Centralized Wireless Data Systems with User Arrivals and Departures – Part II: Applications and Extensions

Rajat Prakash & Venugopal V. Veeravalli

Abstract

In a companion paper [1], we presented a powerful time-scale separation technique for the analysis of centralized wireless data systems. In this paper, we first apply our technique to an example TDMA system, and demonstrate the accuracy of the time-scale separation approximation for a wide range of parameter values via simulations. We then go on to illustrate the usefulness of our theoretical results in cross-layer design, by deriving a power control scheme that attains an optimal balance between the twin goals of reducing the service process variability and increasing the service process mean. We also use the time-scale separation approximations to study the role of fading correlation and channel variability on the performance of TDMA and scheduling systems.

The results that we presented in [1] required a number of somewhat restrictive conditions, and some of these conditions are relaxed in this paper. The negligible discreteness condition is relaxed to study the effect of service discreteness, and it is shown that discreteness causes a reduction in the effective service rate. The independent service condition is relaxed to extend the second order approximation to systems with dependent service (such as CDMA systems with channel correlation across time). A heuristic is given to relax the exponential filesize condition, and show that the second order approximation is unchanged under arbitrary filesize distributions. Finally, the user symmetry condition is relaxed for the special case of TDMA.

Address of corresponding author
V. V. Veeravalli
University of Illinois at Urbana-Champaign
106 CSL
1308 West Main Street
Urbana, IL 61801

Other contact info
Telephone: (217) 333-0144
Fax: (217) 333-1642
e-mail: vvv@uiuc.edu

R. Prakash was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. He is now with Qualcomm Inc, San Diego, CA 92121. V. V. Veeravalli is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana IL 61801; e-mail: rprakash@qualcomm.edu, vvv@uiuc.edu

This research was supported by by NSF under CCR-9980616, through a subcontract with Cornell university, and by the NSF CAREER/PECASE award CCR-0049089.

This paper was presented in parts at the 2002 IEEE International Symposium on Information Theory and the 2003 IEEE International Symposium on Information Theory.

Submitted: January 2004, IEEE Transactions on Information Theory

1
1 Introduction

In Part I of this two-part paper, we introduced a time-scale separation technique for the analysis of centralized wireless systems with user arrivals and departures. In such systems (which we refer to as dynamic-user systems), new users arrive into the system at random instants with a certain service requirement (filesize). Each user in the system is provided service by a centralized server (base station, access point), and the nature of this service is governed by the physical layer. We showed in [1] that although the exact analysis of dynamic-user systems is complicated, an approximate analysis can be performed in a regime where the time taken to transmit one file is large relative to the time-scale of variation of the channel. In particular, our first order approximation result showed that, when file sizes are large, random fluctuations in the service rate can be ignored and the service process can be replaced by one whose rate depends only on the number of users in the system. Our second order approximation result, which is applicable to a more restricted class of dynamic-user systems, incorporated the effects of service variability more precisely. We showed that variability in the service process reduces the effective service rate, thus quantifying the conventional heuristic that service variability degrades system performance.

This paper deals with applications and extensions of the results of Part I. We begin by applying the time-scale separation result to a TDMA system, and compare the first and second order approximations with simulation results. For the TDMA system, the accuracy of the first and second order approximations depends on the system parameters as follows. Large File sizes: The first and second order approximations are nearly identical, and are equally effective. Moderate file sizes, high fading correlation across time: The first order approximation begins to fail, while the second order approximation remains accurate. Moderate file sizes, small fading correlation across time: Both the the first and second order approximations begin to fail.

The reason for the inaccuracy of the second order approximation for the case of moderate file sizes and small fading correlation is due to the fact that the negligible discreteness condition [1] that is required for this approximation is not satisfied. To overcome this problem, we develop a discreteness correction term for the second order approximation. This correction term results in a reduction in the effective service rate, beyond the reduction caused by service stochasticity.

In addition to aiding in the performance analyses of a variety of dynamic-user systems, our approximations may also be used for system design. We considered some system design applications of the first order approximation previously in [2] and [3]. In [3], we introduced the idea of
traffic-load based power control algorithms that vary the transmit power as a function of the number of users in the system. In particular, for some types of dynamic-user systems, we showed that the transmit power should be increased as the number of users increases, and that such control of the transmit power can reduce the average power consumption without increasing the average delay.

In another application of time scale separation, we studied the tradeoff between mean delay and the offered load in [2]. We compared this tradeoff for different physical layer schemes, and introduced the notion of capacity of a dynamic-user system as the maximum offered load that maintains stability. We note that in [2] and [3] we only considered applications of the first order approximation. In this paper, we consider applications of the second order approximation.

First, we consider the problem of designing a power control rule to adjust the transmit power according to the fading level. In designing a power control algorithm, there are two conflicting objectives: maximization of the average throughput (desirable for non-real-time users), and minimization of service variability (desirable for real-time users). Water filling (see, e.g., [4, Chapter 3]) is known to maximize the average throughput at the expense of high service variability, while channel inversion is known to minimize service variability at the expense of low average throughput. The second order approximation suggests that the right balance between maximizing the mean service rate and minimizing the service variance is obtained by maximizing an effective service rate, which is a function of both the mean and the variance of the service process. We give a numerical example to show that the power control that maximizes the effective service rate results in a lower mean delay than that attained by simply maximizing the mean service rate or minimizing the service variance. We show that larger service mean should be favored over smaller service variability when the filesizes are large, and vice-versa when the filesizes are small.

The following clarification about the difference between service variability and channel variability is in order here. The second order approximation says that for two systems with the same mean service rate, the system with lower service variability will perform better, leading to the conclusion that service variability is undesirable. This, however, does not imply that channel variability is undesirable. For example, it is argued in [5], that channel variability can be used to improve performance by scheduling service to the user with the best channel, and thereby increasing the mean service rate. We show that for highly or moderately loaded systems, channel variability is indeed helpful, and improves overall performance. In contrast, for lightly loaded systems, channel variability is undesirable, because the improvements in the mean service rate are small, and come at
the cost of large service variability. Thus, we conclude that artificially induced channel fluctuations along the lines of [5] are useful in the high or moderate load regimes, but not in the low load regime.

In this paper, we also consider relaxing some of the conditions required for the first and second order approximations of [1]. In particular, we focus on relaxing the exponential filesize, independent service, stationary arrivals, and symmetric users conditions. For the sake of brevity, we do not consider relaxing the symmetric resource allocation condition in this paper, and instead refer the reader to [6].

For non-exponential filesize distributions, we show that while our proof of the first order approximation (given in [1]) remains valid, our proof of the second order approximation result breaks down. Nevertheless, we provide a simple heuristic to argue that the second order approximation result may remain valid.

To relax the independent service condition, we give a modified form of the second order approximation that is applicable to systems with dependent service (such as CDMA). With dependent service, we show that the effective service rate is lower than that suggested by the original second order approximation. This leads to the conclusion that dependent service causes a degradation in performance.

Finally, we outline a technique to relax the symmetric user condition. We show that a first order approximation can be attained in the special case of TDMA systems, but other types of systems appear to be intractable.

The rest of the paper is organized as follows. In Sections 2.1 and 2.2, the example TDMA model is introduced, and a survey of previous work on related models is given. Numerical results to verify the approximation results are given in Section 3. In the same section, a simple heuristic is developed to relax the negligible discreteness requirement of the second order approximation result. In Section 4, a power control application of the second order approximation is given, and in Section 4.2, the role of service variability is discussed in the context of scheduling.

The independent service condition for the second order approximation is relaxed in Section 5, and numerical results validating this relaxation are given in Section 5.1. The exponential filesize condition is relaxed in Section 6, and the symmetry condition is relaxed in Section 7. Finally, conclusions are given in Section 8.
2 Application to TDMA Systems

In this section, we define a dynamic-user TDMA system, obtain the first and second order approximations for its performance, and test the accuracy of these approximations using simulations. We begin by describing an information theoretic model for information transmission in a TDMA system.

2.1 Incremental Redundancy

To analyze a practical communication system, we have to account for the actual coding, modulation and power control schemes. However, to illustrate the use of time-scale separation, we adopt an information theoretic simplification. For each time-slot, we calculate the mutual information that is transmitted from the transmitter to the receiver, and consider a file transmission to be complete when the accumulated mutual information at the receiver exceeds the filesize. Thus, we approximate actual data bits by mutual information bits.

This information theoretic simplification approximates the performance of incremental redundancy (IR), which works as follows. Using a feedback link and an error detecting code, the receiver informs the transmitter whether decoding of a codeword was successful. In case of decoding failure, the transmitter sends extra parity bits to help the decoding process, and this process continues until decoding is successful.

IR was introduced by Chase [7] as a technique to improve the performance of traditional ARQ schemes for a single user system. Caire and Tuninetti [8] have recently analyzed many forms of IR for multiuser systems, under the assumption that the number of users stays fixed. There is, however, relatively little analysis of dynamic-user IR systems. For dynamic-user systems with a large number of users, a fixed-point based analysis is discussed in [9], but this work does not provide a way to compute the key quantity of interest, namely the steady-state distribution of the number of users in the system. For a certain special types of dynamic-user systems, [10] and [11] give an intractable Markov chain model, which has to be solved by simulation.

In contrast to these works, we show in the following that our time-scale separation approximations provides an easy way to analyze dynamic-user IR systems. We note that there have been previous applications of time-scale separation to dynamic-user IR systems, including our own work [2], [3], [12] and [13]. But these works do not focus on the accuracy of the time-scale separation approximation, and are restricted to the first order approximation (which they use without proof). In
the remainder of this Section, we describe an example TDMA system (with IR), and for this system, we examine the accuracy of both the first and second order approximations.

### 2.2 System Model

We are interested in dynamic-user systems, where the number of users evolves according to an arrival and departure process. We assume that arrivals occur at the points of a Poisson process with rate $\lambda$, and the filesize of each arriving user is independent and exponentially distributed with mean $\bar{S}$. Users are served by a central server (base station), and depart as soon as service is completed.

We consider a TDMA system with slot duration $\delta$ seconds, with each user transmitting once every $u$ slots. The channel fading for user $j$ at time $k$ is denoted by $h_{j,k}$, and $h_{j,k}$ follows a Gilbert-Elliot model with two states $g_0$ and $g_1$, where $g_0^2 + g_1^2 = 1$. Transitions between the two states are governed by the Markov model in Figure 2.2. Under this model, the channel state stays constant for an average of $G^{-1}$ seconds. We assume that the channel evolves independently for each user in the dynamic-user system, and that the channel state is known to both the receiver and the transmitter. Using this knowledge of the fading state, the transmitter allocates powers $P(g_0)$ and $P(g_1)$ in fading states $g_0$ and $g_1$ respectively. Then, assuming unit noise variance, the information transmitted in a slot with fading level $g$ is

$$C(g) = W\delta \log(1 + |g|^2 P(g)),$$

$$g = g_0, g_1.$$  \hspace{1cm} (2.1)

We assume that the power levels $P(g_0)$ and $P(g_1)$ are selected according to the water-filling formula in [4], that maximizes $E[C(h)]$, subject to $P(g_0) + P(g_1) \leq \text{SNR}$.

![Figure 1: Gilbert-Eliot fading model](image)

It is clear that for the service process defined above, a constant service rate model is not applicable. The system can, however, be described by a Markov chain by defining the state vector

$$X[k] = (u[k], TDM[k], h_{1,k}, \ldots, h_{u,k})$$  \hspace{1cm} (2.2)
where $TDM[k]$ is the index of the user to be served in slot $k$. Evolution of the channel state is governed by the Gilbert-Eliot model of Figure 2.2, and $TDM[k]$ evolves cyclically from 1 to $u$, with $u$ being incremented or decremented at times of arrival or departure.

For the system described above, the quantity of interest is the steady state probability $\pi(u)$ of $u$ users being present in the system. This probability can be computed if we can determine the steady state distribution of the Markov chain $X[k]$. Although the exact steady state analysis of $X[k]$ is not easy, we showed in Part I [1] that time-scale separation can be used to approximate $\pi(u)$ under certain conditions. In the next Section, we verify these conditions, and give the time-scale separation approximation for $\pi(u)$.

### 2.3 Applying Time-Scale Separation

**First Order Approximation:** The first order approximation is applicable under the separability condition of Definition 2, given in Section 3.3 of Part I [1]. To check the separability condition, we need to evaluate the mean service rate of a user in an appropriately constructed fixed-user system. Recall from Part I that the fixed-user system is constructed by restricting the number of users to $u$ (by barring arrivals, and setting the filesize of each user to infinity). The mean service rate can be computed from the steady state distribution $\eta_u(x)$ of the state vector $X[k]$ in the fixed-user system. From the definition of the TDMA system, it is easy to see that under the distribution $\eta_u$, the fading levels $h_{j,k}$ are independent across $j$, taking values $g_0$ and $g_1$ with equal probability, and that the service index $TDM$ is uniformly distributed between 1 and $u$. The mean service rate per user does not depend on the user index $j$, and is given by

$$\frac{\phi(u)}{u} = \frac{1}{2a}W[\log(1 + |g_0|^2P(g_0)) + \log(1 + |g_1|^2P(g_1))] + \log(1 + |g_1|^2P(g_1))] \quad (2.3)$$

**A Technical Conditions:** Apart from the user and resource allocation symmetry conditions, the first order approximation also requires the technical condition that the fixed-user system be aperiodic (see [1, Definition 2]). For the TDMA system of interest, the fixed-user Markov chain does not satisfy the aperiodicity condition due to periodicity of the state component $TDM[k]$. To meet the technical condition of aperiodicity, we modify TDMA slightly. Let the TDMA index be randomized at the points of a Poisson process with a small rate. For small enough rate of this Poisson process, randomization will make little difference to the service process, but will cause the technical condition of aperiodicity to be satisfied. The randomization argument above is for technical purposes only, and we use standard TDMA in our simulations.
Next, we verify the conditions required for the more precise second order approximation, and show that these conditions are satisfied by the TDMA system of interest. Further, we argue that channel correlation in time reduces the effective service rate.

**Second Order Approximation:** To apply the second order approximation of Theorem 2 of Section 3.4 given in Part I [1], we need to check the nice separability condition. The nice separability condition holds for TDMA, as discussed in Section 4 of Part I.

Also, we argued in [1, Section 3.4] that the second order approximation provides useful approximations only when the negligible discreteness condition is met. The negligible discreteness condition depends on the choice of parameters, and we will consider parameter values that result in a varying range of discreteness. For parameter values that do not satisfy the negligible discreteness condition, we will give a heuristic argument to modify the effective service rate later in Section 3.2.

To apply the second order approximation, we need to evaluate the service variance \( \sigma^2(u) \). The variance of the service provided in one slot is

\[
\sigma_{\text{slot}}^2 = \frac{1}{2} \left( C(g_0)^2 + C(g_1)^2 - \frac{1}{2} (C(g_0) + C(g_1))^2 \right). \tag{2.4}
\]

To apply the second order approximation, we are interested in the variance of the service process \( \tilde{I}_j(t) \) for user \( j \) in a fixed-user system with \( u \) users. It is shown in Appendix A that

\[
\sigma_j^2(u) \overset{\text{def}}{=} \lim_{t \to \infty} \frac{1}{t} \mathbb{E} \left[ \left( \tilde{I}_j(t) - t \frac{\phi(u)}{u} \right) \right] \approx \sigma_{\text{slot}}^2 \frac{1 + (1 - 2G\delta)^u}{u(1 - (1 - 2G\delta)^u)}. \tag{2.5}
\]

Using the independence of the fading across users, we can conclude that the correlation between the service processes of different users is zero. Then, the variance of the total service (summed across all users) \( \tilde{I}(t) \) is equal to the sum of the variances of the service \( \tilde{I}_j(t) \) of the \( u \) users in the system. This gives the effective service rate approximation:

\[
\phi^*(u) \approx \phi(u) - \frac{\sigma^2(u)}{2S} = \phi(u) - \sigma_{\text{slot}}^2 \frac{1 + (1 - 2G\delta)^u}{S\delta(1 - (1 - 2G\delta)^u)}. \tag{2.6}
\]

From the mean service rate in (2.3) and the effective service rate in (2.6), we can compute the first and second order approximations. The second order approximation helps us understand the effect of channel correlation. The variance term in (2.6) depends on \( G \), and a simple calculation shows that \( \sigma^2(u) \) depends on \( G \) roughly as \( 1/G \), or in other words, \( \sigma^2(u) \) is proportional to the mean correlation time of the channel. Thus, the effective service rate decreases with increasing
channel correlation, thereby increasing the mean delay. This increase is due to the fact that when the channel correlation is high, some users may stay in a low channel state for a long time, contributing an increase in the mean delay.

**Technical Conditions:** For the second order approximation, we noted in [1, Section 4] that TDMA satisfies the nice separability condition when the TDMA index is reset randomly after each arrival and departure event. When we prove the second order approximation in Section 6.3, we will argue that the second order approximation is applicable to TDMA even when such randomization is not performed (see the comment at the end of Section 6.3). For this reason, in our simulations, we will not carry out randomization of the TDMA index.

### 3 Accuracy of Time-Scale Separation

In this section, we test the accuracy of the first and second order approximations for the example TDMA system. The range of parameter values we consider can be divided into two broad categories. The first category corresponds to independent fading across time slots, while the second category corresponds to correlated fading across time slots.

#### 3.1 Independent Fading

When fading is independent across time-slots, all conditions required for the first order approximation are satisfied (see Section 2.3). Thus, the first order approximation can be applied when the time taken to process a file is much larger that the time-scale of fluctuation of the service process. This can be seen in Figure 2, where the \( \pi^\alpha \) and \( \mu \) coincide, confirming the applicability of the first order approximation in the large filesize regime.

Although the first order approximation is applicable under independent fading, the second order approximation is not applicable because the negligible discreteness condition ([1, Section 2.3]) does not hold. To see this, note that independent fading is obtained by setting \( G\delta = \frac{1}{2} \) in Figure 2.2. When \( G\delta = \frac{1}{2} \), from (2.6) it can be seen that \( \sigma(u)^2 = \sigma^2_{\text{slot}}/\delta \), where \( \sigma^2_{\text{slot}} \) is evaluated in (2.4). Further evaluation shows that \( \sigma(u)^2 \) is proportional to \( W^2\delta \), which is of the same order of magnitude as \( \lambda^2\delta \). Thus, the negligible discreteness condition is not satisfied.

The goal of the second order approximation was to improve the accuracy of the first order approximation when file sizes are not large. When the file size is moderate, and each file requires only a few slots for service, the first order approximation is not accurate. This can be seen in
Figure 2: TDMA with independent fading, large filesize $\tilde{S} = 0.5$, $\lambda = 1.2$, $\delta = 0.01$, $\text{SNR} = 0$ dB, and $g_1^2 = 1.6$.

Figure 3.

In the next section, we develop a second order approximation that is applicable even when discreteness is nonnegligible, and attain a better approximation for moderate file sizes.

### 3.2 Discreteness Correction

This section is devoted to improving the second order approximation when discreteness is not negligible. To begin, we examine the cause of the breakdown of the first order approximation. The dynamic user system deviates from the constant rate service model because the service is discrete and stochastic, and the effects of the stochastic deviation can be accounted for by the second order approximation. However, when discreteness is not negligible, we need to account for both the stochastic and discrete deviations in the service process. In this section, we develop such a correction. Although our treatment is not formal, we believe that a formal treatment can be developed using the techniques described in [1].

**Nonfading System:** To examine the effect of discreteness, consider a TDMA system *without fading*, where the deviation from a service rate that is only a function of the number of users in the system is solely due to discreteness. It is easy to check that the variance time-constant $\sigma^2(u)$ is zero for this system. For this nonfading system, we will calculate the effect of discreteness on the service
rate when $u$ users are present. Recall that the under the TDMA model without fading, a user will be served $\delta \phi(u)$ bits in a slot, and will receive service once every $u$ slots. Now, a typical user will require $\bar{S}/(\delta \phi(u))$ slots to complete service. However, observe that in the slot where service for a user is completed, on an average, half the service goes unused. This is because the residual filesize is less than $\delta \phi(u)$ in the slot where service is completed. Accounting for this wasted service shows, after a simple calculation, that the effective service rate is

$$\phi_{\text{eff}}(u) = \phi(u) \left(1 - \frac{\phi(u)\delta}{2\bar{S}}\right).$$  \hfill (3.1)

Using the effective service rate given above, we can get a modified second order approximation $\mu_{\delta}^\alpha$ for the distribution of interest $\pi$. Figure 4 shows that for the nonfading system, this modified second order approximation closely matches simulation results.

With this additional insight about the effect of discreteness, we now return to the case with fading. We know that stochastic deviation in the service processes reduces the service rate according to (2.6). Also, we have a heuristic that says that discrete deviation in the service rate reduces the service rate according to (3.1). Combining the effects of these two deviations gives the effective service rate

$$\phi_{\alpha}(u) = \phi(u) - \frac{\phi(u)^2\delta}{2\bar{S}} - \frac{\sigma^2(u)}{2\bar{S}}.$$  \hfill (3.2)
Let the steady state distribution for a symmetric queue with this service rate be $\mu_d^\alpha$, where the subscript $d$ denotes that discreteness is also taken into account. Figure 3 shows that this improved approximation gives an acceptable match between $\pi$ and $\mu_d^\alpha$.

3.3 Highly correlated fading

In the previous two sections, we examined the independent fading case. In this section, we consider the case where fading across time-slots is highly correlated, and show that such correlation increases the mean delay. In Section 2.3 we showed that the service variance (approximately) increases linearly with the correlation time of the channel. Thus, when the correlation time is much larger than the discreteness $\delta$, the negligible discreteness condition is satisfied, and the second order approximation will be applicable.

Figures 5 and 6 plot the distributions $\pi$, $\mu$, and $\mu^\alpha$ for different values of mean correlation time. When the mean delay is much larger than the correlation time (Figure 5), both the first and second order approximations are equally accurate. However, when the mean delay is of the same order as the correlation time (Figure 6), the second order approximation is considerably more accurate than the first order approximation, confirming the utility of the second order approximation.

Also, by comparing Figures 5 and 6, we can see that increasing the correlation time of the channel shifts the steady state distribution of the number of users to the right, thereby increasing the mean delay. This confirms our earlier observation (at the end of Section 2.3) that channel correlation has an adverse impact on mean delay.
Figure 5: TDMA with correlated fading. $\bar{S} = 0.5$, correlation time $= 0.5$, $\lambda = 1.2$, $\delta = 0.01$, SNR $= 0$ dB, and $g_{0}^{2} = 1.6$.

Figure 6: TDMA with correlated fading. $\bar{S} = 0.5$, correlation time $= 10$, $\lambda = 1.2$, $\delta = 0.01$, SNR $= 0$ dB, and $g_{0}^{2} = 1.6$. 
Jakes Fading Model: The numerical results given thus far assumed that the channel evolves according to a Gilbert-Eliot Markov model. However, constructive and destructive interference in wireless channels, are better modeled by the Jakes model [14]. Under the Jakes model with \( N \) oscillators, the channel state at time \( t \), \( h(t) \), is the sum of \( N \) complex signals

\[
h(t) = \sum_{n=1}^{N} W_n e^{2\pi \cos(2\pi fn/N)f_{Dop}t}
\]

where \( f_{Dop} \) is the Doppler frequency and \( \{W_n\} \) are i.i.d. zero-mean, unit variance, complex Gaussian random variables. Although the Jakes model accurately characterizes the spectral shape of the signal, from our point of view it has the technical limitation that it is not ergodic. To overcome this technical limitation, we assume that \( W_n \) is a function of time, denoted by \( W_n(t) \). In particular, we assume that \( W_n(t) \) is a zero-mean complex Gauss-Markov process with unit variance. For the channel described above, we compute \( \sigma^2(u) \) and \( \phi(u) \) by simulation, and obtain the first and second order approximations. We see in Figure 7 that the second order approximation is more accurate than the first order approximation. This confirms the utility of the second order approximation for realistic channel models.

![Figure 7: TDMA with correlated fading. \( \bar{S} = 0.5, \lambda = 1.0, \delta = 0.001 \text{ s}, \text{SNR} = 0 \text{ dB, Jakes fading with } N = 10, f_{Dop} = 10 \text{ Hz, mean correlation time for } W_n(t) = 1 \text{ s.} \)
4 Applications of Time-Scale Separation

Reducing the mean delay is one of the important design objectives for dynamic-user systems. The delay can be computed using the steady state distribution of the number of users, and Little’s law. In case the service process has a rate $\phi$ that is only a function of the number of users $u$, it is easy to see that the mean delay is minimized by maximizing $\phi(u)$ individually for each $u$, and the goal of maximizing the mean throughput $\phi(u)$ is given considerable importance in physical layer design.

However, when the service process $\phi$ is not just a function of $u$, roughly speaking, two factors influence the mean delay. The first factor is the mean service rate: the higher the mean service rate, the lower the delay. The second factor is the service variability: in case of high variability, the user might experience a large delay before the service rate reaches a value close to its average, making low service variability desirable. In this Section, we will study the interplay of service mean and variance for TDMA and scheduling based schemes.

For TDMA, we will show that the two goals of maximizing the average service rate and minimizing the service variance are contradictory. In particular, the average service rate is maximized by water filling in time, while the service variance is minimized by channel inversion. We will argue that the right balance between these two contradictory goals is to maximize the effective service rate in (2.6), and we will show that maximizing the effective rate gives better performance than maximizing the mean.

For scheduling, where the user with the largest channel is allocated a slot, it is shown in [5] that channel variability can be used to increase the mean service rate. This suggests that channel variability is desirable. However, the analysis of [5] ignores the service variability which is introduced by larger channel variability. We show that when the offered load is low, channel variability degrades performance because of the high service variability it produces. However, when the offered load is moderate or large, service variability improves performance, as predicted in [5].

4.1 Power Control for TDMA

For the TDMA system of Section 2.2, we wish to find a power control scheme that maximizes the effective service rate in (2.6) under an average power constraint. In other words, we wish to select $P(g_0)$ and $P(g_1)$ to maximize the effective service rate, under the constraint that $P(g_0) + P(g_1)$ is held constant.

We compare the delay under two schemes. The first scheme maximizes the mean service rate
\( \phi(u) \) using water-filling, while the second scheme maximizes the effective service rate (2.6). Numerical results in Figures 8 and 9 show that the second scheme gives as much as 30% improvement in the mean delay. This demonstrates the utility of the second order approximation in system design. With the aid of the second order approximation, we are able to balance the contradictory goals of high mean service rate and low service variance.

![Figure 8: Delay under water filling and second order optimal schemes for a TDMA system with parameters \( G = 0.99, \delta = 0.01, \text{SNR} = 0 \text{ dB}, g_0^2 = 1.6. \)](image)

Figure 8: Delay under water filling and second order optimal schemes for a TDMA system with parameters \( G = 0.99, \delta = 0.01, \text{SNR} = 0 \text{ dB}, g_0^2 = 1.6. \)

![Figure 9: Ratio of water filling delay and second order optimal delay.](image)

Figure 9: Ratio of water filling delay and second order optimal delay.

Note that the relative advantage of the second order design diminishes with increasing load. This is because the variance (2.5) decreases with increasing \( u \), making the effective service rate close to
the mean service rate. However, even in the regime of high loads, the second order optimal design should be preferred because it provides a small gain over water filling at no extra implementation cost.

4.2 Scheduling and Channel Variability

Scheduling refers to a physical layer scheme where the user with the highest channel gain is selected for transmission. It is been suggested in [5] that channel variability improves the performance of scheduling. In this section, we argue that although channel variability improves performance at high offered loads, it may degrade performance at low offered loads. We show that this degradation in performance is caused by two effects: the familiar Jensen’s penalty for fading channels, and the service process variability caused by larger channel variability.

For the study of scheduling, we cannot use the second order approximation. This is because scheduling does not satisfy the independent service condition discussed in [1, Section 4] and is thus not nicely separable (see [1, Definition 3]). For this reason, the arguments in this section will rely on the first order approximation and some heuristics. A detailed discussion on nonnicely separable systems is given in Section 5.

The system model for scheduling is as follows. Consider the same channel model as for the TDMA system in Section 2.2. We will control channel variability by adjusting the fading levels \( g_0 \) and \( g_1 \). Channel variability is maximum when \( g_0 = 0, g_1 = 2 \) (on-off fading), and is minimum when \( g_0 = g_1 = 1 \) (no fading). Following the notation of Section 2.2, let the information transmitted to a user with fading level \( g \) be \( C(g) \), and assume that \( P(g_0) = P(g_1) = \text{SNR} \), i.e., there is no power control.

**First Order Analysis:** Using the independence of fading for different users, the mean throughput under scheduling is

\[
\phi(u) = \left(1 - \frac{1}{2u}\right) C(g_1) + \frac{1}{2u} C(g_0).
\] (4.1)

The way in which channel variability affects the mean throughput depends on the number of users \( u \). For large \( u \), we get \( \phi(u) \approx C(g_1) \), and it is clear that greater channel variability, which increases \( g_1 \), also increases the mean throughput (a point noted in [5]). However, for small \( u \), for example \( u = 1 \), we get \( \phi(u) = \frac{1}{2}(C(g_0) + C(g_1)) \), and due to the concavity of the log function, the mean throughput actually decreases with increasing channel variability. (This decrease is due to the so-called Jensen penalty.) From the above arguments, it follows that channel variability is favorable...
when the number of users is large (large offered load), and unfavorable when the number of users is small (small offered load).

This effect is confirmed by simulation in Figure 10, where we compare the performance of two channels: Channel A with low channel variability, $g_0^2 = 0.8, g_1^2 = 1.2$, and Channel B with high channel variability, $g_0^2 = 0.4, g_1^2 = 1.6$.

![Figure 10: Performance of scheduling under low and high variance channels, $C = \log(1 + \text{SNR})$ model, with $\text{SNR} = 10$, correlation time = 2s.](image)

**Second Order Analysis:** Thus far, we have been concerned with the effect of channel variability on the service mean. However, we have not addressed the effect on service variability. To address this issue, we select the function $C$ as $C = \text{SNR}$, which is a linear function of the received energy, and thus, does not suffer from the Jensen penalty. Now, consider again the Channels A and B described earlier. Under the linear $C$ function, from (4.1) it follows that the mean service rate is higher for Channel B for all values of $u$. Thus, from a mean service rate perspective, Channel B should outperform Channel A, and service variability should improve performance.

We saw in Section 3.3 that the mean service rate does not characterize system performance completely, especially in the regime when the channel is highly correlated across time. In the high correlation regime, service variance affects the performance significantly, and we will show here that service variability can cause Channel A to outperform Channel B under low load conditions.

For the scheduling scheme, the service variance is not easy to compute because the service
process is more complicated than that for TDMA. However, it is easy to see that service variability is always higher for Channel B than it is for Channel A. Further, service variability is affected by the number of users as follows. When \( u \) is large, with high probability, service is provided to a user with fading level \( g_1 \), and the service variability is low. However, when \( u \) is small, service oscillates between users with fading levels \( g_0 \) and \( g_1 \), giving higher service variability.

From the above observations about the mean rate and service variability, we can draw the following conclusions about performance. At high loads, service variability is small and does not effect system performance significantly, and Channel B performs better due to its higher mean service. At low loads, service variability is large, and Channel A may perform better due to its lower service variance. These observations are confirmed by numerical results shown in Figure 11.

![Figure 11: Performance of scheduling under low and high variance channels, \( C' = \text{SNR model} \), with \( \text{SNR} = 1 \), correlation time = 2 s.](image)

The results of this section did not assume any any form of power control. However, if water-filling based power control is used in conjunction with scheduling, both the service rate and the service variability will increase. Although we do not study this problem here, it is a matter of interest to determine the role of service variability when power control is used, and it may be possible to design an optimal power control algorithm for scheduling, along the lines of the TDMA power control algorithm of Section 4.
5 Second Order Approximation with Dependent Service

In this section, we describe how the second order approximation can be extended to systems that do not satisfy the independent service condition described in Section 2.3. For the sake of brevity, the presentation here will not be as rigorous as in the proofs for the first and second order approximation theorems in Part I [1], but we believe a rigorous treatment can be provided. Also, this section will involve several references to the notation and proof technique in [1].

For a system with independent service, the jump chain distribution lemma of [1, Section 6.3] states that the jump chain distribution \( \gamma^\alpha(x) \) can be approximately decomposed as \( \eta(x)\gamma^\alpha(X_u) \). This means that at times of transition into the set \( X_u \), the state is distributed according to the steady state distribution in the set \( X_u \). The jump chain distribution lemma is used in the proof of the holding time and transition probability lemmas through [1, (6.13)], in the part where we use \( \sum_x \eta(x)\dot{F}(x) = 0 \).

When service is not independent, the distribution \( \gamma^\alpha \) may not satisfy the jump chain distribution lemma. We first describe a technique to evaluate \( \gamma^\alpha \), and then use \( \gamma^\alpha \) to develop a more general second order approximation.

For independent service, we argued in [1] that the nice arrivals and departures condition ([1, Definition 3]) hold, and these conditions enabled us to approximate \( \gamma^\alpha \). However, for a general service, arrivals and departures may not be nice. To evaluate \( \gamma^\alpha \) for such a system, we define new distributions \( \eta^\text{dn}_u \) and \( \eta^\text{up}_u \) as follows. When \( X^\alpha[0] \) is initialized by \( \eta_u \),

\[
\eta^\text{dn}_{u-1}(x) = \Pr(X^\alpha[1] = x \mid D_0 = 1, A_0 = 0) \quad (5.1)
\]

and \( \eta^\text{up}_u \) is defined by

\[
\eta^\text{dn}_{u+1}(x) = \Pr(X^\alpha[1] = x \mid D_0 = 0, A_0 = 1) \quad (5.2)
\]

The distribution \( \eta^\text{dn}_u \) is the distribution in \( X_{u-1} \) just after an arrival, while \( \eta^\text{dn}_u \) can be interpreted as the distribution in \( X_{u+1} \) just after a departure. Throughout the discussion in this section, we will assume that \( \eta^\text{dn}_u \) and \( \eta^\text{up}_u \) are known.

The following proposition shows how the approximation \( \gamma^\alpha(x) = \gamma^\alpha(X_u)\eta_u(x) + O(\alpha) \) (which is valid with independent service) is modified for systems with dependent service.

**Proposition 1** The distribution \( \gamma^\alpha(x) \) can be approximated as

\[
\gamma^\alpha(x) = \gamma^\alpha(X_u)\eta^\text{dn}_u(x) + O(\alpha)
\]
where
\[
\eta_u^{av}(x) = \frac{\mu(u-1)\Omega \eta_u^{dp}(x) + \mu(u+1)\phi(u+1)\eta_u^{dn}(x)/S}{\mu(u-1)\Omega + \mu(u+1)\phi(u+1)/S}.
\] (5.3)

We give an informal argument for the above proposition, and omit the detailed proof. If the distributions \(\eta^{dn}\) and \(\eta^{up}\) are the same as \(\eta_u\), then the method of proof for [1, Lemma 3] can be repeated to prove the above Proposition. However, when the two distributions are different, the average state after transition into set \(X_u\) will be a convex combination \(a\eta_u^{dn} + (1-a)\eta_u^{up}\), where \(a\) is the fraction of times state \(X_u\) is entered through a downward transition. From the first order approximation (which does not depend on the above Proposition), we can compute the fraction of time a state is entered through upward or downward transitions. This fraction turns out to be the quantity in (5.3).

When the independent service condition does not hold, \(\gamma^\alpha(x)\) is not well approximated by \(\eta_u\), and the derivation of [1, (6.13)], in particular the step in [1, (6.16)] will not work because

\[
\frac{1}{\gamma^\alpha(X_u)} \sum_{x \in X_u} \gamma^\alpha(x) \hat{F}_u(x) \gg O(\alpha).
\]

Although this summation is no longer zero, it can be approximated using Proposition 1 to give

\[
F_u^{av} \overset{\text{def}}{=} \sum_{x \in X_u} \eta_u^{av}(x) \hat{F}_u(x).
\]

Then, it can be verified that [1, (6.13)] is modified to

\[
\sum_{x \in X_u} \gamma^\alpha(x) \left[ \alpha\delta \sum_{k=1}^{\infty} \Pr(G^\alpha \geq k) \right] = \frac{\gamma^\alpha(X_u)(1 - \alpha F^{av}(u))}{\Omega + \phi^\alpha(u)} + \delta O(\alpha).
\]

It can be further verified that the second order approximation holds for the modified service rate given by

\[
\phi^\alpha = \phi(u) - \frac{1}{2}\alpha u\sigma^2(u) + \alpha F^{av}(u)(\Omega + \phi(u)).
\] (5.4)

In the next section, we apply this extension to an example system.

5.1 Example Nonnicely Separable System

Consider the same channel model as the TDMA system of Section 2.2, but consider a communication scheme based on nonorthogonal access. Let the transmit power of all users be constant at \(P\). Assume that the fading state is known to the receiver and not to the transmitter, and that each user uses i.i.d Gaussian symbols for signaling. Then, assuming single user decoding at the receiver, the
mutual information for user \( j \) in time slot \( k \) is

\[
I_{j,k} = W \delta \log \left( 1 + \frac{|h_{j,k}|^2}{WN_0/P + \sum_{j' \neq j} |h_{j',k}|^2} \right)
\]  

(5.5)

where \( 2W \) is the number of symbols transmitted each second and \( N_0 \) is the noise spectral density. This model can be interpreted as one obtained by introducing fading in the model of [15].

For the system described above, the distribution \( \eta_u^{\text{up}} \) will be the same as \( \eta_u \), if we assume that the fading state of the arriving user is distributed according to the marginal distribution of the fading. However, \( \eta_u^{\text{dn}} \) will not be the same as \( \eta_u \). This is because the independent service condition is not satisfied. In particular, the service given to user \( j \) depends on the future channel condition of other users (due to channel correlation across time). If user \( j \) departs, it is likely that the departure was caused by low channel gains of other users, which reduced the interference to user \( j \). Thus, at the point of departure, the remaining users will have channel gains that are biased below the marginal distribution, making the distribution \( \eta_u^{\text{dn}} \) different from the steady state distribution \( \eta \).

In most cases, it will be difficult to evaluate the distribution \( \eta_u^{\text{dn}} \) and the function \( \hat{F} \), making the evaluation of the correction term difficult. The reason for the simplicity of the second order approximation for nicely separable systems was that we did not need to concern ourselves with the inner working of the Markov chain \( X^\alpha[k] \). The extension to nonnicely separable systems, in contrast, requires us to evaluate detailed properties of the state space. This difficulty notwithstanding, we consider one case where the quantities \( \eta_u^{\text{dn}} \) and \( \hat{F} \) can be determined in closed form. This is the case where fading is governed by the Gilbert-Eliot model of Figure 2.2, and where the channel gain in the bad channel state is set to zero.

For this model, the system is state \( X^\alpha[k] \) consists of a two dimensional vector that contains the total number of users, and the number of users with a good channel. Further, the fixed-user system state consists only of the number of users with a good channel. For such a fixed-user system, the Poisson equation [1, eq. (6.8)] takes a simple form, and it is possible to solve for \( \hat{F} \) by solving a set of linear equations. Similarly, it is possible to compute the distribution of \( \eta^{\text{dn}}(u) \) by numerical means, which further enables the computation of the correction term (5.4). Details about the numerical computation of \( \hat{F} \) and \( \eta^{\text{dn}}(u) \) are omitted for brevity.

Results obtained by applying the modified second order approximation to the example system are given in Figure 12. It can be seen that the nonnice correction term considerably improves the accuracy of the second order approximation.
Useful Bound: Although the extension to the nonnice separable case is complicated, it is useful because it indicates that the second order approximation offers an optimistic estimate of effect of service variability on the overall performance. For example, in Figure 12, it can be seen that the unmodified second order approximation underestimates the delay. The reason for this underestimation lies the nature of the distribution \( \eta^{dn} \). For most physical layer schemes, violation of the independent service condition will mean that the service of one user is enhanced when the channel states of the other users are low. Thus, the channel state of the users remaining after a departure is less favorable compared to the marginal distribution of the channel state. Also, less favorable states will have a negative \( \hat{F} \), as evidenced by property [1, equation (6.10)] of the function \( \hat{F} \). This will make the correction term \( F^{av} \) negative, resulting in a further reduction in the effective service rate, that is in addition to the reduction caused by service variability.

6 Extension to Non-Exponential File Sizes

Until now, we have considered dynamic-user systems with exponentially distributed file size \( S \). In this section, we briefly describe a technique to extend the first order approximation to a more general class of file size distributions.
6.1 First Order Approximation

Kelly [16] gives a technique for the analysis of symmetric queues with general filesize distributions. The analysis makes use of the fact that any distribution can be approximated by a mixture of a finite number of gamma distributions. Let the filesize distribution be a mixture of $M$ gamma distributions, with gamma distribution $m$ corresponding to the sum of $n_m$ exponential random variables, each with mean $\tilde{S}_m$. Let the probability associated with gamma distribution $m$ be $p_m$, with $m = 1, \ldots, M$. Then, the mean filesize is related to the means of the constituent exponential random variables by

$$\bar{S} = \sum_{m=1}^{M} p_m n_m \tilde{S}_m.$$  \hspace{1cm} (6.1)

This model for the filesizes can be interpreted as dividing users into $M$ types, where a type $m$ user wishes to transfer $n_m$ files, each exponentially distributed with mean $\tilde{S}_m$.

In this case, the number of users in the symmetric queue $\bar{U}(t)$ does not form a Markov process, and the state needs to be enlarged to a multidimensional Markov process $\bar{V}(t)$ in a finite state-space $\mathcal{V}$. Details about the structure of $\mathcal{V}$ be found in [16].

Returning to the dynamic-user system, earlier we considered exponentially distributed $S$. The method of proof for the first order approximation result relies on splitting the state space $\mathcal{X}$ into a partition $\mathcal{X}_u$, with $u = 0, 1, \ldots$, where $\mathcal{X}_u$ corresponds to state $\bar{U}(t) = u$ for the symmetric queue.

With a general distribution for $S$, the definition of dynamic-user systems in [1, Section 3.1] needs to change. In particular, for a non-exponential dynamic-user system, the state space $\mathcal{X}$ should be split into a partition $\mathcal{X}_v$, where $v$ lies in the set $\mathcal{V}$. For this partition of the state space, a new version of the holding time and transition probability lemmas can be proved to show that transitions probabilities and mean holding times for the sets $\mathcal{X}_v$ are well approximated by the transition probabilities and mean holding times of states $v$ in the symmetric queue. These new lemmas can be used to modify the proof of the first order approximation theorem to prove that the steady state distribution of the sets $\mathcal{X}_v$ is the same as the steady state distribution of the Markov process $\bar{V}(t)$. In this manner, the first order approximation can be extended to non-exponential dynamic-user systems.

6.2 Second Order Approximation

The second order approximation makes explicit use of the exponential nature of the filesize distribution, and hence cannot be readily extended to arbitrary filesize distributions. However, it can be argued that the second order approximation continues to provide a good heuristic for system design.
To develop this heuristic, we consider a single-user system, where the service process has mean $\phi$ and variance $\sigma^2$. Let $\tau$ be the service time for the user. For large $S$, we will show that

$$E[\tau] \approx \frac{S}{\phi - \frac{\sigma^2}{2S}} \quad (6.2)$$

This mean service time is the same as that given by using the effective service rate $\phi - \sigma^2/2S$. Thus, even though the service time distributions depend on the exact filesize distribution, we establish that the service time mean depends only on the filesize mean. This gives some justification for the use of the second order approximation for general filesize distributions.

To show (6.2), we use the result in [17] that the service $\tilde{I}(t)$ can be approximated by a Gaussian random variable with mean $\phi(u)$ and variance $t\sigma^2$. Then, the distribution of $\tau$ is given by

$$\Pr(\tau > t) = \Pr(\tilde{I}(t) < S) = E_S[\Pr(\tilde{I}(t) < s \mid S = s)]$$

To evaluate $E[\tau]$, we integrate the above expression from $t = 0$ to $t = \infty$. For a given value $S = s$, the probability above can be evaluated from the Gaussian assumption for $\tilde{I}(t)$, and it can be verified easily that

$$E[\tau \mid S = s] = \frac{s}{\phi} + \frac{\sigma^2}{2\phi^2}$$

After some manipulation, this proves (6.2).

To verify the heuristic that the second order approximation is valid for general filesize distributions, numerical results for the TMDA system with constant file sizes are given in Figure 13. It can be seen from the figure that the second order approximation performs well when file sizes are constant.

### 7 Multiple User Classes

The key element of our analysis thus far has been the symmetric queue, which has a simple closed form solution. The symmetric queue model, however, breaks down when users are divided into different classes, with different mean service rates for users of each class. Borst [13] has used a multiclass symmetric queue model to analyze proportionally fair scheduling algorithms. However, Borst uses the time-scale separation in his analysis without proof. In this section we argue that first order time-scale separation continues to hold for multiclass symmetric queues. However, the extension of the second order approximation to multiclass systems is not that clear, and remains a topic for future work.
Consider the following model for a multiclass symmetric queue. Users are divided into \( \kappa \) classes, where each class corresponds to a different channel characteristic. Class \( c \) users arrive into the system according to a Poisson process of rate \( \lambda_c \) and have exponentially distributed file size with mean \( S_c \). Arrivals are blocked if there are more than \( u_0 \) users in the system. At time \( t \), let the number of users of class \( c \) be \( u_c(t) \), with \( U(t) = \{U_1(t), U_2(t), \ldots U_\kappa(t)\} \). Let the combined service rate for all users in class \( c \) be \( \phi_c(u(t)) \) bits/s. Then, \( U(t) \) forms a Markov process. Let the steady state distribution of this Markov process be \( \pi(u) \).

Now consider a multiclass dynamic-user system with \( \kappa \) user classes, and a variable rate service model. For the corresponding fixed-user system, let the the mean service rate for class \( c \) customers depend on the number and type of customers in the system through \( \phi_c(u) \). Then, in the limit of large filesizes, we can argue that this system is approximated by the multiclass symmetric queue of the previous paragraph. The argument for the validity of this approximation is similar to the one for nonexponential filesizes in Section 6.1. The state space \( \mathcal{X} \) can be partitioned into sets corresponding to different values of \( U \), and it can be shown that transitions between these sets are similar to the transitions between the states \( U \) in the multiclass symmetric queue. The first order result will then take the following form: if the steady state distribution for the multiclass dynamic-user system is \( \pi^\alpha(u) \), we will have

\[
\pi^\alpha(u) = \mu(u) + O(\alpha).
\]

The above approximation allows us to replace a multiclass dynamic-user system by a more
tractable multiclass symmetric queue. Unfortunately, however, for general functions $\phi_c(u)$, there may be no closed form solution for $\mu$, and numerical techniques may be required to estimate it. A simple product form for $\mu$ exists only under strict conditions on the service process (see, e.g., Kelly [16, Section 3.5]). Recently, Borst [13] has given an example of a practical multiclass dynamic-user system that has a product form solution. These product form solutions are valid for a class of orthogonal multiaccess systems where each user is served for an equal fraction of time.

8 Conclusions

For a dynamic-user centralized wireless network, we showed that the time-scale separation technique that we introduced in [1] provides an effective way to evaluate the quality of service (delay, blocking probability) in terms of the statistics of the service provided by the physical layer. This evaluation can enable cross-layer design, where service process statistics at the physical layer are selected to optimize a quality of service metric such as delay. We gave one example of cross-layer design by selecting power control to attain an optimal tradeoff between the twin goals of low service variability and high service mean.

We believe that time-scale separation can lead to other interesting cross-layer design applications. For example, the blocking probability can be expressed in terms of the distribution of the number of users in the system, $\pi(u)$, and the maximum load that can be sustained for a given blocking probability can be determined. Time-scale separation may also be useful in the cellular context, to study handoff, inter-cell interference statistics and multi-cell capacity. Applications of time-scale separation we have considered in earlier work include throughput-load tradeoff [2], and traffic-load based power control [3].

In addition to studying some key applications of the results of [1], we relaxed some of the conditions required in the analysis given in [1]. We showed that the first order approximation, and (less rigorously) the second order approximation remain unchanged under non-exponential filesizes. Further, we gave a technique to extend the second order approximation to cases with non-negligible discreteness and dependent service. These extensions show that time-scale separation is applicable to a variety of dynamic-user systems (TDMA, Nonorthogonal, Scheduling), under a variety of parameter values (large to moderate fileizes).

However, we were unable to relax the assumptions of user symmetry or symmetric resource allocation, and there is a need for future work to relax these assumptions. In the asymmetric user
case, the first order approximation can be computed only under strict conditions on the service process. Even when the service process satisfies these strict conditions, determining the form of the second order approximation remains an unsolved problem. In case an exact result along the lines of the second order approximation for symmetric users is not possible, it may be possible to obtain a heuristic similar to the non-exponential filesizes case of Section 6.2.

The analysis of asymmetric resource allocation schemes is a difficult problem, and some preliminary results are given in [6]. Although filesize based prioritization has the advantage of reducing the mean delay, it has the drawback in that it encourages users to gain an unfair advantage by breaking their files into smaller fragments. Also, schemes which minimize the mean delay often do so at the expense of fairness, and there is a possibility for future work on asymmetric resource allocation that attains an optimal tradeoff between mean delay and fairness.

A Variance for the TDMA Example

Fix the number of users at $u$, and without loss of generality consider user 0 who transmits in slots $nu, n = 0, 1, 2, \ldots$. With some abuse of notation, denote the channel state of this user at time $k$ by $h_k = h_{0,k}$, with the initial channel state $h_0$ taking the value $g_0$ or $g_1$ with equal probability. The information transmitted in fading state $h$ is $C_h$, e.g., when $h = g_0$, the information transmitted in a slot is $C_{g_0}$. The cumulative service process till time $t$, with $n = \lfloor t/\delta \rfloor$ is

$$I_j(t) = \sum_{k=0}^{n} C_{h_k}.$$

The distribution of $Y[k]$ conditioned on the initial state is

$$\Pr(h_k = h_0 \mid h_0) = \sum_{\ell=0, k \text{ even}}^{k} (G\delta)^\ell (1 - G\delta)^{n-\ell} \binom{k}{\ell}.$$

(A.1)

We are interested in the variance of $I_j(t)$. To keep the notation simple, we scale $C(h)$ as $C(g_0) = -1$ and $C(g_1) = 1$. This notation, together with the conditional distribution in (A.1), gives

$$\mathbb{E}[C(h_k)C(h_\ell)] = (1 - 2G\delta) |k-\ell| \mathbb{I}_{\{k\mid u\}} \mathbb{I}\{\ell\mid u\}$$

where the indicator functions indicate that the service is zero when it is not the turn of the user of interest to be served.

The variance of interest is then computed as:
Simplifying the above summation gives

\[\sum_{k=0}^{[n/u]} E[C(h_k)]^2 + 2 \sum_{k=0}^{[n/u]} \sum_{\ell=1}^{k-1} E[C(h_k)C(h_{\ell})] = \sum_{k=0}^{[n/u]} (1 + 2 \sum_{\ell=1}^{k-1} (1 - 2G\delta)^u(k-\ell)).\]

which leads to (2.5).

References


