Delay and Rate Constrained Transmission Policies over Wireless Channels

Dinesh Rajan, Ashutosh Sabharwal and Behnaam Aazhang
Department of Electrical and Computer Engineering, Rice University,
6100, S.Main, Houston, Texas 77005.{dinesh,ashu,aaz}@rice.edu

Abstract—**In this paper, we study delay and rate-constrained transmission of bursty traffic over wireless channels. We characterize the minimum power requirements via bounds for both single user and multiuser downlink problems, using a class of randomized first-come first-serve policies. We show that larger tolerable delay leads to power reduction, even for single-user Gaussian channels; a source coding interpretation is offered for the result. Further, we show that traffic with max-delay constraints requires more power than the same traffic with average-delay constraints.**

Keywords— **Traffic regulation, broadcast scheduling, wireless channels, randomized policies.**

I. Introduction

In this paper, we study the role of traffic model in the design of delay-limited transmission policies over wireless channels. The importance of incorporating traffic model in the design of wireless communication systems was recognized in [1], [2]. The effect of queuing for Gaussian multiaccess channels was studied in [3]; a buffered random access model with feedback is assumed. Schemes without feedback for two-state Gilbert-Elliot fading channels were considered with an average delay constraint to minimize average power in [4], and to minimize packet outage in [5].

In this paper, we characterize minimal power requirements to achieve zero outage [6] using first-come first-serve (FCFS) transmission policies for delay- and rate-constrained traffic. A *transmission policy* chooses the number of packets to transmit during every time slot based on the state of the queue and channel. In [5], we studied the throughput variation under average delay and power constraints. In this paper, we study the dual problem of minimizing the average power with constraints on delay and throughput, using a larger class of policies.

The main results of the paper can be summarized as follows. First, we prove the convexity of the delay-power region for randomized FCFS policies under zero outage. Second, we show that increasing the allowable delay leads to lesser average transmit power. Third, we show that policies which guarantee max-delay require more queue state information and higher average transmit power than policies with average delay guarantees.

In this paper, we limit our attention to Gaussian channels, although the methods easily extend to arbitrary finite-state channels [4], [5]. The simplified analysis brings out the role of delay constraints for bursty traffic in both single and multiuser case. The single user problem provides methods for traffic regulation in circuit-switched TDMA systems, and the multiuser problem forms the basis for scheduling in CDMA systems. Further, we restrict our attention to independent and identically distributed packet arrivals in each slot; again, the analysis easily extends to markov arrival processes which are good models for multimedia applications like video transmission.

In Section II, we introduce the system and notation used, and formulate the problem. In Section III, we study policies with absolute delay guarantees. Section IV deals with the transmission policies for the two user case. We provide discussion on protocol information needed for transmitting variable rate traffic in Section V, and conclude the paper in Section VI.

II. System and Problem Formulation

In this section we focus on the transmission of packets arriving at a single finite queue and being transmitted over a wireless channel. We find lower and upper bounds for the average power required under given average delay constraints among the class of FCFS, zero outage policies.

The following notation is used: We consider a time-slotted transmission system where L represents the maximum buffer size, M is the maximum number of packet arrivals during any time slot, n represents the time slot, x_n is the number of packets in the buffer, β_n is the number of packets transmitted, P_n is the transmission power, a_n is the number of packet arrivals, A_n represent the channel fading state,[1] and $\lambda = \mathbb{E}[a_n]$ is the average packet arrival rate. These are illustrated in Figure 4. α is a stochastic transmission policy specified by a $(L+1)\times(L+1)$ upper triangular matrix. $\alpha_{i,j}$ represents the probability of transmitting j packets when the buffer is in state i, where $0 \leq j \leq i \leq L$. From requirement of normalization of probability distributions we have,

$$\sum_{j=1}^{i} \alpha_{i,j} = 1.$$

In this paper we consider both *deterministic* and *randomized* policies. The former policies use a single valued mapping from the state of the system to the action under that state. On the other hand, *randomized* policies use a probabilistic mapping from each system state to a set of possible actions under each state. In the case of *deterministic* policies, all elements of α are either one or zero. The transition probabilities between the different buffer states under a given policy α is denoted by $\mathbf{C}$ and has dimension $(L+1)\times(L+1)$. The probability of transition from buffer state $x_n = i$ to buffer state $x_{n+1} = j$ is denoted by $C_{i,j}$. In this paper we will limit our attention to stationary policies and hence α and $\mathbf{C}$ are not functions of the time index n.

We assume that all transmission occurs in packets (with no packet subdivision), and each packet is long enough to allow re-

[1] In our simulations this is a constant, but the analysis holds for Markovian fading models.

liable communication close to mutual information; this is motivated by outage versus capacity analysis in [7]. Define Shannon capacity function ρ as

$$\rho(P_n, A_n, \sigma^2) = \frac{1}{S}\log_2\left(1 + \frac{|A_n|^2 P_n}{\sigma^2}\right), \tag{1}$$

where σ^2 represents the noise variance at receiver and S is a normalization factor that depends on the size of the packets.

We assume that all packets which arrive for transmission during a certain time slot are available for transmission at the next time slot. The buffer state is then given by

$$x_{n+1} = \min(x_n + a_n - \beta_n, L). \tag{2}$$

The stationary probability of buffer state i is denoted by s_i with $\mathbf{s} = [\ s_0\ s_1\ s_2\ ...\ s_L\ ]^T$. From definition of $\mathbf{C}$ and s, it follows that

$$\mathbf{C}\,\mathbf{s} \ = \ \mathbf{s}. \tag{3}$$

The following proposition provides necessary and sufficient conditions for zero packet loss, *i.e.* zero outage.

Proposition A : Consider a queue with finite buffer size L and an input process with no more than M packet arrivals in one slot. A stationary stochastic policy is zero outage if and only if for each state $x_n = i$ one of the following two conditions are satisfied.

1. The stationary distribution is zero for state i, *i.e.* $s_i = 0$.
2. If z denotes the minimum number of packets transmitted in state i under the given policy, then $(i - z) \ \leq \ (L - M)$. In the notation we have introduced, $z \ = \ \arg\min_j \alpha_{i,j} \ \neq \ 0$.

Proof: [8]. □

Since we are considering only zero-outage policies, there are certain constraints on β_n and P_n [5] namely,

$$\beta_n \ = \ \lfloor \rho(P_n, A_n, \sigma^2) \rfloor. \tag{4}$$

In other words, packets are not lost in the channel. Under a given policy α and a given input distribution, using Little's result [9] the average packet delay can be shown to be

$$D_{avg}(\alpha) \ = \ \frac{1}{\lambda}\left(\sum_{i=0}^{L} s_i\, i\right). \tag{5}$$

Average delay is specified in units of number of time slots and the minimum possible value is one, corresponding to the policy where all packets are transmitted in the next slot after they arrive. The average transmission power is given by

$$P_{avg}(\alpha) \ = \ \sum_{i=0}^{L}\sum_{j=0}^{i} \alpha_{i,j}\ s_i f(j), \tag{6}$$

where

$$f(j) \ = \ \frac{\sigma^2}{|A_n|^2}(2^{Sj} - 1) \tag{7}$$

Unlike [5], in this paper we consider only zero outage policies and minimize the average transmit power under given average delay constraint. We denote the set of all randomized zero outage policies as Θ, *i.e.*

$$\Theta = \{\alpha : \alpha \text{ is a zero outage }\} \tag{8}$$

The problem we are interested in can be stated as

$$\min_{\substack{\Theta \\ \mathbb{E}[x_n] \ \leq \ \lambda D_{limit}}} \quad \mathbb{E}[P_n] \tag{9}$$

Deterministic as well as randomized policies are allowed in the above optimization process. A point (d, p) on the average delay limit versus average transmit power limit plane is achievable if there exists a transmission policy (deterministic/randomized) that achieves zero outage at an average delay less than d and average power less than p.

Clearly, the policy which guarantees the minimum average delay of one is characterized by α with ones along the main diagonal in the first $M + 1$ columns. This corresponds to a policy which completely flushes the queue at all times.

Definition: We define an equivalence class as follows $\mathbf{Q_{D_0,P_0}} = \{\alpha : \alpha \in \Theta, D_{avg}(\alpha) = D_0; P_{avg}(\alpha) = P_0\}$. Restating the same, an equivalence class is a set of stationary policies having zero outage and exactly the same average delay and average power. More than one deterministic policy could belong to the same equivalence class $\mathbf{Q_{D_0,P_0}}$. This is illustrated in Example 1.

Example 1 :

$$\alpha_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \alpha_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\alpha_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \alpha_4 = \begin{bmatrix} 1 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If $M = 2$, then $\alpha_1 \equiv \alpha_2$ where $\equiv$ is used to represent that the two policies have same average delay and average power and hence belong to the same equivalence class. $\alpha_4 = 0.6\ \alpha_1 + 0.4\ \alpha_3$ This is true for any $M \leq 3$. If $M = 2$, then $\alpha_4 \equiv 0.6\ \alpha_2 + 0.4\ \alpha_3$.

Observation: Given a finite buffer length, the set of all possible deterministic policies form a basis set for all possible randomized policies.

Proposition B: Consider a queue with finite buffer size L and an input process with no more than M packet arrivals in one slot. If a stationary deterministic policy having zero outage belongs to $\mathbf{Q_{D_0,P_0}}$ and satisfies condition 1 of **Proposition A** then there exists a zero outage stationary deterministic policy in $\mathbf{Q_{D_0,P_0}}$ that satisfies condition 2 of **Proposition A**.

Proof: [8]. □

Proposition C: Consider a queue with finite buffer size L and an input process with no more than M packet arrivals in one slot. The average packet delay achieved by any stationary randomized policy having zero outage probability is given by the convex combination of the average packet delays achieved by all possible stationary deterministic policies having zero outage. In addition, the same convex combination of the average powers of all stationary deterministic policies having zero outage gives the average power required by that stationary randomized policy.

Proof: [8]. □

Example 2 illustrates the application of this proposition.

Example 2: Let $M = 2$, and

$$\alpha_5 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Time sharing between α_5 and α_3 might not guarantee a policy with zero outage. The probability of being in buffer state 3 is non-zero under policy α_3, and if we then switch to policy α_5, we will remain in buffer state 3 indefinitely and continue to loose packets due to buffer overflow. Instead we can time share between policies α_2 and α_3 so that outage becomes zero and still achieve the same average delays and average powers as those achieved by a linear combination of α_2 and α_5.

Using the above proposition, all vertices of the d-p achievable region can be obtained using deterministic policies. Also, the region is convex. Time sharing of policies can be used to convexify the region if it was non-convex.

Now, in the solution of (9) we will restrict ourselves to *deterministic* policies to find the vertices of the achievable region and then find the convex hull of the same. Solving (9) in the general case is difficult and we resort to finding lower and upper bounds to the same using techniques similar to [4] [5]. The lower bound gives the non-achievable region while the upper bound gives the achievable region. We use Dynamic programming methods [10] to find the bounds. In particular we use the Value Iteration Algorithm [11] to find near-optimal policies. Two examples of the transmission policy given by this optimzation procedure is shown in Figure 1 for achieving two different points on the boundary of the achievable region.

The plot of the lower bound and upper bound for transmission power versus average delay limit is shown in Figure 2. Transmission policies exists which achieve average delays and powers in the region above the solid curve (*i.e.* upper bound) while no policies can guarantee delays and powers in the region below the dotted line (*i.e.* lower bound). Since the lower and upper bounds are almost the same for all average delay limits, it is clear that for all practical purposes these curves represent the capacity regions. It is clear that the power required reduces as we relax the delay constraint. For finite buffer sizes, it is not possible to achieve arbitrarily large delays and this explains the floor of the curve. Also shown is the minimum power required to transmit packets that arrive at a uniform rate without any burstiness (indicated as Shannon's limit in Figure 2). The excess power required is an indication of the loss due to finite delay requirements. As expected, with increasing delays, the loss due to burstiness goes down. For comparison we have also included the power required versus delay while transmitting at some finite constant or zero power(indicated by o in Figure 2) which is specified as follows

$$\beta_n = \begin{cases} w & \text{if } x_n \geq w \\ 0 & \text{if } x_n < w \end{cases} \tag{10}$$

This gives an indication of the gains in varying the transmit power based on the queue state. Only certain values of w can guarantee zero outage and hence the curve has step like behavior.

If the available power P_0 is greater than the power specified by the upper bound for a given D_{avg} then we can design a transmission policy which satisfies the given constraints. If P_0 is lower than the power specified by the lower bound for a given D_{avg} then such powers can not be achieved under zero outage and one can design policies which minimize the outage and meet the desired delay and power constraints [5].

III. Policies for absolute delay guarantees

In this section, we will study zero outage policies with absolute delay constraint instead of average delay constraint; absolute delay is better suited for multimedia applications like video and audio.

In order to find a schedule that minimizes the probability of outage under given absolute delay and average power constraints, the length of the buffer does not carry sufficient information to make a 'good' scheme. For example, if during time slot n, there are x packets in the buffer, one does not know the time stamp when the packets in the Queue arrived. So to guarantee any absolute delay, the policy would have to transmit all packets in the buffer and this may not be very efficient. It is easy to see that the profile on the absolute delay limit versus average power limit plane will be a straight line of constant power corresponding to average delay limit of one.[2]

If in addition to the length of the buffer, one has knowledge of the time stamps of the different packets, then a policy with lower transmit power can be formulated for a given absolute delay limit. In this case we treat the state of the system $\mathbf{B_n}$ to be a D_{abs}-dimensional vector as follows

$$\mathbf{B_n} = \{(b_{1,n}, b_{2,n}, ..., b_{D_{abs},n}) : 0 \leq b_{i,n} \leq M\}, \tag{11}$$

where D_{abs} is the maximum allowed delay for any packet and $b_{i,n}$ represents the number of packets that arrived during the $(n-i)^{th}$ slot that have not been transmitted before time slot n. It can be easily shown that in any time slot we do not need information of time arrivals of all packets before that time but only that of D_{abs} time slots prior to the current slot [8]. We pose the optimization problem as follows,

$$\min_{\substack{\Theta \\ d_{max} \leq D_{abs}}} \mathbb{E}[P_n], \tag{12}$$

where d_{max} represents the maximum of the actual delay suffered among all packets. As in (9), we use the value iteration

[2] Since we are assuming causal systems, average delay limit of one gives the same policy as absolute delay limit of one.

algorithm [11] to solve the above problem. The result of this optimization is indicated in Figure 2. The three points marked by x correspond to absolute delay guarantees of one, two and three time instants. For each of those D_{abs} values, we plot the average power required versus the average delay achieved. For a given average delay, the difference between the curves marked average delay - achievable and absolute delay - achievable indicates the excess transmit power required at a certain average delay to additionally guarantee an absolute delay limit. The required power reduces with increasing the absolute delay limit. However, just like in the average delay case, the power required has a floor due to finite buffer sizes.

It is clear that this representation of state (11) carries more information than just the buffer length which was used in optimizing under average delay constraints. In the case of average delay constraints the number of *states* as defined in (11) would become non-trivial to handle. Also, Little's result [9] gives a simple relation between average delay and average buffer length whereas there is no simple relation between average buffer length and absolute delay.

Having an average delay constraint fixes the mean of the departure distribution of packets from the queue while an absolute delay constraint fixes the tail of the departure distribution. In practise, we can control both the mean and tail of the departure distribution by optimizing the following.

$$\min_{\substack{\Theta \\ d_{max} \leq D_{abs}}} \quad \mathbb{E}[P_n + \delta x_n] \tag{13}$$

For different values of δ we would get different average transmit powers and average delays under some absolute delay guarantees.

IV. Two users with separate queues

In this case we assume there are two separate queues for the two different users and we focus on the broadcast scenario where a single transmitter sends the information to two different receivers. The channels seen by the two receivers as well as the input arrivals to the two queues are assumed to be independent. The transmitter is assumed to have complete information of both the channels and the receivers are assumed to have full knowledge of only their channels. The capacity region for a degraded broadcast channel is given by [12]

$$R_1 \leq \rho\left(\gamma P,\ A_1, \sigma_1^2\right) \tag{14}$$

$$R_2 \leq \rho\left((1-\gamma) P,\ A_2, \sigma_2^2 + \gamma P\right) \tag{15}$$

In this case, we minimize the total transmit power while guaranteeing the average delay constraints for both the users. The lower and upper bounds on the total power are shown in Figure 3. The figure shows the total average power required versus average delay limit of user 1 for two different average delay limits of user 2. For low values of delay limits for both users, there is a gap between the two bounds. If either of the two users have high delay limits, then the bounds are almost identical. The results can be extended to max-delay case just like in the single user case. In the case of more than two users in the system, although our method is still applicable, the computational costs can be quite high. We have proposed and analyzed the performance of closed form suboptimal policies having low complexity in [8].

V. Source coding interpretation of delay-power tradeoff

In any system where the information transmission rate varies from one time slot to another, the receiver needs to know the instantaneous rate to decode correctly. This *extra* information needs to be transmitted across the channel and is commonly denoted as protocol information [13]. Since our transmission policies are multi-rate, we need this protocol information to be transmitted. The entropy of the source consists of the entropy due to contents of the packets, the entropy in their arrival times and entropy in length of the packets. Since we are assuming constant size packets, there is zero entropy in the length of the packets and the departure times conditioned on the arrival times. A policy which guarantees an average or absolute delay of one transmits the information about packet arrivals to the receiver (irrespective of whether the receiver uses this information). As delay requirements are relaxed, our transmission policy smoothens the burstiness of the packet arrivals which in turn reduces the entropy due to packet arrivals. This is reflected in Figure 2 as reduced power requirement with increasing delays. In the limit when the delay goes to infinity, no information about packet arrivals is transmitted to the receiver and hence the power required asymptotically reaches the power required to transmit constant rate traffic.

VI. Conclusions

In this paper we presented a lower bound for the power required to guarantee any average packet delay for transmitting bursty traffic having independent and identically distributed packet arrivals in each slot through Gaussian channels. We showed achievablity of this bound in the single user case. The power requirements decreases with relaxing delay constraints and we offered a source-coding interpretation of this result. For max-delay guarantees, we showed that we need more information about packet arrivals than just buffer state and more power than the average delay case. In the two user broadcast scenario, we showed the limits of total transmit power under constrained average packet delays for both users. We showed the achievability of this bound when the delay requirements of atleast one of the users is large. The results are easily extended to MArkovian type fading channels.

References

[1] Robert G. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 124–142, Mar. 1985.

[2] Anthony Ephremides and Bruce Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[3] I. Emre Telatar and Robert G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on Select Areas in Communications*, vol. 13, no. 6, pp. 963–969, Aug. 1995.

[4] B.E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," *Proceedings of Allerton International Conference on Communication, Control and Computing*, pp. 709–717, 1999.

[5] Dinesh Rajan, Ashutosh Sabharwal, and Behnaam Aazhang, "Transmission policies for bursty sources through time varying channels," *35th An-*

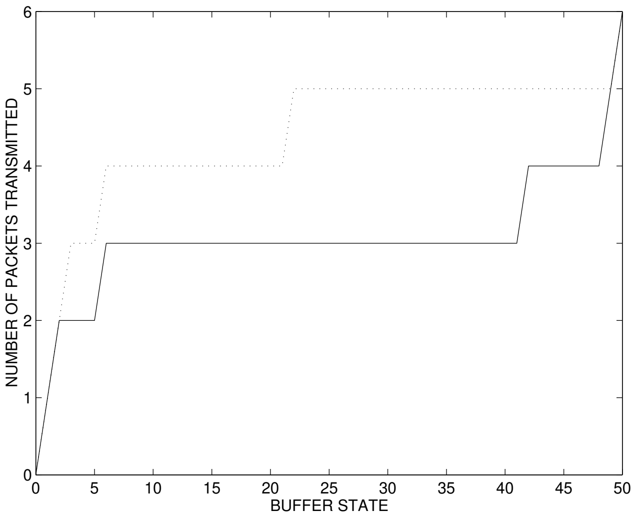

Fig. 1. Transmission policies for achieving different average delay constraints at minimum possible power. Buffer size L=50, maximum arrival M=6, uniform arrival distribution.

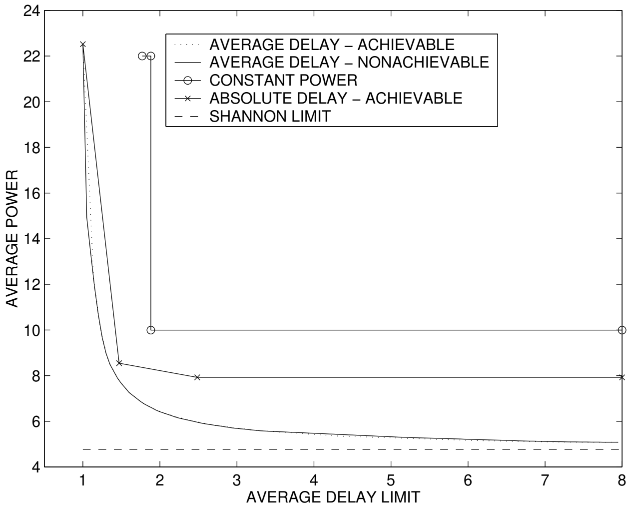

Fig. 2. Achievable and Non-achievable average delay versus average power profile under zero outage probability, AWGN channel. The points marked by x represent the lowest average power achievable under absolute delay limits of 1, 2 and 3. Buffer size L=50, maximum arrival M=6, uniform arrival distribution.

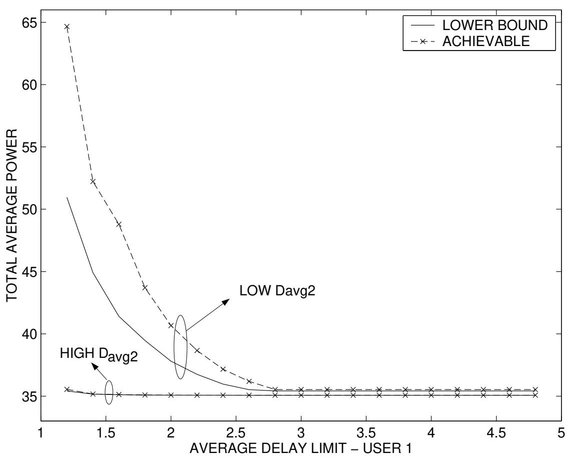

Fig. 3. Achievable and Non-achievable regions for 2 user case. The plots show variation of total transmit power versus average delay limit of user 1 for different values of average delay limit of user 2. Buffer sizes for user 1 ans user 2 equals 10. i.i.d. arrival process for both users with M = 4, uniformly distributed.

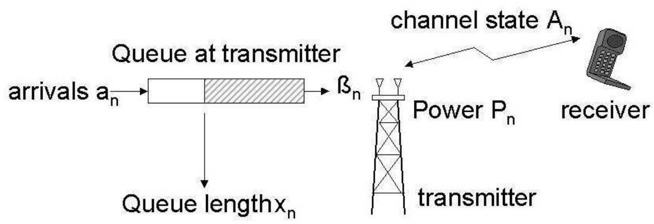

Fig. 4. Schematic of transmitter queue and receiver

nual Conference on Information Sciences and Systems, Baltimore, Mar. 2001.

[6] Leonidas Georgiadis, Roch Guerin, and Abhay Parekh, "Optimal multiplexing on a single link: Delay and buffer requirements," *IEEE Transactions on Information Theory*, pp. 1518–1535, September 1997.

[7] Lawrence H. Ozarow, Shlomo Shamai, and Aaron D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[8] Dinesh Rajan, Ashutosh Sabharwal, and Behnaam Aazhang, "Delay bounded scheduling of bursty traffic over wireless channels," *To be sumitted to IEEE Transactions on Information Theory*.

[9] Dimitri P. Bertsekas and Robert Gallager, *Data Networks*, Prentice Hall, 1992.

[10] Dimitri P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic models*, Prentice Hall, 1987.

[11] Linn I. Sennott, *Stochastic Dynamic Programming and the control of Queueing Systems*, John Wiley and Sons, Inc., 1999.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[13] Robert G. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 385–399, Jul. 1976.