) \\ &= \frac{R}{R-a}\left(1 - e^{-\eta^2}\right)\exp\left(\frac{\kappa a - R\kappa e^{-\eta^2}}{a(R-a)}x\right).\end{aligned}$$

As seen in (10), the eigenvalue problem $\phi Q = z\phi D$ applied to the present two-state system contains only one negative solution. The large deviation principle governing the probability of buffer overflow is therefore immediately seen to equal

$$\begin{aligned}-\lim_{x\to\infty}\frac{\log\Pr\{L > x\}}{x} &= -\frac{a\lambda + a\mu - s\lambda}{a(s-a)} \\ &= -\frac{\kappa a - R\kappa e^{-\eta^2}}{a(R-a)}.\end{aligned}$$

The large deviation principle governing the distribution of a queue is sometimes preferable as a design criteria because it admits a simpler form.

The effective capacity is the dual concept of effective bandwidth [4]. Given specific system parameters and an exponential decay rate θ, the effective capacity is the maximum arrival rate for which the QoS requirement θ is fulfilled. Mathematically, this can be expressed as

$$\alpha(\theta) = \sup\left\{a \geq 0 : -\lim_{x\to\infty}\frac{\log\Pr\{L > x\}}{x} \geq \theta\right\}. \quad (11)$$

Taking into account condition (9), an explicit formula for the effective capacity of the Gilbert-Elliott channel is given by

$$\alpha(\theta) = \frac{\theta R + \kappa - \sqrt{(\theta R + \kappa)^2 - 4\theta R\kappa e^{-\eta^2}}}{2\theta}. \quad (12)$$

B. On-Off Information Sources

Some traffic sources are better modeled as on-off sources. Voice, for instance, is a good example of an information process that can be accurately modeled as an on-off source. When two people are carrying a conversation, they are unlikely to speak simultaneously. On average, a person involved in a discussion speaks at most half of the time. Other data sources such as instant messaging applications and wireless sensors can also be modeled as on-off sources. As such, we extend the analysis of the previous section to the case where the data source features an on-off behavior.

Suppose that the data arrives in the buffer at a rate $a(t)$, where $a(t)$ is a two-state Markov-modulated source. We assume that the arrival rate is equal to $a > 0$ when the source is ON; it is equal to zero otherwise. The generator matrix of the underlying Markov chain for this arrival process can be written as

$$Q_a = \begin{bmatrix} -\lambda_a & \lambda_a \\ \mu_a & -\mu_a \end{bmatrix}.$$

Again, we assume that the service offered through the wireless channel is a Markov-modulated process with generator matrix Q_s, as defined in (1). The aggregate system is therefore a stochastic fluid process modulated by a four-state Markov chain. The evolution of the buffer content is governed by (5), where the generator matrix Q is equal to

$$\begin{aligned}Q &= \begin{bmatrix} -\lambda_a - \lambda & \lambda_a & \lambda & 0 \\ \mu_a & -\mu_a - \lambda & 0 & \lambda \\ \mu & 0 & -\lambda_a - \mu & \lambda_a \\ 0 & \mu & \mu_a & -\mu_a - \mu \end{bmatrix} \\ &= Q_s \otimes I + I \otimes Q_a.\end{aligned}$$

Throughout, we use $A \otimes B$ to denote the Kronecker product of matrices A and B. The net arrival rate in the buffer is represented by $D = -sE \otimes I + aI \otimes E$, where the matrix E is defined by

$$E = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

We assume that the system is stable; that is, the following condition is satisfied

$$\frac{\lambda_a}{\lambda_a + \mu_a}a < \frac{\lambda}{\lambda + \mu}s.$$

The equilibrium distribution of this system is governed by (6), and its spectral representation follows the general form of (7).

This problem can be solved using techniques similar to the ones employed for a constant source. Consider an exponential decay rate requirement of $\theta > 0$ on the probability of buffer overflow,

$$-\lim_{x\to\infty}\frac{\log\Pr\{L > x\}}{x} \geq \theta. \quad (13)$$

This requirement will be fulfilled if and only if the value of v corresponding to the largest negative eigenvalue of $\phi_v = (s-v, v)\otimes(a-v, v)$ is less than $\alpha(\theta)$ but greater than $\beta(\theta)$, where $\alpha(\theta)$ is the effective capacity introduced earlier

$$\begin{aligned}\alpha(\theta) &= \sup\left\{\nu \geq 0 : \frac{\nu\lambda + \nu\mu - s\lambda}{\nu(s-\nu)} \leq -\theta\right\} \\ &= \sup\left\{\nu \geq 0 : \theta\nu^2 - (\theta s + \lambda + \mu)\nu + s\lambda \geq 0\right\}\end{aligned}$$

and $\beta(\theta)$ is the effective bandwidth of a two-state Markov-modulated fluid source [8], [1], [9]

$$\begin{aligned}\beta(\theta) &= \inf\left\{\nu \geq 0 : \frac{a\lambda_a - \nu\lambda_a - \nu\mu_a}{\nu(a-\nu)} \leq -\theta\right\} \\ &= \inf\left\{\nu \geq 0 : \theta\nu^2 - (\theta a - \lambda_a - \mu_a)\nu - a\lambda_a \geq 0\right\}.\end{aligned}$$

Clearly, the QoS requirement of (13) can only be met if $\alpha(\theta) \geq \beta(\theta)$. A standard buffer decoupling argument shows that this inequality is, in fact, a necessary and sufficient condition for (13) to hold. That is, the exponential decay rate requirement $\theta > 0$ will be satisfied if and only if $\alpha(\theta) \geq \beta(\theta)$. This observation greatly facilitates the performance analysis contained in the next section. For a QoS requirement such as (13), an on-off source shares the same service needs as a constant source with rate $\beta(\theta)$. Thus, for a given $\theta > 0$, the allocation of system resources can be studied in terms of fixed arrival rates, whether the source rate is a constant or a Markov modulated fluid model.

IV. Performance Analysis of Gilbert-Elliott Systems

We proceed to analyze the performance of the Gilbert-Elliott system as a function of physical resources. Following the literature on effective bandwidth, we use the exponential decay rate as our primary performance measure. Figure 2 shows the maximal supported arrival rate $\alpha(\theta)$ as a function of the QoS constraint θ for the Markov decay parameters $\kappa \in \{10^2, 10^3, 10^4\}$. This figure also includes the optimal coderate R as a function of θ. Not surprisingly, more stringent QoS constraints result in lower effective capacities for fixed system parameters. This is intuitive since a lower arrival rate reduces the expected queue length. More importantly, we see that the optimal coderate R is also a function of the QoS requirements. Under strict QoS constraints, an error-control code with a lower rate performs better as it reduces the probability of the channel being OFF. This analysis provides a new and systematic way to select the coderate as a function of the channel profile and the QoS requirements of a specific system.

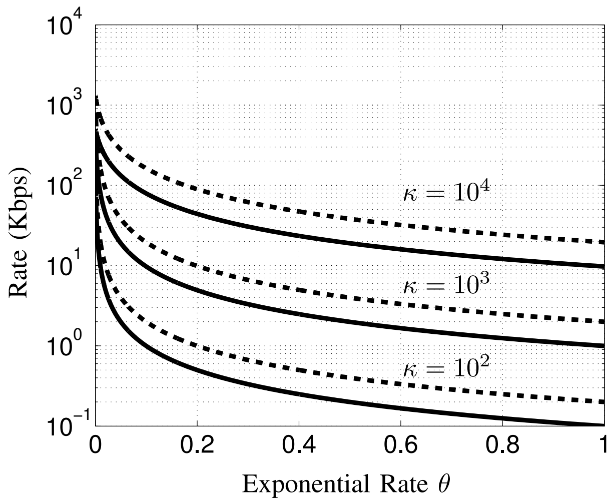

Fig. 2. Optimal coderate (dashed) and effective capacity (solid) as a function of exponential decay rate θ for various Markov decay parameters $\kappa \in \{10^2, 10^3, 10^4\}$.

It is interesting to note that the maximum throughput and the corresponding coderate are independent of the Markov decay parameter κ,

$$\lim_{\theta \to 0} \alpha(\theta) = \lim_{\theta \to 0} \frac{\theta s + \lambda + \mu - \sqrt{(\theta s + \lambda + \mu)^2 - 4\theta s \lambda}}{2\theta}$$
$$= \lim_{\theta \to 0} \frac{\theta R + \kappa - \sqrt{(\theta R + \kappa)^2 - 4\theta R \kappa e^{-\eta^2}}}{2\theta}$$
$$= R e^{-\eta^2}.$$

However, the effective capacity for $\theta > 0$ depends heavily on the statistical profile of the channel. Correlation impairs effective capacity. The higher the correlation coefficient, the lower the effective capacity. In other words, a throughput analysis of this system is not sufficient to provide an accurate assessment of supported rates under strict QoS constraints. This also implies that the common assumption that channel realizations are independent and identically distributed through time may lead to over-optimistic performance predictions on effective capacity.

The effective capacity shown in Figure 2 decays rapidly as a function of θ. It is therefore of interest to look at the reverse problem: for a given arrival rate a, we wish to characterize the amount of physical resources necessary to meet a prescribed QoS constraint $\theta > 0$. First, we note that a necessary condition for a solution to exist is the stability criterion $a < Re^{-\eta^2}$. However, this condition may not be sufficient. A solution exists if and only if we can find a power P and a bandwidth W such that $\theta a^2 - (\theta R + \kappa) a + R\kappa e^{-\eta^2} = 0$. This equation can be rearranged as

$$\eta^2 = \frac{N_0 W}{P}\left(2^{\frac{R}{W}} - 1\right)$$
$$= -\log\left(\frac{\theta R a + \kappa a - \theta a^2}{R\kappa}\right).$$

Since $\eta^2 > 0$ the following inequality must apply,

$$0 < \theta R a + \kappa a - \theta a^2 < R\kappa.$$

A necessary and sufficient condition for a solution to exist is $\theta < \frac{\kappa}{a}$. We emphasize that, even with an unlimited power and spectral bandwidth budget, only a finite arrival rate can be supported for a QoS constraint $\theta > 0$. Furthermore, this bound is independent of the actual coderate R used in the system. This fact is in sharp contrast with Shannon capacity, which goes to infinity as power and spectral bandwidth grow unbounded. This limitation is partly due to the fact that, in the system under study, the transmitter has no knowledge of the channel gain. Thus, it cannot transmit at the (error-free) instantaneous channel capacity. Without channel state information, the best decay rate θ is limited by the ratio of κ to the arrival rate a.

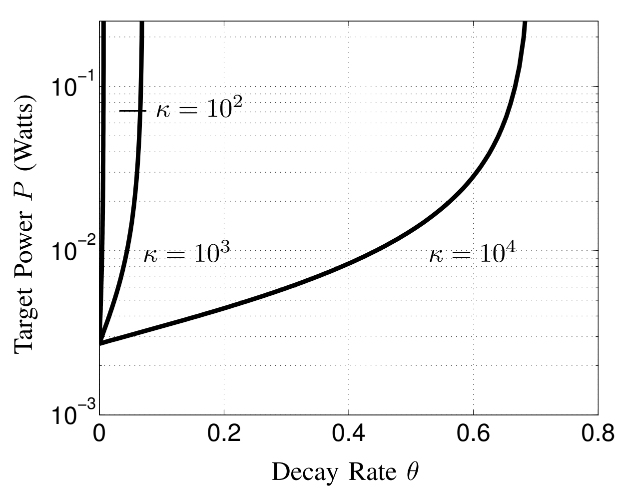

Fig. 3. Signal power as a function of exponential decay rate θ for various Markov decay parameters $\kappa \in \{10^2, 10^3, 10^4\}$.

Figure 3 shows the target power P as a function of the QoS constraint θ for an arrival rate $a = 14.4$ Kbps. Note that power P can be obtained in closed form as

$$P = -\frac{N_0 W \left(2^{\frac{R}{W}} - 1\right)}{\left(\log\left(\theta R a + \kappa a - \theta a^2\right) - \log\left(R\kappa\right)\right)}.$$

As expected, the required power goes to infinity for a finite θ. The set of supported exponential decay rates is intimately connected to the Markov decay factor κ. Not only does correlation decrease effective capacity, it also limits the rates and qualities of service that can be sustained on a given wireless channel. Figure 4 shows minimum spectral bandwidth as a function of decay rate θ for an arrival rate of $a = 14.4$ Kbps. Again, the amount of resource required goes to infinity for a finite θ.

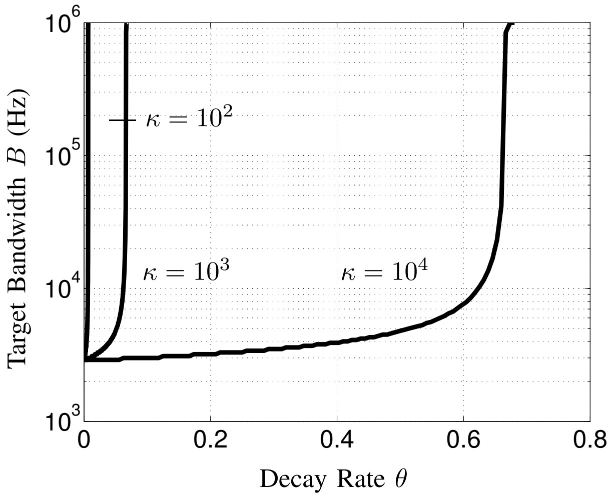

Fig. 4. Spectral bandwidth as a function of exponential decay rate θ for various Markov decay parameters $\kappa \in \{10^2, 10^3, 10^4\}$.

The speech process of an interlocutor involved in an English conversation can be modeled as an on-off information source [10]. Exponentially distributed talk spurts with a mean duration of $\mu_a^{-1} \approx 352$ ms are followed by silent periods with mean $\lambda_a^{-1} \approx 650$ ms. Using advanced signal processing techniques, active speech can be compressed to a rate of 14.4 Kbps. The average throughput of an encoded speech process is therefore 5.06 Kbps. However, for a delay constraint of 20 ms or an approximately equivalent QoS constraint $\theta = 0.5$, the effective bandwidth of speech is essentially equal to 14.4 Kbps. The very strict delay constraint imposed on speech traffic forces the effective bandwidth to be nearly equal to its peak rate, which is much higher than the average throughput. As seen on Figure 3, voice traffic cannot be successfully transmitted over highly correlated channels without sophisticated power control. This partly explains why power control is critical to cellular telephony.

V. CONCLUSIONS

We considered the allocation of system resources in the context of wireless communications under QoS constraints. A Markov model was introduced to capture the unreliable nature of wireless systems. For a given error correcting code, the behavior of the overall wireless connection is assumed equivalent to a continuous-time Markov chain. While throughput represents a major component of user satisfaction; queue length distribution, packet loss probability, and delay are also important factors that influence the quality of service perceived by the users. One of the main QoS measures present in the literature is the large deviation principle governing the probability of buffer overflow,

$$\lim_{x \to \infty} \frac{\left(\log \Pr\{L > x\}\right)}{x}.$$

This definition of quality of service was used to conduct a performance analysis of the Gilbert-Elliott system as a function of physical resources. The effective capacity is found to decay sharply as a function of QoS constraint $\theta > 0$. Furthermore, the optimal code selection for a wireless system depends on its QoS requirements. A more stringent constraint on θ lowers the optimal coderate R.

Correlation is also found to have a major impact on performance. The effective capacity of a slowly varying channel can be very small. For communications under QoS constraints, the popular assumption of independent and identically distributed channel realizations results in an over-optimistic assessment of system performance. The impact of correlation on system performance is perhaps best exemplified by the fact that arrival rate a can only be supported if $\theta < \frac{\kappa}{a}$. That is, even with unlimited amount of physical resources, the maximum arrival rate supported under QoS constraint θ is bounded.

.

REFERENCES

[1] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," *IEEE/ACM Trans. on Networking*, vol. 1, no. 3, pp. 329–343, June 1993.

[2] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. on Automatic Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[3] F. P. Kelly, S. Zachary, and I. Ziedins, *Stochastic Networks: Theory and Applications*, ser. Royal Statistical Society Lecture Note Series. Oxford University Press, 1996.

[4] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of services," *IEEE Trans. on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003.

[5] ——, "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *IEEE Trans. on Vehicular Technology*, vol. 54, no. 3, pp. 1198–1206, May 2005.

[6] C.-S. Chang, *Performance Guarantees in Communication Networks*, ser. Telecommunication Networks and Computer Systems. Springer, 1995.

[7] F. P. Kelly, *Reversibility and Stochastic Networks*. Wiley, 1979.

[8] G. Kesidis, J. Walrand, and C.-S. Chang, "Effective bandwidths for multiclass markov fluids and other ATM sources," *IEEE/ACM Trans. on Networking*, vol. 1, no. 4, pp. 424–428, August 1993.

[9] M. M. Krunz and J. G. Kim, "Fluid analysis of delay and packet discard performance for QoS support in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 384–395, February 2001.

[10] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 856–868, September 1986.

[11] Q. Zhang and S. A. Kassam, “Finite-state markov model for rayleigh fading channels,” *IEEE Trans. on Communications*, vol. 47, no. 11, pp. 1688–1692, November 1999.

[12] C.-D. Iskander and P. T. Mathiopoulos, “Analytical level crossing rates and average fade durations for diversity techniques in nakagami fading channels,” *IEEE Trans. on Communications*, vol. 50, no. 8, pp. 1301–1309, August 2002.

[13] ——, “Fast simulation of diversity nakagami fading channels using finite-state markov models,” *IEEE Transactions on Broadcasting*, vol. 49, no. 3, pp. 269–277, September 2003.

[14] H. S. Wang and N. Moayeri, “Finite-state markov channel – A useful model for radio communication channels,” *IEEE Trans. on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, February 1995.

[15] W. Turin and R. van Nobelen, “Hidden markov modeling of flat fading channels,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 9, pp. 1809–1817, December 1998.

[16] D. Mitra, “Stochastic theory of a fluid model of producers and consumers coupled by a buffer,” *Advances of Applied Probability*, vol. 20, pp. 646–676, September 1988.

[17] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communications aspects,” *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998.

[18] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer, 1996.

[19] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., ser. Universitext. Springer, 2003.

[20] J. G. Kim and M. M. Krunz, “Bandwidth allocation in wireless networks with guaranteed packet-loss performance,” *IEEE/ACM Trans. on Networking*, vol. 8, no. 3, pp. 337–349, June 2000.