WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2012; **12**:1215–1224
Published online 28 December 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/wcm.1050

RESEARCH ARTICLE

Power efficient scheduling over fading channel for cross-layer optimization

Xiaofeng Bai[1], Abdallah Shami[2]* and Serguei Primak[2]

[1] Motorola Inc., USA
[2] Department of Electrical and Computer Engineering, The University of Western Ontario, Canada

ABSTRACT

We consider the minimization of long-term average power consumption for packet transmission between a mobile station and the base station over Nakagami-*m* fading channel. Power consumption is minimized by intelligent transmission scheduling design, with the average queuing delay and joint packet loss across MAC and physical layers being confined below certain levels. The problem is formulated as an infinite horizon constrained Markov decision problem and solved by linear programming (LP) method. The primary intention of this paper is to provide a visible paradigm on using LP method to optimize the performance of mobile wireless communication systems. We elaborate the detailed mathematical solution with consistent simulation experiments and emphasize the effectiveness of adaptive transmission scheduling for cross-layer QoS provisioning. Copyright © 2010 John Wiley & Sons, Ltd.

KEYWORDS

finite state Markov channel; transmission power control; cross-layer optimization; constrained Markov decision process; randomized policy; linear programming

***Correspondence**

Abdallah Shami, Department of Electrical and Computer Engineering, The University of Western Ontario, Canada.
E-mail: ashami@eng.uwo.ca

[†]A preliminary version of this paper appeared in the Proceedings of IEEE ICC 2008 [1]

1. INTRODUCTION

In mobile broadband communication systems, adaptive transmission has been accepted as an effective solution to provide quality-of-service (QoS) for non-uniform traffic flows over time-varying wireless fading channel. The advantages of adaptive transmission in such systems are two-fold. On one side, by varying the media access control (MAC) layer data rate released to the physical (PHY) layer interface, performance impairments due to channel variation and bursty traffic profiles can be minimized. By varying the PHY layer transmission power along with MAC layer data rate, on the other side, energy consumption for providing required QoS can be optimized. Therefore, an efficient transmission scheduling strategy can best utilize the wireless channel dynamic and real-time traffic characteristic to improve MAC-PHY cross-layer performance, with minimal average power consumption. In previous research, power adaptation schemes either require error-free transmission over the channel or enforce dropping-exempt MAC queuing, both cost increased energy budget due to higher instantaneous transmission power required to achieve such conditions. Moreover, though many research works have studied on constrained Markov Decision Process and Linear Programming (LP) theories and acknowledged their applicability to wireless communication systems, few paper could clearly picture such an application. In this work, we aim to fill this gap with the designed power-efficient transmission scheduling method that simultaneously considers channel loss, MAC dropping and average power drain. The major contributes of this work are as follows: (1) this paper illustrates the effectiveness of scheduling-based transmission power control using discrete power levels. (2) This paper considers dynamic performance adjustment across MAC and PHY layers to provide cross-layer QoS guarantee. (3) This paper presents detailed mathematical solution for the researched problem, which can serve as a paradigm on solving similar problems.

The rest of this paper is organized as follows. In Section 2 we briefly review some related research work. In Section

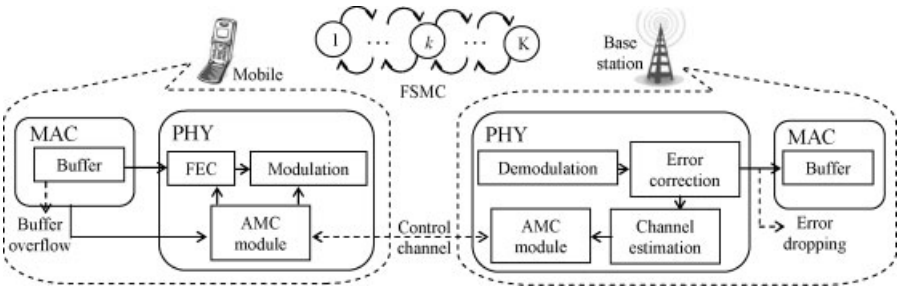

Figure 1. Communication system model.

3 we give the system model used for this study. Section 4 formulates the power-efficient transmission scheduling design problem as constrained Markov decision problem (MDP) and introduces the LP method. Section 5 elaborates detailed solution on applying LP for the researched problem. Finally, Section 6 verifies the proposed design with simulation experiments and further comments on the optimality, boundedness and parameterization issues of the designed scheme, before concluding this study in Section 7.

2. RESEARCH BACKGROUND

Power-efficient adaptive transmission has been studied by many researchers for mobile wireless communication systems. These research works can be roughly categorized as QoS-relevant and non-QoS-relevant designs. The research on QoS-relevant power control mainly focus on dynamically varying available data rate in terms of changing transmission power to achieve certain deterministic/statistical QoS criteria, such as required signal-to-interference ratio (SIR) [2], average queuing delay [3–5], average buffer overflow probability [6,7], QoS exponent [8,9]. Studies on non-QoS-relevant power control [10,11], though less concerned on MAC layer performance, are interested in maximizing channel capacity and bandwidth efficiency with limited average power supply.

Most of the previous QoS-relevant adaptive transmission schemes applied continuous data rates and power levels for the corresponding optimal design. The effect of scheduling based power optimization with discrete power/rate levels has not been well studied. Moreover, in existing research the MAC layer QoS metrics were confined exclusive of a zero (*e.g.*, [6]) or bounded (*e.g.*, [7]) bit error rate (BER) preset at the PHY layer. This limits the capability of transmission scheduling by simultaneously balancing the costs at both MAC and PHY layers, while leaving the designed optimality subject to PHY layer presumptions. In this study, we elaborate a multi-mode transmission scheduling algorithm that minimizes long-term average power usage with joint packet loss constraint defined across MAC and PHY layers. We believe this work combined with previous research on this topic[‡] will usher a visible guideline for solving other similar problems.

3. MODEL DESCRIPTION

3.1. Communication system

We consider packet transmission from a wireless mobile station to the base station, as illustrated in Figure 1. In this model, a single-transmitting antenna and a single-receiving antenna are separated by the time-varying wireless fading channel. At the transmitter side, packet stream arriving from upper layers are enqueued into a limited buffer space at the MAC layer, where overflowed packets are dropped. Backlogged packets are dequeued, on first-come-first-serve basis, and further processed by the forward error correction (FEC) encoder. Codewords are then mapped into certain number of PHY layer channel symbols at the digital modulator and finally released into the wireless channel *via* the transmitting antenna. Undergoing distortion and attenuation, at the receiver side, received symbols are sequentially passed through demodulator and decoder, to retrieve MAC layer packets. Packets that cannot be recovered due to uncorrectable transmission errors are dropped at the PHY layer.

The PHY layer operation runs on (time)frame-by-(time)frame basis, where each time frame is a fixed time duration containing multiple channel symbols. Symbols transmitted within each time frame are functionally grouped into frame control section and data transmission section, as illustrated in Figure 2. The frame control section consists of a certain number of pilot symbols permitting the receiver to estimate instantaneous channel gain, followed by symbols carrying uplink control information indicating transmission

[‡] In Reference [7], the LP-based adaptive transmission scheduling is introduced however with detailed mathematical derivations omitted and yet to be experimentally verified. For example, in order to correctly define the sought control policy in Equation (25), the packet dropping probability cost defined in Equation (16) should be weighted by probability $p(B_n - \Psi(U_n) = k)$ for each $I_{(B_n-\Psi(U_n)=k)}$. However, the value of $p(B_n - \Psi(U_n) = k)$ is in turn contingent on the control policy decided by Equation (25). Such ambiguities are clearly addressed in this paper with simulation verifications.

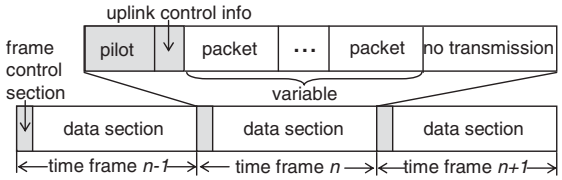

Figure 2. Illustration of time frame and its components.

mode index for different modulation and coding schemes used to process the following data symbols. Frame control symbols are transmitted with robust modulation/coding and constant power level. We assume a fast feedback channel that delivers the current channel state information (CSI) to the mobile station before the termination of transmitting frame control symbols. Hence the adaptive modulation and coding (AMC) module in the mobile station can decide the transmission mode to be applied, based on the current CSI and its buffer occupancy, before the start of data packet transmission in the same time frame. Moreover, packets arriving from upper layers at the MAC layer during $[nT_{\mathrm{f}}, (n+1)T_{\mathrm{f}})$ are enqueued at time $(n+1)T_{\mathrm{f}}$ and will be serviced in one of subsequent time frames, where T_{f} denotes the duration of one time frame.

3.2. Channel model

The dynamic of wireless channel is modeled as Nakagami-*m* fading channel [12], which is defined by the probability density function (PDF) of the signal-to-noise ratio (SNR) by detecting received pilot symbols as:

$$p_{\gamma}(\gamma) = \frac{m^m \gamma^{m-1}}{\rho^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\rho}\right) \tag{1}$$

with

$$\Gamma(m) = \int_0^{\infty} t^{m-1} e^{-t} \mathrm{d}t \tag{2}$$

being the Gamma function, $m \geq 1/2$ the Nakagami fading parameter, and ρ the average received SNR per symbol, respectively. The continuous fading process is discretized by finite state Markov channel (FSMC) model [13]. FSMC is a block fading model for slow-varying flat fading channels, where the channel is assumed to stay in the same state within one block period. Specifically, let $0 = \Gamma_0 < \Gamma_1 < \Gamma_2 < \cdots < \Gamma_{K-1} < \Gamma_K = \infty$ be the SNR thresholds partitioning channel status, the channel is said to be in state $k = 1, 2, \cdots K$, if the received SNR falls into the interval of $[\Gamma_{k-1}, \Gamma_k)$. We define a fading block as one time frame duration depicted in subsection 3.1. As in [7,13,14], moreover, the channel fading is assumed slow enough such that state transition between adjacent time frames only occurs over neighboring states. Namely, denoting $p_{i,j}$ as the transition probability from state i to state j where $i, j = 1, 2, \cdots K$, then $p_{i,j} = 0, \forall\ |i-j| \geq 2$. In such case, other state transition probabilities of the FSMC for Nakagami-*m* fading channel can be found in Reference [14]. In this study, we partition the channel with the *equal probability partition* method introduced in Reference [13], although other channel partitions are also applicable. This method partitions the received SNR beyond a threshold level γ_{cutoff} that prevents from using deep channel fades into FSMC states such that the stationary probability for the channel to dwell in each usable state is equalized.

3.3. Rate and power adaptation

Transmission rate and power are jointly adapted by applying AMC with different power levels. Particularly, power level $\Psi(a) \triangleq \{\psi_1(a), \psi_2(a), \cdots \psi_n(a)\}$ is available to AMC mode a that provides data rate $R(a)$. In general, a larger data rate promptly vacates MAC buffer space but suffers from more transmission errors, while a higher power level reduces transmission errors however consumes more energy. By applying different rate-power combinations *via* multiple AMC modes, therefore, the PHY layer error performance and energy consumption, as well as MAC layer QoS metrics (*e.g.*, service delay or buffer overflow probability) can be concurrently balanced through the same multi-mode transmission scheduling design.

4. PROBLEM STATEMENT

Taking discretized FSMC states into consideration, the transmission power applied during any time frame is now a function of both the AMC mode a being selected and the instantaneous channel state k, *i.e.*, $\psi(a,k) \in \Psi(a)$. The objective of power-efficient transmission scheduling is to minimize the long-term average power consumption of the mobile while maintaining the following QoS constraints: (1) joint packet loss rate at both PHY and MAC layers, *i.e.*, either by erroneous transmission or by buffer overflow, is below a certain level ε; (2) MAC layer average packet queuing delay is less than the specified value δ.

To investigate such cross-layer QoS constrained transmission scheduling, we convert the joint packet loss rate constraint into the constraint on MAC layer throughput. This avoids computing the correlated packet loss costs at both layers separately and leads to a constrained MDP formulation of the designed problem. Concurrently, the stationary probability $p(B_n - \Psi(U_n) = k)$ required in Equation (16) of Reference [7] is not needed.

4.1. Mathematical formulation

Let us consider a Poisson arrival process at the transmitter's MAC, which is given as:

$$p\{A(\tau) = i\} = \frac{e^{-\lambda\tau}(\lambda\tau)^i}{i!} \quad i = 0, 1, \cdots \tag{3}$$

where $A(\tau) = i$ denotes the event that i arrivals occur within time duration of τ and λ is the average arrival rate in packets/s. Let J_n be the total energy consumed by the mobile for packet transmission[§] in time frame $n = 1, 2, \cdots$, *i.e.*, during time $[(n-1)T_f, nT_f)$. Since the arrival process and channel variation are mutually independent and ergodic, the long-term average power consumption ψ_{ave} by process $\{J_n\}$ is then computed as:

$$\psi_{ave} = \lim_{m\to\infty} \frac{1}{mT_f} \sum_{n=1}^{m} J_n = \frac{1}{T_f}\mathbb{E}(J_n) \tag{4}$$

Therefore, minimizing ψ_{ave} is equivalently to minimize $\mathbb{E}(J_n)$, with the same average delay and throughput constraints. This average energy minimization problem can be formulated as the optimization of an infinite horizon constrained MDP and the optimal policy μ^* for such a problem is randomized [15]. Namely, the action a taken in state $s\in\mathcal{S}$ is probabilistically selected over its action space $\mathcal{A}(s)$ with distribution function $\theta^*(a|s)$, where $\mathcal{S}$ denotes the entire state space of the system. Let $\pi^*(a, s)$ be the stationary probability driven by policy μ^* of taking $a\in\mathcal{A}(s)$ and the system is in state $s\in\mathcal{S}$, and $\pi^*(s)$ its marginal distribution with respect to s, then μ^* is obtained by defining

$$\theta^*(a|s) = \frac{\pi^*(a, s)}{\pi^*(s)} = \frac{\pi^*(a, s)}{\sum_{a\in\mathcal{A}(s)} \pi^*(a, s)} \tag{5}$$

For the aforementioned energy minimization problem[‖], we define the state space as $\mathcal{S} \triangleq \mathcal{K} \times \mathcal{Q}$, where $\mathcal{K}$ and $\mathcal{Q}$ are the sets of channel states and possible queue sizes in *packets* at the transmitter's MAC, respectively. The AMC modes applicable to state $s\in\mathcal{S}$ then constitute the action space $\mathcal{A}(s)$. The performance constraints are given by the MAC-PHY joint packet loss probability ξ, which implies that the long-term MAC layer throughput is $\lambda(1-\xi)$ packets/s, and the transmitter side MAC layer average queuing delay $\mathcal{D}$ measured in time frames. From Equation (5) it is known that to decide $\theta^*(a|s)$ and thereby the optimal randomized policy μ^* is equivalent to know each $\pi^*(a, s)$. Given performance constraints as $\xi \le \varepsilon$ and $\mathcal{D} \le \delta$, therefore, the quantities to be decided are $\pi^*(a, s)$ for each $a\in\mathcal{A}(s)$ and $s\in\mathcal{S}$, that collaboratively minimize $\mathbb{E}(J_n)$ and satisfies the following

(1) *performance constraints*:
  (a) average throughput

$$\chi^* \ge \lambda(1-\varepsilon) \tag{6}$$

  (b) average packet queuing delay

$$\mathcal{D}^* \le \delta \tag{7}$$

(2) *causality constraints*: [7,15]
  (a) unity property

$$\sum_{a\in\mathcal{A}(s),\, s\in\mathcal{S}} \pi^*(a, s) = 1 \tag{8}$$

  (b) balance property

$$\sum_{s'\in\mathcal{S}} \sum_{a'\in\mathcal{A}(s')} \pi^*(a', s') p^{a'}_{s';s} = \sum_{a\in\mathcal{A}(s)} \pi^*(a, s) \quad \forall s\in\mathcal{S} \tag{9}$$

  (c) nonnegativity

$$\pi^*(a, s) \ge 0 \quad \forall a\in\mathcal{A}(s),\ \forall s\in\mathcal{S} \tag{10}$$

where $p^a_{i;j}$ denotes the transition probability from state $i\in\mathcal{S}$ to state $j\in\mathcal{S}$ when AMC mode $a\in\mathcal{A}(i)$ is applied.

4.2. Objective solution

The above linearly constrained MDP problem can be solved by solving an associated LP problem [15], which can be presented in matrix form as:

$$\text{Minimize} \quad z = \mathbf{c}^T\mathbf{x} \text{Subject to} \quad \mathbf{A}\mathbf{x} \le \mathbf{b}, \mathbf{A}'\mathbf{x}=\mathbf{b}', \mathbf{x} \ge 0$$

where z is the *objective function* to be optimized, vector $\mathbf{c}$ the *cost vector*, vector $\mathbf{x}$ the *decision vector*, respectively. Matrix $\mathbf{A}$ and vector $\mathbf{b}$, matrix $\mathbf{A}'$ and vector $\mathbf{b}'$ define the inequality and equality constraints, respectively. LP problems can be solved efficiently by the simplex method or the interior point method [16]. The major challenge for above stated power-efficient transmission scheduling design lies in formulating various performance metrics of the particular problem into the standard LP form.

5. POWER EFFICIENT TRANSMISSION SCHEDULING

From previous discussion it is clear that the power-efficient transmission scheduling is achieved by applying the optimal randomized policy μ^*, which is available by solution of the associated LP problem presented in subsection 4.2. To form such an LP problem, it is thus equivalent to define the components of $\mathbf{A}$, $\mathbf{A}'$ matrixes and $\mathbf{c}$, $\mathbf{x}$, $\mathbf{b}$, $\mathbf{b}'$ column vectors in the LP problem. We next derive these elements of the LP in context of the packet transmission system given in Section 3.

[§] Since the frame control section is fixed-length and transmitted with constant power, we omit the energy consumption by this part in the following discussion.

[‖] It is helpful to note that for an ergodic process as the problem considered in this study, the denominator $\pi^*(s)$ in Equation (5) is non-zero in any policy design, which otherwise implies the existence of transient states.

5.1. Reward, cost, and state transition probabilities

We first define the *reward* (*i.e.*, expected number of packets that will be correctly received at the receiver's MAC) and *cost* (*i.e.*, total energy consumption) in each time frame, as well as the state transition probability $p^a_{(k,q);(k',q')}$, when the system is in state $s_{(k,q)}$ and AMC mode $a \in \mathcal{A}(s_{(k,q)})$ is selected. Here, $k = 1, 2, \cdots, K$ and $q = 0, 1, \cdots, B$ denote the instantaneous channel state and transmitter side MAC buffer occupancy, respectively, and B is the capacity of the transmitter's MAC buffer in *packets*. Assuming mutually independent packet loss over the channel, then depending on the set $\{a, k, q\}$ for a particular time frame n, the reward $\mathcal{R}_n(a, k, q)$ is computed as:

$$\mathcal{R}_n(a,k,q) = \sum_{l=0}^{\nu_n} l \binom{l}{\nu_n} [1-\mathrm{PER}_n(a,k)]^l \mathrm{PER}_n^{\nu_n - l}(a,k) \tag{11}$$

where ν_n, the number of packets transmitted in time frame n, is determined as $\nu_n = \min(q, \varphi^a_{\max})$, with $\varphi^a_{\max}$ denoting the maximum number of packets that can be serviced by AMC mode a within T_f. The packet error rate $\mathrm{PER}_n(a, k)$ in Equation (11) is determined by $\{a, k\}$ as:

$$\mathrm{PER}_n(a,k) = \frac{1}{P(k)} \int_{\Gamma_{k-1}}^{\Gamma_k} \mathrm{PER}^{(a)}\left[\gamma, \psi(a,k)\right] p_\gamma(\gamma)\,\mathrm{d}\gamma \tag{12}$$

where $\mathrm{PER}^{(a)}\left[\gamma, \psi(a, k)\right]$ denotes the PER by applying AMC mode a with transmission power level $\psi(a, k)$ when the detected SNR level by pilot symbols is γ, and probability $P(k)$ is computed using Equation (1) as:

$$P(k) = \int_{\Gamma_{k-1}}^{\Gamma_k} p_\gamma(\gamma)\mathrm{d}\gamma = \frac{\Gamma\left(m, \frac{m\Gamma_{k-1}}{\rho}\right) - \Gamma\left(m, \frac{m\Gamma_k}{\rho}\right)}{\Gamma(m)} \tag{13}$$

where

$$\Gamma(m,x) = \int_x^\infty t^{m-1} e^{-t}\mathrm{d}t \tag{14}$$

is the complementary incomplete Gamma function [17].

Depending on the set $\{a, k, q\}$ for a particular time frame n and noticing that only data-carrying portion of the time frame consumes energy (see Figure 2), the cost $\mathcal{C}_n(a, k, q)$ is given as:

$$\mathcal{C}_n(a,k,q) = \left\lceil \frac{L_p \nu_n}{\eta^{(a)}} \right\rceil \frac{1}{R_s} \psi(a,k) \tag{15}$$

where we denote by $\eta^{(a)}$ the symbol efficiency of AMC mode a in *bits/symbol*, L_p the packet size in bits and R_s the PHY layer symbol rate in *baud*, respectively.

The state transition probability from $s_{(k,q)}$ to $s_{(k',q')}$ by applying AMC mode $a \in \mathcal{A}(s_{(k,q)})$, *i.e.*, $p^a_{(k,q);(k',q')}$, is defined according to the set $\{a, k, q, k', q'\}$ as:

(1) if $|k-k'| < 2$ and $q - \min\left(q, \varphi^a_{\max}\right) \le q' < B$, then

$$p^a_{(k,q);(k',q')} = p\left\{A(T_\mathrm{f}) = q' - \left[q - \min\left(q, \varphi^a_{\max}\right)\right]\right\} \times p_{k;k'} \tag{16}$$

(2) if $|k-k'| < 2$ and $q - \min\left(q, \varphi^a_{\max}\right) \le q' = B$, then

$$p^a_{(k,q);(k',q')} = \sum_{j=B-\left[q-\min\left(q,\varphi^a_{\max}\right)\right]}^{\infty} p\{A(T_\mathrm{f}) = j\} \times p_{k;k'} \tag{17}$$

(3) for any other sets of $\{a, k, q, k', q'\}$, $p^a_{(k,q);(k',q')} = 0$.

These definitions are based on the following facts:

(1) channel state transitions only occur between neighboring states, as given in subsection 3.2;
(2) in any time frame n, $\nu_n = \min\left(q, \varphi^a_{\max}\right)$ packets are serviced;
(3) for any state transition, the ending queue state q' is no less than $q - \nu_n$;
(4) packet arrival process can be viewed as independent from channel fading dynamic.

The channel state transition probability $p_{k;k'}$ in Equations (16) and (17) can be found in Reference [14].

5.2. Forming LP elements

As discussed in subsection 4.1, the multiple variables to be concurrently decided are the stationary probabilities $\pi(a, k, q)$, where $a \in \mathcal{A}(s_{(k,q)})$, $k \in \mathcal{K}$, $q \in \mathcal{Q}$, *i.e.*,

$$\mathbf{x} = \left[\pi\left(\mathcal{A}(s_{(1,0)}), 1, 0\right), \cdots, \pi\left(\mathcal{A}(s_{(K,B)}), K, B\right)\right]^T \tag{18}$$

with each $\pi\left(\mathcal{A}(s_{(k,q)}), k, q\right)$ being the row vector consisting of all $\pi(a, k, q)$ items where $a \in \mathcal{A}(s_{(k,q)})$.

The objective function z to be minimized, *i.e.*, $\mathbb{E}(J_n)$ defined in subsection 4.1, can be computed by Equation (15) as:

$$z = \mathbb{E}(J_n) = \sum_{\substack{a \in \mathcal{A}(s_{(k,q)}) \\ k \in \mathcal{K},\ q \in \mathcal{Q}}} \mathcal{C}_n(a,k,q)\pi(a,k,q) \tag{19}$$

Therefore, the cost vector $\mathbf{c}$ is given as:

$$\mathbf{c} = \left[\mathbf{C_n}\left(\mathcal{A}(s_{(1,0)}), 1, 0\right), \cdots, \mathbf{C_n}\left(\mathcal{A}(s_{(K,B)}), K, B\right)\right]^T \tag{20}$$

with each $\mathbf{C_n}\left(\mathcal{A}(s_{(k,q)}), k, q\right)$ being the row vector consisting of all $\mathcal{C}_n(a, k, q)$ items where $a \in \mathcal{A}(s_{(k,q)})$.

Following the same manner, matrix $\mathbf{A}$ and vector $\mathbf{b}$ that define the inequality constraints given in Equations (6) and (7) are obtained, respectively, as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \tag{21}$$

where $\mathbf{w}_1$ and $\mathbf{z}_1$ are given respectively as:

$$\mathbf{w}_1 = -\left[\mathbf{R_n}\left(\mathcal{A}(s_{(1,0)}), 1, 0\right), \cdots, \mathbf{R_n}\left(\mathcal{A}(s_{(K,B)}), K, B\right)\right] \tag{22}$$

$$\mathbf{z}_1 = -\lambda(1-\varepsilon)T_{\mathrm{f}} \tag{23}$$

Now we consider the Little's Theorem [18] that translates the delay constraint in Equation (7) as:

$$\mathcal{D} = \frac{\overline{q}}{\lambda_q T_{\mathrm{f}}} \leq \delta \tag{24}$$

where $\overline{q}$ and λ_q denote the average queue size and average enqueued arrival rate (*i.e.*, exclusive of the overflowed packets) in packets/s, respectively. The value of $\overline{q}$ can be obtained by marginal distribution of the buffer occupancy as:

$$\overline{q} = \sum_{q \in \mathcal{Q}} q \times \sum_{k \in \mathcal{K}} \sum_{a \in \mathcal{A}(s_{(k,q)})} \pi(a, k, q) \tag{25}$$

To form $\mathbf{w}_2$ in Equation (21), define a zero-one matrix $\mathbf{e}_k$ for states $s_{(k,q)} (q \in \mathcal{Q})$ as:

$$\mathbf{e}_k = \begin{bmatrix} 1_{1\times\omega(k,0)} & 0 & 0 & 0 \\ 0 & 1_{1\times\omega(k,1)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1_{1\times\omega(k,B)} \end{bmatrix}$$

where $1_{u\times v}$ is a $u \times v$ matrix with all elements are 1 and $\omega(k, q)$ the size of $\mathcal{A}\left(s_{(k,q)}\right)$. Then Equation (25) can be represented as:

$$\overline{q} = \mathbf{Q} \times \Phi_0 \times \mathbf{x} \tag{26}$$

with $\mathbf{Q}$ and Φ_0 being defined as: $\mathbf{Q} = [0, 1, \cdots, B]$, $\Phi_0 = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_K]$. Since the buffer size is finite as B packets, the enqueued arrival rate λ_q in Equation (24) may be less than the actual arrival rate λ, due to buffer overflow. Consider that when the system reaches stable state, the average enqueued arrival rate always matches the average departure rate into the channel. Therefore, we obtain λ_q by evaluating the average service rate κ, which is scheduling dependent and can be computed as:

$$\lambda_q = \kappa = \frac{\mathbf{U} \times \mathbf{x}}{T_{\mathrm{f}}} \tag{27}$$

where $\mathbf{U}$ is in the form of $\mathbf{U} = \left[\mathbf{v}^{\mathcal{A}(s_{(1,0)})}, \mathbf{v}^{\mathcal{A}(s_{(1,1)})}, \cdots, \mathbf{v}^{\mathcal{A}(s_{(K,B)})}\right]$, with each $\mathbf{v}^{\mathcal{A}(s_{(k,q)})}$ being a row vector consisting of all $\nu = \min\left(q, \varphi^a_{\max}\right)$ items, as defined in subsection 5.1, and $a \in \mathcal{A}(s_{(k,q)})$. Substituting Equations (26) and (27) into (24), the terms of $\mathbf{w}_2$ and $\mathbf{z}_2$ in Equation (21) are finally presented as: $\mathbf{w}_2 = \mathbf{Q} \times \Phi_0 - \delta\mathbf{U}$, $\mathbf{z}_2 = 0$.

Next, the balance property given in Equation (9) can be organized in matrix form as: $\mathbf{P} \times \mathbf{x} = \Phi_1 \times \mathbf{x}$, where

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}^{A(s_{(1,0)})}_{(1,0);(1,0)} & \cdots & \cdots & \mathbf{p}^{A(s_{(K,B)})}_{(K,B);(1,0)} \\ \vdots & \mathbf{p}^{A(s_{(1,1)})}_{(1,1);(1,1)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{p}^{A(s_{(1,0)})}_{(1,0);(K,B)} & \cdots & \cdots & \mathbf{p}^{A(s_{(K,B)})}_{(K,B);(K,B)} \end{bmatrix}$$

and

$$\Phi_1 = \begin{bmatrix} \mathrm{e}_1 & 0 & 0 & 0 \\ 0 & \mathrm{e}_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathrm{e}_k \end{bmatrix}$$

with each $\mathbf{p}^{\mathcal{A}(s_{(k,q)})}_{(k,q);(k',q')}$ being a row vector consisting of all $p^a_{(k,q);(k',q')}$ items where $a \in \mathcal{A}(s_{(k,q)})$.

Combining Equations (8) and (9), the matrix $\mathbf{A}'$ and vector $\mathbf{b}'$ that define equality constraints of the LP problem are achieved respectively as:

$$\mathbf{A}' = \begin{bmatrix} \mathbf{P} - \Phi_1 \\ \mathbf{1}_{1\times\Omega} \end{bmatrix} \qquad \mathbf{b}' = \begin{bmatrix} \mathbf{0}_{1\times S} & 1 \end{bmatrix}^T \tag{28}$$

where $\Omega = \sum_{(k,q)\in\mathcal{K}\times\mathcal{Q}} \omega(k, q)$, *i.e.*, the size of $\mathbf{x}$, and $S = K \times (B + 1)$, *i.e.*, the size of $\mathcal{S}$.

Noting that Equation (10) is guaranteed by the definition of $\mathbf{x}$ for LP in subsection 4.2, with $\mathbf{A}$, $\mathbf{A}'$, $\mathbf{b}$, $\mathbf{b}'$, $\mathbf{c}$, and $\mathbf{x}$ obtained above, the LP problem that decides the optimal transmission scheduling design can be solved using the simplex method [16] or other commercial LP packages. The optimal randomized policy μ^* discussed in subsection 4.1 is finally obtained by applying the solution of $\mathbf{x}$ into Equation (5).

6. DESIGN VERIFICATION AND ANALYSIS

In this section, we verify the optimal scheduling policy design proposed in Section 5 with Monte Carlo simulations, and observe its properties with further analysis.

6.1. Simulation setup

In the simulation we apply the five convolutionary coded M_n-ary rectangular quadrature-amplitude modulation (QAM) transmission modes as the TM2 defined in Reference [14], with presence of additive white Gaussian noise (AWGN) to represent multiple AMC modes. In rate

ascending order, these AMC modes are referred to as mode 1 to mode 5, respectively. Moreover, there is a *null* mode, referred to as mode 0, where no packet transmission takes place due to deteriorated channel condition or empty buffer. The average PER by different transmission modes of TM2 over AWGN is approximated as [[14], Equation (3)]:

$$PER_{\hat{n}}(\varrho) \approx \begin{cases} 1, & 0 < \varrho < \varrho_{p\hat{n}} \\ \alpha_{\hat{n}} \exp\left(-g_{\hat{n}}\varrho\right), & \varrho \geq \varrho_{p\hat{n}} \end{cases} \quad (29)$$

where $\hat{n} = 1, \cdots, 5$ is the transmission mode index and ϱ is the instantaneous received SNR per symbol. The corresponding value of fitting parameters $\alpha_{\hat{n}}$, $g_{\hat{n}}$, and $\varrho_{p\hat{n}}$ for packet length $L_{\mathrm{p}} = 1080$ bits are given by Table II in Reference [14] and applied in our simulations.

Let ψ_0 be the constant power level for transmitting pilot symbols. The SNR level ϱ for receiving data packets transmitted by AMC mode a and power level $\psi(a,k)$, when SNR level γ is detected by pilot symbols in the frame control section, is then determined as $\varrho = \frac{\psi(a,k)}{\psi_0}\gamma$. . Without loosing generality, we can assume $\psi_0 = 0$ dB. Therefore, the value of $\mathrm{PER}_n(a,k)$ in Equation (12) is obtained by substituting $PER_{\hat{n}}(\varrho)$ determined in Equation (29) with $\varrho = \psi(a,k)\gamma$. In the simulation, we set the value of $\psi(a,k)$ as the power level to achieve average PER of 10^{-3} (denoted as $P_{\mathrm{PHY}} = 10^{-3}$)¶ by Equation (29) at the lower bound SNR threshold (detected by pilot symbols) of FSMC state k, *i.e.*, Γ_{k-1}, with AMC mode a.

The channel status is partitioned into seven FSMC states, where state $k = 1$ corresponding to SNR range $\gamma \in [0, \Gamma_1)$ is not used for data transmission to avoid deep channel fades. The other six channel states are partitioned over $\gamma \in [\Gamma_1, +\infty)$ by the equal probability partition method given in Reference [13]. We set the value of Γ_1 as the value of $\varrho_{p\hat{n}}$ defined in Equation (29) for transmission mode 1. With these definitions, the action space for each system state $s_{(k,q)} \in \mathcal{S}$ is then decided as:

$$\mathcal{A}\left(s_{(k,q)}\right) = \begin{cases} \text{mode } 0 & (k = 1 \text{ or } q = 0) \\ \text{mode } 0,\ 1,\ \cdots,\ 5 & (k > 1,\ q > 0) \end{cases} \quad (30)$$

In the simulation we set the average queuing delay constraint as $\delta = 2$ time frames and average packet loss rate constraint as $\varepsilon = 10^{-1}$. Other relevant parameters used for simulation are listed in Table I.

Table I. Simulation Parameters.

ρ (pilot symbols)	15 dB
f_m	10 Hz
T_{f}	1 ms
λ	$10^3 - 8 \times 10^3$ packets/s
B	15 packets
L_p	1080 bits
Symbol rate	2.26 MBaud (2260 symbol/frame)
Frame control	100 symbol/frame

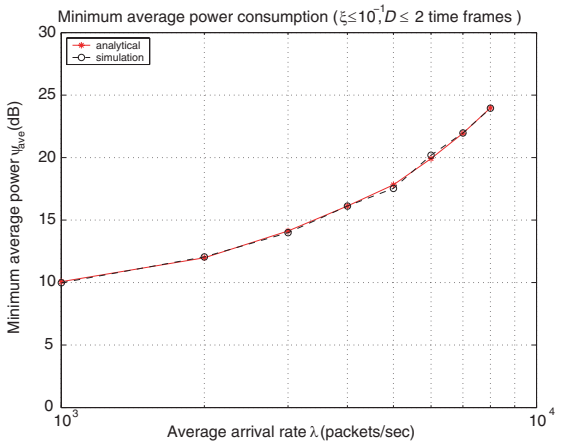

Figure 3. Minimum power consumption by optimal scheduling policy.

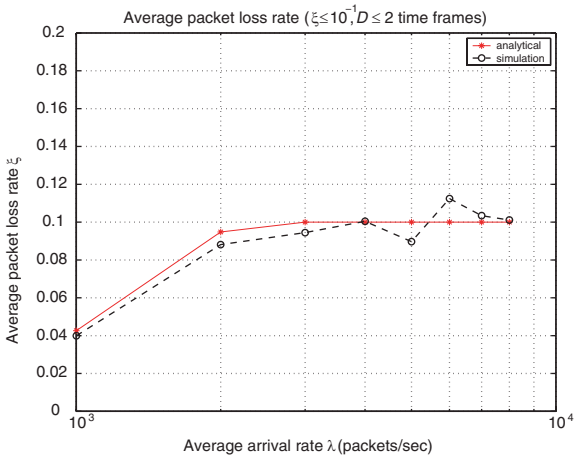

Figure 4. Average packet loss rate by optimal scheduling policy.

6.2. Results and discussion

Figures 3–5 show the minimum average power ψ_{ave}, average packet loss rate ξ and average packet queuing delay $\mathcal{D}$, respectively, by computing the proposed solution in Section 5 and by simulating the generated optimal policy given in Equation (5). It is seen in the figures that analytical computations and simulation experiments agree on the same minimal average power consumption while both required performance metrics, *i.e.*, ξ and $\mathcal{D}$, are constantly below the specified levels. With successful verification of the proposed solution, the following discussion is only based on analytical computations.

¶ Note that this value only serves to define the power level of $\psi(a,k)$. The packet loss constraint ξ is defined across MAC and PHY layers.

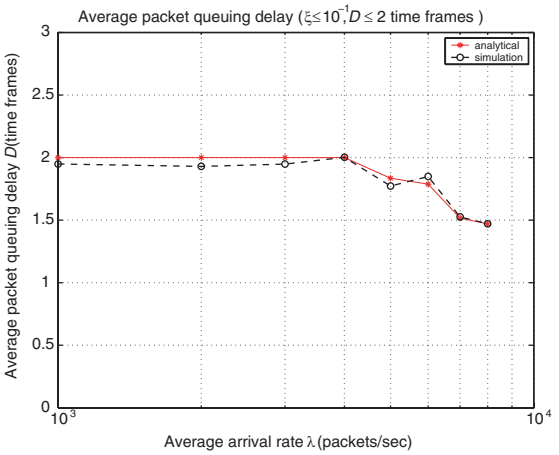

Figure 5. Average packet queuing delay by optimal scheduling policy.

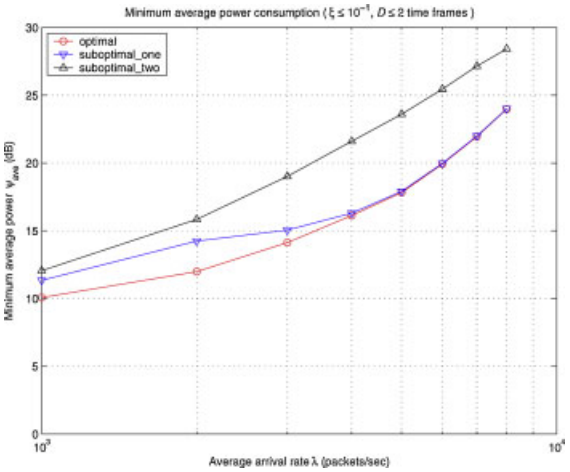

Figure 6. Minimum power consumption by optimal and two suboptimal scheduling policies.

6.2.1. Cross-layer dynamics.

It is interesting to note that in Figures 4 and 5 though the optimal transmission scheduling policy constantly maintains the two performance constraints under their specified level, in Figure 4 the average packet loss rate drops off 10^{-1} at light load points ($\lambda < 3 \times 10^3$ packets/s), while in Figure 5 the average packet queuing delay fades away from 2 time frames at heavy load points ($\lambda > 4 \times 10^3$ packets/s). This behavior reveals the MAC-PHY cross-layer dynamics associated with the control policy. Particularly, when the offered traffic load is light, the very small queue size excludes MAC layer buffer overflow. Hence, packet loss is only incurred by PHY layer channel errors. This allows the scheduler to, for the sake of saving power, avoid using bad channel states and transmit only when good channel states appear, till $\overline{q}$ is enlarged so that further increasing queuing delay is not allowed by the specified delay constraint δ. Therefore, in light load scenario the control policy is uniquely driven by the average queuing delay constraint, as shown in Figure 4. Similarly, when the offered traffic load becomes heavy, packets are pushed to pass the MAC buffer faster in order to alleviate buffer overflow and thereby to maintain the specified packet loss constraint ε. Though the scheduler endeavors to drop more packets at the MAC layer to save power, the value of $\mathcal{D}$ is still pulled down by increased traffic load. Consequently, in heavy load scenario the control policy is uniquely driven by the cross-layer packet loss constraint, as shown in Figure 5.

6.2.2. Optimality comparison.

Next, we continue to see the optimality of the policy determined in Section 5. In this experiment, the optimal policy is compared with two suboptimal policies:

(1) policy suboptimal_one enforces low rate transmission mode (mode 0 or mode 1) when the MAC queue occupancy is less than 4 packets. This policy variant intends to reduce channel errors and still achieve required cross-layer QoS constraints, emulating PHY layer error-free cross-layer power control schemes in the literature (*e.g.*, [6]).

(2) policy suboptimal_two enforces high rate transmission mode (mode 5) when the MAC queue occupancy is more than 11 packets. This policy variant intends to minimize buffer overflow without violating required cross-layer QoS constraints, emulating MAC layer dropping-exempt cross-layer power control schemes in the literature (*e.g.*, [7]).

This can be done by applying another performance constraint in addition to Equations (6) and (7) as $\pi(a, s) = 0$ where $a =$ mode 2, 3, 4, 5 and $s \in s_{(k,q<4)}$ for policy suboptimal_one, and $a =$ mode 0, 1, 2, 3, 4 and $s \in s_{(k,q>11)}$ for policy suboptimal_two.

Figure 6 compares the minimum power consumption by these suboptimal policies and the optimal policy. It is shown that when the offered traffic load increases, the MAC buffer tends to be fully occupied irrespective the scheduling effort and hence policy suboptimal_one becomes identical to the optimal policy. However, since the additional constraint applied in policy suboptimal_two limits the scheduling diversity when the MAC buffer tends to be fully occupied, this affects the scheduling decision even when the buffer is lightly occupied. Therefore, policy suboptimal_two behaves differently from the optimal policy for all loading scenarios. Nevertheless, different performance between the suboptimal policies and the optimal policy is always recognized by more power consumption in the former.

6.3. Boundedness analysis

In the previous sections, we have assumed that the performance constraints on average packet queuing delay and average packet loss rate are appropriately given and achievable. In this section, we investigate the boundedness of these

two constraints and to define their feasible range for a particular system setting. We first define the feasible range of $\mathcal{D}$ without constraining on ξ. Then for each feasible delay constraint value δ, the lower bound of delay limited average packet loss rate, *i.e.*, ξ_L, can be found[#].

6.3.1. Unconstrained delay bound.

When only the average packet queuing delay $\mathcal{D}$ is considered, it is apparent that the minimum value of $\mathcal{D}$ can be achieved by applying the highest rate transmission mode $a_H(k,q) \in \mathcal{A}\left(s_{(k,q)}\right)$ for each state $s_{(k,q)}$, which in the experimental setup of subsection 6.1 means $a_H(k,q) =$ mode 0 for $k = 1$ or $q = 0$, and $a_H(k,q) =$ mode 5 otherwise. This leads to an unconstrained MDP formulation driven by deterministic policy μ_H that applies $a_H(k,q)$ for each state $s_{(k,q)} \in \mathcal{S}$. The lowest delay bound $\mathcal{D}_L$ can be computed by Equation (24) as

$$\mathcal{D}_L = \frac{\sum_{i=1}^{K}\sum_{j=1}^{B}(j \times \sigma_{(i,j)})}{\sum_{i=2}^{K}\sum_{j=1}^{B}\left(\min\left(j, \varphi_{\max}^{a_H(i,j)}\right) \times \sigma_{(i,j)}\right)} \tag{31}$$

where $\sigma \triangleq \left\{\sigma_{(k,q)} | k \in \mathcal{K}, q \in \mathcal{Q}\right\}$ denotes the stationary distribution of the resultant Markov chain and can be obtained from the state transition probability matrix of the Markov chain as given in [[19], Proposition 1].

6.3.2. Delay limited packet loss bound.

For a given value of $\delta \in [\mathcal{D}_L, +\infty)$, $\xi_L(\delta)$ can be solved by minimizing ξ with the constraint on average queuing delay, using the same LP method discussed in Section 5. For example, removing $\mathbf{w}_1$ and $\mathbf{b}_1$ from Equation (21) and setting $\mathbf{w}_1$ as the objective function of the LP, $\xi_L(\delta)$ is obtained as:

$$\xi_L(\delta) = 1 + \frac{\chi_H(\delta) * T_{\mathrm{f}}}{\lambda} \tag{32}$$

where $\chi_H(\delta) = arg_{\mathbf{x}} \min(\mathbf{w}_1 \times \mathbf{x})$, subject to Equations (7–10).

6.3.3. Determining instantaneous power levels.

The above boundedness analysis can be used to determine the instantaneous power level $\psi(a,k)$ applied to each transmission mode and channel state. Namely, the selection of $\psi(a,k)$ should ensure that the required constraints on average queuing delay and packet loss rate are achievable. Assuming the set of available power levels for a particular system is $\mathcal{T}$, to determine the optimal combination of $\psi(a,k)$ that can satisfy the required delay and packet loss constraints, the following procedure applies:

(1) for the given performance constraints $\mathcal{D} < \delta$ and $\xi < \varepsilon$, find subset $\hat{\mathcal{T}} \in \mathcal{T}$ that achieves $\mathcal{D}_L \leq \delta$ and $\xi_L(\delta) \leq \varepsilon$ with selected combinations of $\psi(a,k)$;
(2) for the estimated traffic load, find the optimal combination of $\psi(a,k)$ within $\hat{\mathcal{T}}$ as $\tau = \arg_{\tau \in \hat{\mathcal{T}}} \min(\psi_{\mathrm{ave}})$ using the proposed solution in Section 5.

7. CONCLUSION

In this paper, we have studied the power-efficient transmission scheduling from a mobile station to the base station over wireless fading channel. The detailed solution of applying classical MDP and LP theories to such a packet transmission system is presented. Alongside minimizing the average power consumption, the proposed solution allows dynamical balancing PHY and MAC layer behaviors to achieve required cross-layer performance. Finally, the designed transmission scheduling scheme is experimentally verified and investigated with more insightful analysis.

REFERENCES

1. Bai X, Shami A, Primak SL. Optimal power control over fading channel with cross-layer performance constraint. *Proceedings IEEE ICC* 2008; 2265–2269.
2. Kim DI, Hossain E, Bhargava VK. Dynamic rate and power adaptation for provisioning class-based QoS in cellular multirate WCDMA systems. *IEEE Transactions on Wireless Communications* 2004; **3**(5): 1590–1601.
3. Berry RA, Gallager RG. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory* 2002; **48**(5): 1135–1149.
4. Bettesh I, Shamai S. Optimal power and rate control for fading channels. *IEEE VTC* 2001; **2**: 1063–1067.
5. Goyal M, Kumar A, Sharma V. Power constrained and delay optimal policies for scheduling transmission over a fading channel. *IEEE INFOCOM* 2003; **1**: 311–320.
6. Wang H, Mandayam NB. A simple packet-transmission scheme for wireless data over fading channels. *IEEE Transactions on Communications* 2004; **52**(7): 1055–1059.
7. Karmokar AK, Djonin DV, Bhargava VK. Optimal and suboptimal packet scheduling over correlated time varying flat fading channels. *IEEE Transactions on Wireless Communications* 2006; **5**(2): 446–456.
8. Tang J, Zhang X. Quality-of-service driven power and rate adaptation over wireless links. *IEEE Transactions on Wireless Communications* 2007; **6**(8): 3058–3068.
9. Wu D, Negi R. Effective capacity: a wireless link model for support of quality of service. *IEEE Transactions on Wireless Communications* 2003; **2**(4): 630–643.

[#] It is obvious that here only the lower bound of $\mathcal{D}$ and ξ are interested.

10. Goldsmith AJ, Varaiya PP. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory* 1997; **43**(6): 1986–1992.
11. Goldsmith AJ, Chua S-G. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications* 1997; **45**(10): 1218–1230.
12. Simon MK, Alouini M-S. Digital Communication over Fading Channels: A Unified Approach to Performance Analysis, Proakis JG (ed). John Wiley & Sons: New York, 2000.
13. Wang HS, Moayeri N. Finite-state markov channel–a useful model for radio communication channels. *IEEE Transactions on Vehicular Technology* 1995; **44**(1): 163--171.
14. Liu Q, Zhou S, Giannakis GB. Queuing with adaptive modulation and coding over wireless links cross-layer analysis and design. *IEEE Transactions on Wireless Communications* 2005; **4**(3): 1142–1153.
15. Puterman M. Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons: New York, 1994.
16. Sposito VA. Linear and Nonlinear Programming, Ames, Iowa 50010: The Iowa State University Press, 1975.
17. Abramowitz M, Stegun IA (eds). Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables, U.S. Govt. Print. Off: Washington, 1970.
18. Bertsekas D, Gallager R. Data Networks (2nd edn), Prentice Hall: Englewood Cliffs, N.J, 1992.
19. Bai X, Shami A. Two dimensional cross-layer optimization for packet transmission over fading channel. *IEEE Transactions on Wireless Communications* 2008; **7**(10): 3813–3822.

AUTHORS' BIOGRAPHIES

Xiaofeng Bai received the B.E. degree in electrical engineering from Shandong University, China, in 1997, and the Ph.D. degree in electrical and computer engineering from The University of Western Ontario, Canada, in 2009. He is currently a research scientist at Motorola Inc., U.S.A. His research interests are in the area of quality-of-service in broadband access networks, cross-layer protocol design in wireless communication systems, radio resource control in LTE networks.

Abdallah Shami received the B.E. Degree in Electrical and Computer Engineering from the Lebanese University, Beirut, Lebanon in 1997, and the Ph.D. Degree in Electrical Engineering from the Graduate School and University Center, City University of New York, New York, NY in September 2002. In September 2002, he joined the Department of Electrical Engineering at Lakehead University, Thunder Bay, ON, Canada as an Assistant Professor. Since July 2004, he has been with The University of Western Ontario, London, ON, Canada where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. His current research interests are in the area of wireless/optical networking.

Serguei Primak graduated from St. Petersburg University of Telecommunications, Russia (MEngSci, 1991) and completed his Ph.D. at the department of Electrical and Computer Engineering, Ben-Gurion University of the Negev (1997). Currently, he is Associate Professor at The University of Western Ontario. His area of interests lies in modeling and simulation of wireless communications channels, non-Gaussian and non-rayleigh random processes, stochastic differential equations.