Channel-Quality-Based Opportunistic Scheduling with ARQ in Multi-Rate Wireless Networks: Modeling and Analysis

Teerawat Issariyakul, *Student Member, IEEE* and Ekram Hossain, *Senior Member, IEEE*

***Abstract*— In this paper, we develop a novel framework for analyzing radio link level performance for opportunistic scheduling with automatic repeat request (ARQ)-based error control in multi-rate wireless networks. The multi-rate transmission is assumed to be achieved through adaptive modulation and coding (AMC) to adjust the transmission rate according to the channel condition. The residual error effect due to each AMC setting is counteracted by means of a limited persistence ARQ protocol. The novelty of the proposed analytical framework lies in the fact that we are able to derive complete statistics (in terms of *probability mass function*) for both short-term and long-term performance measures such as system throughput, per-flow throughput, inter-success delay under both uncorrelated and correlated wireless channels. These performance measures can also be obtained in case of non-identical channels for different users. Analytical results are validated through simulations and the impacts of channel behavior on the different radio link level performance metrics are investigated.**

***Index Terms*— Finite state Markov channel, opportunistic scheduling, adaptive modulation and coding (AMC), discrete-time Markov chain (DTMC), automatic repeat request (ARQ).**

I. INTRODUCTION

IN A MULTIUSER cellular wireless network, the limited available radio resources must be allocated among all mobiles in the most effective manner. Opportunistic scheduling is a scheduling algorithm which exploits the time-varying nature of a wireless channel. A class of opportunistic scheduling which allows only one-by-one transmission rather than simultaneous transmissions can provide high average network throughput in a wireless network by exploiting the gain due to multiuser diversity [1], which depends on the asynchronous channel variations among mobile users. This type of opportunistic scheduling was shown to maximize network capacity when the network is not limited by available rate set and/or transmission power [2]. However, when the available rate set is finite and/or the transmission power is limited, multiple simultaneous transmissions (e.g., in a code division multiple access (CDMA) system) may maximize the capacity (i.e., increase the *frame fill efficiency* [3]).

That the opportunistic scheduling can maximize a wireless system performance stochastically even under certain resource allocation fairness constraint, was proven in [4] using the notion of 'utility'. Simulation-based forward link data throughput performance in the Qualcomm CDMA-HDR (High Data Rate) system [5], which uses an opportunistic scheduling scheme based on the *proportional fairness (PF)* criterion, was presented in [6]. The PF algorithm was designed to share the wireless channel resources fairly as well as to maximize channel throughput for best-effort data services. Scheduling mechanisms such as *exponential (EXP) rule* [7] and *modified largest weighted delay first (M-LWDF)* [8] were proposed for quality of service (QoS)-sensitive data service. While most of the works on opportunistic scheduling aimed at enhancing the scheduling algorithm in different ways [9]-[12] (and the references therein), little attention has been paid on modeling and analyzing the basic scheduling mechanism under different channel dynamics and its impact on overall radio link level performance.

Again, to effectively utilize the scarce radio bandwidth, adaptive modulation and coding (AMC), which has been adopted in systems such as 3GPP, 3GPP2, IEEE 802.16, IEEE 802.11 and HIPERLAN/2, can be used to adjust the modulation index as well as the strength of the error correcting codes according to the current value of signal-to-noise-plus-interference ratio (SINR) at the receiver [13], [14]. Since in general, AMC is not designed for absolute integrity, reliability can be provided at the radio link level by using ARQ-based error control mechanism which invokes a retransmission procedure in case of transmission failure.

In practice, choices of AMC adjustment correspond to a set of non-continuous transmission rates. In most AMC implementations, the wireless channel is modeled by a *finite state Markov channel* (FSMC) [15], [16] which divides the entire range of SINR into C non-overlapping intervals. For each interval, an AMC is selected to satisfy the target packet error rate constraint. For a single user system (i.e., without scheduling), the effects of traffic source on AMC-based system parameters were studied in [13] and [17]. The effects of different ARQ policies on radio link level buffer management were investigated in [14] and [18] considering a general FSMC model and a *Gilbert-Elliot* channel (i.e., an FSMC with two states), respectively. However, the problem of scheduling was

Manuscript received July 31, 2004; revised December 22, 2004 and February 19, 2005; accepted March 24, 2005. The associate editor coordinating the review of this paper and approving it for publication was M. Zorzi.

E. Hossain is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada R3T 5V6 (e-mail: ekram@ee.umanitoba.ca).

T. Issariyakul was with the Department of Electrical and Computer Engineering, University of Manitoba and is currently with the Department of Electrical and Computer Engineering, University of British Columbia, Canada (e-mail: teerawat@ece.ubc.ca).

Digital Object Identifier 10.1109/TWC.2006.04014.

not addressed in these works.

This paper presents a novel analytical framework to evaluate radio link level performance. The framework incorporates channel-quality-based opportunistic scheduling mechanism (and also round-robin scheduling), ARQ-based error recovery, and multi-rate transmission under both correlated and time-independent wireless channels. We determine the *probability mass function*s (pmfs) of performance metrics including system and per-flow throughput, inter-access delay, inter-success delay, and connection reset delay We present simulation results which validate the numerical results obtained from the analytical model and we investigate the impacts of system and channel parameters on the different radio link level performance measures thoroughly.

The rest of the paper is organized as follows. Section II provides a summary of the system model and the underlying assumptions. The analytical framework is presented in Section III. The numerical and the simulation results as well as their useful implications are presented in Section IV. Finally, conclusions are stated in Section V.

II. System Model and Assumptions

A. System Description

We consider channel-quality-based opportunistic scheduling for a multi-rate time-division multiplexing (TDM) cellular wireless network. During a fixed-size time slot, the base station transmits/receives data to/from a mobile perceiving the best channel condition. We assume perfect channel state information at the base station, which might be achieved by a training-based channel estimation [19]. We also assume continuously backlogged data flows corresponding to each mobile. Since our main focuses are on the inter-relationship among channel parameters, scheduling mechanism, AMC, and ARQ, and on their impact on radio link level throughput and delay performance, we do not address the issues related to radio link level buffer management in this paper.

B. Wireless Channel Model

In this paper, the wireless channel is considered to be *discrete* where the received SINR is divided into C intervals. Each interval corresponds to a 'channel state' ($\mathscr{C} \in \{1, 2, \cdots, C\}$), which is assumed to remain unchanged during one time slot [14]-[17]. When the channel state is c, a maximum of r packets can be transmitted during a time slot such that the packet error rate does not exceed $p_{err}^{(c)}$. Although we assume that $r = c$ in this paper, other channel-to-rate mapping functions (e.g., those in [13]) can be modeled in a similar manner. Hereafter, we will refer to higher channel states which can accommodate higher transmission rates as *good* or *better* channel states.

An FSMC is a discrete channel model which captures correlated channel fading characteristics by using a discrete-time Markov chain (DTMC). An FSMC is completely described by ($\boldsymbol{\alpha}$,$\mathbf{P}$), where $\boldsymbol{\alpha}$ is an initial probability row vector and $\mathbf{P}$ is the transition probability matrix (TPM)[1] for which the entry (i,j) is denoted by p_{ij}.

[1]Throughout this paper, we use regular and boldface type letters to represent scalar values and matrices, respectively.

Due to the time-correlation property of a DTMC, the channel state of an FSMC at time t depends on that at time $t-1$. Mathematically, $\boldsymbol{f}_{\mathscr{C}}^{(t)} = \boldsymbol{f}_{\mathscr{C}}^{(t-1)} \cdot \mathbf{P}$, where $\boldsymbol{f}_{\mathscr{C}}^{(t)} = [f_{\mathscr{C}}^{(t)}(1), f_{\mathscr{C}}^{(t)}(2), \cdots, f_{\mathscr{C}}^{(t)}(C)]$ and $f_{\mathscr{C}}^{(t)}(c)$ is the probability that the channel state at time t is c. At steady state, $\boldsymbol{f}_{\mathscr{C}}^{(t_s)} = \boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_C]$, where π_i is the steady-state probability that the channel is in state i and t_s is time at which the FSMC reaches the steady state. Due to the limiting behavior of an ergodic DTMC, $\boldsymbol{f}_{\mathscr{C}}^{(t_s+T)} = \boldsymbol{\pi} \cdot \mathbf{P}^T = \boldsymbol{\pi}, T \geqslant 0$. In other words, $\boldsymbol{f}_{\mathscr{C}}^{(t)}$ will be time-invariant if we start observing the channel from the point in steady state (i.e., $\boldsymbol{\alpha} = \boldsymbol{f}_{\mathscr{C}}^{(0)} = \boldsymbol{\pi}$), or if we observe the long-term behavior of the channel in the case that $\boldsymbol{f}_{\mathscr{C}}^{(0)} \neq \boldsymbol{\pi}$. For example, long-term expected throughput is averaged over a long period of time, and is therefore, time-invariant. Under these situations, the wireless channel is considered to be a *random state channel* (RSC) which is simply characterized by $\boldsymbol{\pi}$ (i.e., $\boldsymbol{f}_{\mathscr{C}}^{(t)} = \boldsymbol{\pi}, \forall t$).

The approaches for estimating FSMC parameters depend on the underlying physical layer model. For example, estimation of the FSMC parameters for Rayleigh and Nakagami-m fading channels were provided in [15] and [16], respectively. To free our model from physical layer assumptions, we assume that all the stationary parameters are available. For an FSMC, we abstract the channel model by using $\boldsymbol{\pi}$ and average channel state correlation ρ ($= \sum_{i=1}^{C} \pi_i \cdot p_{ii}$). Given $\boldsymbol{\pi}$ and ρ, all the values of $p_{ij} \in [0, 1]$ for an FSMC are calculated by solving the following optimization problem [9]:

$$\min_{p_{ij}} \sum_{i=1}^{C} \pi_i \cdot (p_{ii} - \rho)^2 \tag{1}$$

$$\text{s.t.} \quad \sum_{j=1}^{C} p_{ij} = 1 \ (\forall i), \quad \sum_{i=1}^{C} \pi_i \cdot p_{ij} = \pi_j \ (j \neq 1). \tag{2}$$

Note that, the solutions of p_{ij} for the above optimization problem can be obtained by solving (2) and then choosing among those which satisfy the constraint $p_{ii} = \rho, \forall i$. If such solutions do not exist, then the values of p_{ij} which minimize (1) are chosen in which case all values of p_{ii} will be very close to ρ.

In this paper, we consider both RSC and FSMC models. In an RSC, the channel state in each time slot is randomly selected according to $\boldsymbol{\pi}$. In an FSMC, we assume that the channel state in time slot 0 is known. Channel states in subsequent time slots vary according to $\mathbf{P}$. Again, an FSMC which starts with probability vector $\boldsymbol{\pi}$ is equivalent to an RSC. In presence of n mobiles, we study three following cases:

- **Case I** (*All-RSC*): Channel state for each mobile varies according to $\boldsymbol{\pi}$,
- **Case II** (*All-FSMC*): Channel state for each mobile follows the FSMC model. For these mobiles, we specify $\boldsymbol{\pi}$ as well as ρ, and use (1) and (2) to calculate the TPM ($\mathbf{P}$),
- **Case III** (*FSMC-RSC*): Channel states for n_r and n_f mobiles follow the RSC and the FSMC model, respectively.

C. Opportunistic Scheduling Policy and Automatic Repeat reQuest (ARQ) Mechanism

In each time slot, there are k *eligible* mobiles whose channel states are the best among those of all n mobiles. The base station randomly selects one of the eligible mobiles for transmission in the current time slot. Therefore, the transmission probability for each eligible mobile is $1/k$. If selected, the base station/mobile will transmit only $r(= c)$ packets in the current time slot so that the requirement on the packet error probability can be satisfied.

To combat with the residual error probability, ARQ-based error recovery with limited persistence is employed to retransmit only erroneous packets. For each mobile, a retransmission counter is maintained to keep track of the number of time slots in which all the transmissions have failed. When all the transmitted packets during a time slot are lost, the counter is incremented, and when the counter exceeds a certain limit K, the corresponding connection is reset. The counter is reset to zero when the connection is initiated or reset or when at least one of the transmitted packets is received successfully. By adopting the above ARQ mechanism, a connection tends to be reset only when the mobile is turned off or experiences extremely bad channel condition. In both the cases, data packets in the buffer are subject to extremely long delay and might be discarded during a connection reset process[2].

III. Model Analysis

In this section, we analyze the performance of an opportunistic scheduling under the three above channel assumptions: *All-RSC*, *All-FSMC*, and *FSMC-RSC*. For each case, we divide the analysis into two parts. The first part assumes error-free wireless channel, and derives statistics for the following performance parameters:

- *system throughput* (γ_{sys}) defined as the number of packets successfully transmitted during a time slot,
- *inter-access delay* ($\mathscr{D}_{acc}$) defined as the number of time slots between two channel access opportunities corresponding to the same mobile, and
- *per-flow throughput* (γ_{flow}) defined as the number of packets successfully transmitted to/from a particular mobile per time slot.

In the second part, we introduce non-zero packet error probability to the wireless channel, and use an ARQ mechanism for error recovery at the radio link level. In such a case, at least one packet might be successfully transmitted (sc), or the connection will be reset (rst) after some time due to limited persistence of the ARQ mechanism. For a particular mobile, we measure the conditional delay to the occurrence of either sc or rst as follows:

- *inter-success delay* ($\mathscr{D}_{sc}$) defined as the number of time slots between the points where the retransmission counter is zero and the point where at least one packet is successfully transmitted, given that sc will occur before rst, and
- *connection reset delay* ($\mathscr{D}_{rst}$) defined as the number of time slots between the points where the retransmission counter is zero and the point where the connection is reset, given that rst will occur before sc.

[2]A similar approach is used in IEEE 802.11 [20], where the buffer is flushed after seven transmission failures.

A. Case I: All-RSC

In this section, we assume that channel states of all mobiles are independent and identically distributed (i.i.d.) and are modeled by an RSC model. The results for non-identical channels can also be obtained by the modification suggested in Appendix I.

PROPOSITION 1 *For* **Case I**, *each mobile is scheduled for transmission at rate c in time slot t with probability $p_{tx}^{(t)}(c)$ (in (3)), where $F_c = \sum_{i=1}^{c} \pi_i$ (Appendix I).*

$$p_{tx}^{(t)}(c) = \frac{(F_c)^n - (F_{c-1})^n}{n}. \quad (3) \quad \square$$

When identical to each other, each mobile has the same channel access probability $\sum_{c=1}^{C} p_{tx}^{(t)}(c) = 1/n$. Therefore, the probability that a mobile is not scheduled for transmission is $1 - 1/n$. Since the evolution of an RSC is a memoryless process, $p_{tx}^{(t)}(c) = p_{tx}(c), \forall t$ (i.e., time-invariant).

1) Error-Free Wireless Channel: We assume that $p_{err}^{(c)} = 0$ $(\forall c)$ and derive statistics for γ_{sys}, $\mathscr{D}_{acc}$, and γ_{flow}.

THEOREM 1 *For* **Case I**, *the pmf of system throughput can be calculated from $f_{\gamma_{sys}}(c)$. The joint pmf that a particular mobile acquires channel access at time slot d, and perceives channel state c in that time slot can be calculated from $f_{\mathscr{C},\mathscr{D}_{acc}}(c,d)$, where*

$$f_{\gamma_{sys}}(c) = (F_c)^n - (F_{c-1})^n,$$
$$f_{\mathscr{C},\mathscr{D}_{acc}}(c,d) = \frac{f_{\gamma_{sys}}(c)(n-1)^{d-1}}{n^d}. \quad (4)$$

$\square$

Proof: The probability that the best channel state among all mobiles (and therefore system throughput) is c can be calculated from (5) below

$$\begin{aligned} f_{\gamma_{sys}}(c) &= Pr\{(\exists i : \mathscr{C}_i^{(t)} = c) \; and \; (\mathscr{C}_i^{(t)} < c+1, \forall i)\} \\ &= \left(1 - (1 - f_{\mathscr{C}}^{(t)}(c|c))^n\right) \cdot \left(F_{\mathscr{C}}^{(t)}(c)\right)^n \quad (5) \end{aligned}$$

where $\mathscr{C}_i^{(t)}$ is the channel state of mobile i in time slot t, $f_{\mathscr{C}}^{(t)}(c|c) = f_{\mathscr{C}}^{(t)}(c)/F_{\mathscr{C}}^{(t)}(c)$ and $F_{\mathscr{C}}^{(t)}(c) = \sum_{i=1}^{c} f_{\mathscr{C}}^{(t)}(i)$. The probability $f_{\mathscr{C},\mathscr{D}_{acc}}(c,d)$ that the mobile is not selected for first $d-1$ time slots and is selected in time slot d can be calculated in (6), where $p_{tx}^{(t)} = \sum_{\forall c} p_{tx}^{(t)}(c)$ and $p_{tx}^{(t)}(c)$ can be calculated using (3).

$$f_{\mathscr{C},\mathscr{D}_{acc}}(c,d) = p_{tx}^{(d)}(c) \prod_{t=1}^{d-1} \left(1 - p_{tx}^{(t)}\right) \quad (6) \quad \blacksquare$$

COROLLARY 1 *The statistics of system throughput (γ_{sys}), inter-access delay ($\mathscr{D}_{acc}$), and per-flow throughput (γ_{flow}) for* **Case I** *can be calculated from (7), where $E[\cdot]$ is an expectation function.*

$$E[\gamma_{sys}] = C - \sum_{c=1}^{C-1} (F_c)^n, \quad E[\mathscr{D}_{acc}] = n,$$
$$f_{\mathscr{D}_{acc}}(d) = \frac{(n-1)^{d-1}}{n^d}, \quad E[\gamma_{flow}] = \frac{E[\gamma_{sys}]}{E[\mathscr{D}_{acc}]}. \quad (7)$$

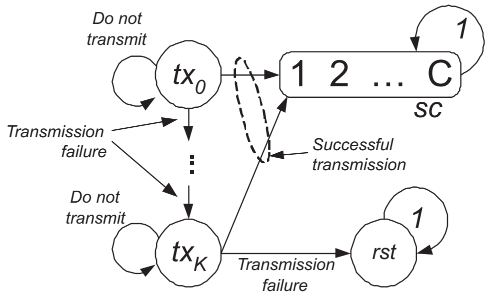

Fig. 1. A DTMC representing retransmission process for **Case I** with ARQ.

Inter-access delay ($\mathscr{D}_{acc}$) is conditioned on that the mobile is granted channel access at time slot 0. Therefore, its *pmf* can be calculated by summing $f_{\mathscr{C},\mathscr{D}_{acc}}(c,d)$ in (4) over all possible channel states, that is, $f_{\mathscr{D}_{acc}}(d) = \sum_{\forall c} f_{\mathscr{C},\mathscr{D}_{acc}}(c,d)$. Now, $E[\gamma_{sys}]$ and $E[\mathscr{D}_{acc}]$ can be obtained by averaging γ_{sys} and $\mathscr{D}_{acc}$ over probabilities $f_{\gamma_{sys}}(c)$ and $f_{\mathscr{D}_{acc}}(d)$, respectively. When identical to each other, each mobile has to wait for n time slots for another channel access opportunity, which leads to a harmonic reduction (i.e., at the rate of $1/n$) in per-flow throughput.

2) Error-Prone Wireless Channel and ARQ mechanism: We introduce error probability $p_{err}^{(c)} = p_{err}(\forall c)$ in the wireless channel. In this case, the *pmf* ($f_{\gamma_{sys}^{err}}(s)$) and the expected value ($E[\gamma_{sys}^{err}]$) of system throughput in presence of channel error can be calculated from (8).

$$f_{\gamma_{sys}^{err}}(s) = \sum_{c=1}^{C} q(s|c)\cdot f_{\gamma_{sys}}(c),\ E[\gamma_{sys}^{err}] = \sum_{s=1}^{C} s\cdot f_{\gamma_{sys}^{err}}(s),$$
$$q(s|c) = \binom{c}{s}\cdot(1-p_{err})^{s}\cdot(p_{err})^{c-s}. \quad (8)$$

We model opportunistic scheduling with ARQ by an absorbing DTMC $\mathscr{X}^{(t)}$. At time slot t, $\mathscr{X}^{(t)} \in \{tx, rst, sc\}$ (Fig. 1) represents transmitting, connection reset, or successful transmission states. The states tx and sc are divided into sub-states tx_i and sc_j representing the retransmission counter ($i \in \{0,\cdots,K\}$) and the number of successfully received packets ($j \in \{1,\cdots,C\}$). We set all the sub-states $tx_i(\forall i)$ to be transient states and all other states to be absorbing states. In effect, there are $C+1$ absorbing states: rst and sc_s ($s \in \{1,\cdots,C\}$) corresponding to connection reset and successful transmission of s packets, respectively.

The above process always starts with the retransmission counter set to zero in the sub-state tx_0, which implies three possibilities: connection initiation, connection reset, and successful transmission. At each transition (time slot), the mobile is granted and not granted a channel access with probabilities $1/n$ and $1-1/n$, respectively. The process finishes in state sc_s with probability $\overline{q_s} = f_{\gamma_{sys}^{err}}(s)/n$, where the mobile acquires a channel access and s data packets are successfully transmitted. With probability $\overline{q_0}$, the mobile acquires a channel access but no packet is successfully transmitted. In this case, the retransmission counter will be incremented. If the transmission fails when $\mathscr{X}^{(t)} = tx_K$, the process will finish in state rst where the connection is reset. Therefore, we set the initial probability vector to $[1,\mathbf{0}]$[3] and obtain the TPM ($\mathbf{W}$) as given by (9), where Ω_{ij} and ω_{ij} are the (i,j) entries of $\mathbf{\Omega}$ and $\boldsymbol{\omega}$.

$$\mathbf{W} = \begin{pmatrix} \mathbf{\Omega} & \boldsymbol{\omega} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$
$$\Omega_{ij} = \begin{cases} \frac{n-1}{n}, & \forall i = j \\ \overline{q_0}, & j = i+1, i = \{1,\cdots,K\}, \\ 0, & otherwise \end{cases}$$
$$\omega_{ij} = \begin{cases} \overline{q_{j-1}}, & j = \{2,\cdots,C+1\}, \forall i \\ \overline{q_0}, & (i,j) = (K+1,1) \\ 0, & otherwise. \end{cases} \quad (9)$$

Since $\mathbf{\Omega}$ is digaonal dominant, a closed-form solution can also be obtained from THEOREM 2 below. Hereafter, we denote all-zero, all-one, and identity matrices by $\mathbf{0}$, $\mathbf{1}$, and $\mathbf{I}$, respectively.

THEOREM 2 *The probability that the DTMC process representing* **Case I** *with ARQ is absorbed to state sc_s at time slot d ($f_{\mathscr{S},\mathscr{D}}(s,d)$), the absorbing probability to state sc_s ($f_{\mathscr{S}}(s)$), and the expected time to absorption to state sc_s ($E[\mathscr{D};s]$) can be calculated by using (10)-(12).*

$$f_{\mathscr{S},\mathscr{D}}(s,d) = \begin{cases} \binom{d-1}{K}\left(\frac{n-1}{n}\right)^{d-K-1}(\overline{q_0})^{K+1}, s = 0 \\ \overline{q_s}\sum_{j=1}^{K+1}\binom{d-1}{j-1}\left(\frac{n-1}{n}\right)^{d-j}(\overline{q_0})^{j-1}, s \geqslant 1. \end{cases} \quad (10)$$

$$f_{\mathscr{S}}(s) = \begin{cases} f_{\gamma_{sys}^{err}}(0)^{K+1}, & s = 0 \\ f_{\gamma_{sys}^{err}}(s)\left(\frac{1-f_{\gamma_{sys}^{err}}(0)^{K+1}}{1-f_{\gamma_{sys}^{err}}(0)}\right), & s \geqslant 1. \end{cases} \quad (11)$$

$$E[\mathscr{D};s] = \begin{cases} n(K+1)f_{\gamma_{sys}^{err}}(0)^{K+1}, & s = 0 \\ \left(1 - f_{\gamma_{sys}^{err}}(0)^{K+1}(2+K-(K+1)f_{\gamma_{sys}^{err}}(0))\right) \\ \cdot\left(\frac{nf_{\gamma_{sys}^{err}}(s)}{(1-f_{\gamma_{sys}^{err}}(0))^2}\right), & s \geqslant 1. \end{cases} \quad (12)$$

□

Proof: See Appendix II. ■

COROLLARY 2 *For* **Case I** *with ARQ, the connection reset and successful transmission probability (p_{rst} and p_{sc}) can be calculated from (13). Also, the conditional* pmf *and the expected values of connection reset and successful transmission delay can be calculated from (14) and (15), respectively.*

$$p_{rst} = f_{\mathscr{S}}(0) = f_{\gamma_{sys}^{err}}(0)^{K+1}, p_{sc} = 1 - p_{rst}. \quad (13)$$
$$f_{\mathscr{D}_{rst}}(d) = \frac{f_{\mathscr{S},\mathscr{D}}(0,d)}{p_{rst}} = \binom{d-1}{K}\frac{(n-1)^{d-K-1}}{n^d},$$
$$E[\mathscr{D}_{rst}] = \frac{E[\mathscr{D};0]}{p_{rst}} = n(K+1). \quad (14)$$
$$f_{\mathscr{D}_{sc}}(d) = \frac{\sum_{s=1}^{C} f_{\mathscr{S},\mathscr{D}}(s,d)}{p_{sc}},$$
$$E[\mathscr{D}_{sc}] = \frac{\sum_{s=1}^{C} E[\mathscr{D};s]}{p_{sc}}. \quad (15)$$

□

Eqs. (13)-(15) can be obtained simply by applying *Bayes' theorem* to (10)-(12).

[3]The operation $[a,b]$ denotes the concatenation of the matrix (or scalar) a to the left of the matrix (or scalar) b.

Having obtained the results for *Case I*, we now discuss some insightful implications. First, knowing that either rst or sc will occur, we define n_{rst} and n_{sc} as the number of access opportunities until the occurrence of rst and sc, respectively. Again, for rst to occur, the transmission must fail for $K+1$ consecutive access opportunities. In other words,

$$n_{rst} = \frac{E[\mathscr{D}_{rst}]}{E[\mathscr{D}_{acc}]} = K+1. \quad (16)$$

Similarly,

$$\begin{aligned} n_{sc} &= \frac{E[\mathscr{D}_{sc}]}{E[\mathscr{D}_{acc}]} \\ &= \frac{1 - f_{\gamma_{sys}^{err}}(0)^{K+1}(2 + K - (K+1) f_{\gamma_{sys}^{err}}(0))}{(1 - f_{\gamma_{sys}^{err}}(0)^{K+1}) \cdot (1 - f_{\gamma_{sys}^{err}}(0))} \end{aligned} \quad (17)$$

where the expected inter-access delay ($E[\mathscr{D}_{acc}] = n$) can be calculated from (7).

Secondly, from (12)-(17), we can obtain unconditional expected delay ($E[\mathscr{D}_{fin}]$) and number of access opportunities (n_{fin}) until the process finishes (either in rst or sc) by using the *total probability theorem* as follows:

$$E[\mathscr{D}_{fin}] = \sum_{s=0}^{C} E[\mathscr{D}; s] = n \cdot \frac{1 - f_{\gamma_{sys}^{err}}(0)^{K+1}}{1 - f_{\gamma_{sys}^{err}}(0)}. \quad (18)$$

$$n_{fin} = \frac{E[\mathscr{D}_{rst}]}{E[\mathscr{D}_{acc}]}. \quad (19)$$

From (18) we can obtain $E[\mathscr{D}_{acc}]$ by setting K in (18) to zero (i.e., no retransmission).

Thirdly, conditioned on the occurrence of sc, the per-flow throughput, which is the number of packets successfully transmitted by a mobile per unit time, can be calculated as follows:

$$E[\gamma_{flow}^{sc}] = \frac{E[\gamma_{sys}^{err}]}{E[\mathscr{D}_{sc}]}. \quad (20)$$

$E[\gamma_{flow}^{sc}]$ above also represents the average number of packets transmitted by a mobile until the connection is reset. To prove this statement, let a *transmission cycle* be an interval between the occurrence of either rst or sc and the next occurrence of either rst or sc. Let $\mathscr{Y}_i \in \{sc, rst\}$ be the state at the end of transmission cycle i. Then, the probability that the process will be in state rst for the first time at the end of i cycle is $p_{sc}^{i-1}(1-p_{sc})$, and the corresponding expected number of cycles is $E[k_{\mathscr{Y}}] = 1/(1-p_{sc})$. Also, the average number of successfully transmitted packets in each access opportunity is $E[\gamma_{sys}^{err}]$. Therefore, the expected per flow throughput until the connection is reset can be calculated from $(E[\gamma_{sys}^{err}] \cdot E[k_{\mathscr{Y}}])/(E[\mathscr{D}_{sc}] \cdot E[k_{\mathscr{Y}}])$. Now, it is clear that this metric is the same as $E[\gamma_{flow}^{sc}]$ in (20). We can also calculate two more useful metrics. One is the number of successfully transmitted packets before connection reset, which can be calculated from $E[\gamma_{sys}^{err}] \cdot E[k_{\mathscr{Y}}]$. Another is the delay until the connection is reset, which can be calculated from $E[\mathscr{D}_{sc}] \cdot E[k_{\mathscr{Y}}]$.

B. Case II: All-FSMC

In this section, we assume that channel states of all n mobiles follow the FSMC model. Denoted by $\boldsymbol{f}_{\mathscr{C}_i}^{(0)}$ and $\mathbf{P}_i$ are the initial probability vector and TPM representing the channel of mobile i. When the initial state is c_i, $\boldsymbol{f}_{\mathscr{C}_i}^{(0)} = \boldsymbol{e}_{c_i}$, where $\boldsymbol{e}_{c_i}$ is a row vector whose c_i^{th} entry is one and all other entries are zero.

1) Error-Free Wireless Channel: Again, we start the analysis with $p_{err} = 0$ and later extend the results for the case with non-zero packet error probability and with ARQ. In the absence of transmission error, the process consists of two steps: channel variation and scheduling. Since the channel model in this case exhibits time correlation, the events corresponding to each mobile (i.e., granted channel access or not) in two successive time slots are not independent. Therefore, the joint *pmf* $f_{\mathscr{S},\mathscr{D}_{acc}}(s,d)$ cannot be factored as in (6).

We use an n-dimensional DTMC $\mathscr{C}^{(t)} = (\mathscr{C}_1^{(t)}\mathscr{C}_2^{(t)} \cdots \mathscr{C}_n^{(t)})$ to keep track of channel states of all n mobiles, where $\mathscr{C}_i^{(t)} \in \{1, \cdots, C\}$ is the channel state of mobile i at time slot t. Correspondingly, the initial probability vector and TPM of all n mobiles can be calculated from $\boldsymbol{f}_{\mathscr{C}}^{(0)} = \boldsymbol{f}_{\mathscr{C}_1}^{(0)} \otimes \cdots \otimes \boldsymbol{f}_{\mathscr{C}_n}^{(0)}$ and $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$, respectively, where $\otimes$ denotes the Kronecker product.

Due to the Kronecker product, the resulting states in the DTMC are arranged such that $\mathscr{C}_i^{(t)}$ increases in the reverse order of i. For example, with $n = 3$ and $C = 2$, $\mathscr{C}^{(t)} \in \{(111), (112), (121), (122), (211), \cdots, (222)\}$. Hereafter, we will refer to each entry in the matrices involving the FSMC process, by using an n-digit label $\mathbf{c} = c_1 c_2 \cdots c_n$, where the i^{th} digit is the channel state of mobile i.

Having obtained the model for channel variation, we now incorporate the scheduling mechanism into the model. Due to time-correlation, we model the transmission process by using an absorbing DTMC $(\mathscr{X}^{(t)}, \mathscr{C}^{(t)})$, where $\mathscr{X}^{(t)} \in \{wait, tx\}$ and $\mathscr{C}^{(t)}$ represent the transmission state of the mobile 1 and channel variations of all mobiles, respectively. In the following analysis, we drop $\mathscr{C}^{(t)}$ inherited in each $\mathscr{X}^{(t)}$ for the sake of explanation. At time slot t, $\mathscr{X}^{(t)} = wait$ and $\mathscr{X}^{(t)} = tx$ imply that the mobile is waiting for and is granted a channel access opportunity, respectively.

Let $\mathbf{G}$ and $\mathbf{G}'$ be diagonal matrices with the same dimension as $\mathbf{P}$. Conditioned on the channel states $\mathbf{c} = (c_1 \cdots c_n)$, the diagonal entries ($g(\boldsymbol{c})$ and $g'(\boldsymbol{c})$) of $\mathbf{G}$ and $\mathbf{G}'$ represent the probabilities that mobile 1 will be and will not be granted channel access opportunity, respectively, and can be calculated from

$$g(\boldsymbol{c}) = \begin{cases} 0, & \exists i > 1 : c_i > c_1 \\ \frac{1}{mult_{\mathbf{c}}(c_1)}, & c_i \leqslant c_1 \forall i \end{cases} \quad (21)$$

$$g'(\boldsymbol{c}) = 1 - g(\boldsymbol{c}) \quad (22)$$

where the multiplicity of λ in $\mathbf{c}$ ($mult_{\mathbf{c}}(\lambda)$) is the number of digits in $\mathbf{c}$ which are equal to λ. Given $\boldsymbol{f}_{\mathscr{C}}^{(t-1)}$, the probability that mobile 1 will be and will not be scheduled for transmission at time slot t, is given by $\boldsymbol{f}_{\mathscr{C}}^{(t-1)}\mathbf{PG}$ and $\boldsymbol{f}_{\mathscr{C}}^{(t-1)}\mathbf{PG}'$, respectively.

We assume that mobile 1 acquires a channel access at time slot 0, reset $\mathscr{X}^{(0)} = wait$, and set every state with

$\mathscr{X}^{(t)} = tx$ as the absorbing state. Therefore, the TPM $\mathbf{W}$ for the $(\mathscr{X}^{(t)}, \mathscr{C}^{(t)})$ process can be expressed by (23) below

$$\mathbf{W} = \begin{pmatrix} \mathbf{\Omega} & \boldsymbol{\omega} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{PG}' & \mathbf{PG} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \quad (23)$$

Since $(\mathscr{X}^{(t)}, \mathscr{C}^{(t)})$ always starts from $\mathscr{X}^{(0)} = wait$, the initial probability vector is $\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0]$.

THEOREM 3 *For* **Case II** *without ARQ, if mobile 1 is granted a channel access opportunity at time slot 0,*

- *the probability that mobile 1 will be granted the next opportunity at time slot d and perceive channel state c is*

$$\boldsymbol{f}_{\mathscr{C},\mathscr{D}_{acc}}(c, d|\boldsymbol{f}_{\mathscr{C}}^{(0)}) = \sum_{\forall c_1 = c} \left(\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0]\right) \mathbf{\Omega}^{d-1} \boldsymbol{\omega} \quad (24)$$

- *the probability that mobile 1 will perceive channel state c in the next access opportunity is*

$$\boldsymbol{f}_{\mathscr{C}}(c|\boldsymbol{f}_{\mathscr{C}}^{(0)}) = \sum_{\forall c_1 = c} \left(\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0]\right) (\mathbf{I} - \mathbf{\Omega})^{-1} \boldsymbol{\omega} \quad (25)$$

- *the expected inter-access delay conditioned on $\boldsymbol{f}_{\mathscr{C}}^{(0)}$ is*

$$E[\mathscr{D}_{acc}|\boldsymbol{f}_{\mathscr{C}}^{(0)}] = \left(\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0]\right) (\mathbf{I} - \mathbf{\Omega})^{-1} \mathbf{1}. \quad (26) \quad \square$$

Proof: Since the time to absorption of this DTMC is equivalent to the number of time slots mobile 1 has to wait for its next channel access opportunity, (24)-(26) can be obtained from standard formulae of an absorbing DTMC [21]. Since the absorbing state space for the above DTMC consists of several channel states, the possibilities attributed to all the states are incorporated through the summation. ∎

COROLLARY 3 *For* **Case II**,

$$E[\gamma_{sys}] = C - \sum_{c=1}^{C-1} (F_c)^n,$$

$$E[\mathscr{D}_{acc}|\boldsymbol{f}_{\mathscr{C}}^{(0)}] = \boldsymbol{f}_{\mathscr{C}}^{(0)} (\mathbf{I} - \mathbf{\Omega})^{-1} \mathbf{1}, E[\gamma_{flow}] = \frac{E[\gamma_{sys}]}{E[\mathscr{D}_{acc}]}. \quad (27)$$

Since the throughput is defined at the steady state, the result is the same as in (7). The result for $E[D_{acc}|\boldsymbol{f}_{\mathscr{C}}^{(0)}]$ follows directly from THEOREM 3.

2) Error-Prone Wireless Channel and ARQ Mechanism: We use an approach similar to that in Section III-A.2 to model *Case II* with ARQ. Again, when channel state is c, s packets will be successfully transmitted with probability $q(s|c)$ in (8). We define two matrices ($\mathbf{Q}$ and $\mathbf{Q}'$) which will be used to model *Case II* with ARQ. Matrix $\mathbf{Q}$ maps channel states during an access opportunity to the number of successfully transmitted packets. The entry in row c and column s of $\mathbf{Q}$ are submatrices $q(s|c) \cdot \mathbf{1}$, where $\mathbf{1}$ is an all-one column vector with size C^{n-1}. Matrix $\mathbf{Q}'$ is a diagonal matrix whose c^{th} entry is $q(0|c) \cdot \mathbf{I}$, where $\mathbf{I}$ is an identity matrix with size C^{n-1}.

By allowing only K consecutive transmission opportunities without any successful packet delivery, the TPM ($\mathbf{W}$) for *Case II* with ARQ is formulated as (28) below

$$\mathbf{W} = \begin{pmatrix} \mathbf{\Omega} & \boldsymbol{\omega} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$
$$\mathbf{\Omega}_{ij} = \begin{cases} \mathbf{PG}' & , i = j \\ \mathbf{PGQ}' & , j = i + 1, i = \{1, \cdots, K\}, \\ \mathbf{0} & , otherwise \end{cases}$$
$$\boldsymbol{\omega}_i = \begin{cases} [\mathbf{0}, \mathbf{PGQ}] & , i = \{1, \cdots, K\} \\ \mathbf{PG}[\mathbf{q}_0, \mathbf{Q}] & , i = K + 1 \end{cases} \quad (28)$$

where $\mathbf{\Omega}_{ij}$ and $\boldsymbol{\omega}_i$ are the sub-matrices in row i and column j of $\mathbf{\Omega}$ and $\boldsymbol{\omega}$, and $\mathbf{q}_0 = \mathbf{Q}' \cdot \mathbf{1}$.

Compared to that in Section III-B.1, the model in this section divides a channel access state (tx) based on the transmission result. The probability that no packet is transmitted successfully and that the ARQ increments the retransmission counter is $\mathbf{PGQ}'$. If the transmission fails when the counter is K, the process will finish with probability $\mathbf{PGq}_0$. On the other hand, the process will finish with s successfully transmitted packets with probability located in column s of $\mathbf{PGQ}$.

The above process always starts with retransmission counter set to zero and finishes when at least one packet is successfully transmitted or when the counter exceeds K. Therefore, we set the initial probability vector to $[\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0], \mathbf{0}]$. The row vector $\mathbf{0}$ appended to $\boldsymbol{f}_{\mathscr{C}}^{(0)} \otimes [1, 0]$ represents that the process must start with the retransmission counter set to zero. Connection reset and inter-success delay correspond to the time that the process requires to be absorbed to a state with zero and at least one packet is successfully transmitted. We observe that $\mathbf{W}$ has the same form as in (23). Therefore, the results in this case are the same as in Theorem 3. By replacing the above results into COROLLARY 2, we can obtain the performance results in terms of $\mathscr{D}_{sc}$ and $\mathscr{D}_{rst}$.

C. Case III: FSMC-RSC

In this section, we assume that channel states for n_f and n_r mobiles follow the FSMC and RSC models, respectively[4]. Again, the channel states of mobile i are characterized by steady state probability vector $\boldsymbol{\pi}_i$ and by TPM $\mathbf{P}_i$ as well as initial probability vector $\boldsymbol{f}_{\mathscr{C}_i}^{(0)}$ in case of RSC and FSMC, respectively. Since an RSC model is equivalent to an FSMC when $\boldsymbol{f}_{\mathscr{C}_i}^{(0)} = \boldsymbol{\pi}_i$ (Section II-B), the performance results in this case can be obtained by using the model in Section III-B and setting $\boldsymbol{f}_{\mathscr{C}_i}^{(0)}$ of mobiles with RSC to $\boldsymbol{\pi}_i$.

In general, the worst case complexity is $O(k^3)$, where k is the size of $\mathbf{\Omega}$. With the above solution, the size of $\mathbf{\Omega}$ is C^n and $(K+1) \cdot C^n$ for the scenarios with and without ARQ, respectively. To reduce the complexity, we reduce the size of $\mathbf{\Omega}$ to C^{n_f} and $(K+1) \cdot C^{n_f}$ in each corresponding case. We also obtain the model for special case with $n_f = 1$ and $n_r = n - 1$ where the size of $\mathbf{\Omega}$ becomes C and $(K+1)C$, respectively.

For *Case III*, we only need to keep track of channel states of the mobiles with FSMC. Therefore, $\mathscr{C}^{(t)}$ in Section III-B reduces to $(\mathscr{C}_1^{(t)} \cdots \mathscr{C}_{n_f}^{(t)})$. Correspondingly, $\boldsymbol{f}_{\mathscr{C}}^{(0)}$ and $\mathbf{P}$ become $\boldsymbol{f}_{\mathscr{C}_1}^{(0)} \otimes \cdots \otimes \boldsymbol{f}_{\mathscr{C}_{n_f}}^{(0)}$ and $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_{n_f}$,

[4]*Case III* converges to *Case I* when $n_f = 0$ and $n_r = n$ and to *Case II* when $n_f = n$ and $n_r = 0$.

respectively. Entries $g(\mathbf{c})$ and $g'(\mathbf{c})$ in $\mathbf{G}$ and $\mathbf{G}'$ are also modified to

$$g(\boldsymbol{c}) = \begin{cases} 0, & \exists i > 1 : c_i > c_1 \\ \sum_{k=0}^{n_r} \frac{\binom{n_r}{k}(\pi_{c_1})^k (F_{c_1-1})^{n_r-k}}{k+mult_{\mathbf{c}}(c_1)}, & c_i \leqslant c_1 \forall i \end{cases} \quad (29)$$

$$g'(\boldsymbol{c}) = 1 - g(\boldsymbol{c}) \quad (30)$$

where F_{c_1-1} is the probability that channel state of a mobile in RSC will be less than c_1. The proof of (29) is given in Appendix I. The performance results after this point can be derived in the same way as in Section III-B.

For the case[5] with $n_f = n-1$ and $n_r = 1$, we only keep track of the channel state of mobile 1. In Appendix I, we prove that $g(c) = \frac{f_{\gamma_{sys}}(c)}{n\pi_c}$. We set $\mathbf{P}$ to $\mathbf{P}_1$, and set the (c, s) and (c, c) entries of $\mathbf{Q}$ and $\mathbf{Q}'$ to $q(s|c)$ and $q(0|c)$, respectively. After these basic matrices are obtained, we use the same methodology as for the *All-FSMC* case to obtain the relevant results.

IV. Performance Evaluation

A. Numerical and Simulation Settings

For the different mobiles in a cell, we assume C-state i.i.d. wireless channels with each state being equally likely. The steady state probability vector for each mobile is denoted by $\boldsymbol{\pi}$. We first set $p_{err} = 0$ and vary the number of mobiles (n) and the number of channel states (C) to study $f_{\gamma_{sys}}(c)$, $E[\gamma_{sys}]$, and $E[\gamma_{flow}]$ in *Case I* and *Case II*. In the former case, the expected inter-access delay ($E[\mathscr{D}_{acc}]$) is always equal to the number of mobiles. For the latter case, we show numerical results only for $n_f = 1$ and $n_r = n-1$ (the special case in Section III-C). The results for more general cases can be generated from our framework as well.

For $n = 2$, we study the effect of average channel correlation (ρ) and initial channel state on $\mathscr{D}_{acc}$ for the mobile with FSMC. When the initial state of the mobile is i, the initial probability vector is set to $\boldsymbol{e}_i$ whose i^{th} entry is 1 and all other entries are 0. We then investigate the impact of n and C on $E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$. Again, $f_{\gamma_{sys}}(c)$, $E[\gamma_{sys}]$, and $E[\gamma_{flow}]$ are long-term performance metrics and are the same as those in *Case I*.

Next, we introduce non-zero $p_{err}^{(c)} = p_{err}(\forall c)$, and show the results for expected inter-success delay ($E[\mathscr{D}_{sc}]$) and connection reset probability (p_{rst}) as a function of p_{err} and maximum number of retransmissions (K). The results for p_{sc} are complement to those for p_{rst}. From (14), $E[\mathscr{D}_{rst}]$ always equals to $n(K+1)$. System throughput and per-flow throughput are expected to decrease monotonically with increasing p_{err} and decreasing K. Based on the results of p_{rst}, p_{sc}, $E[\mathscr{D}_{rst}]$, $E[\mathscr{D}_{sc}]$, and $E[\mathscr{D}_{acc}](= n)$, we can verify the formulation of n_{rst}, n_{sc}, n_{fin}, and $E[\mathscr{D}_{fin}]$ in (16)-(19). These results as well as those in *Case II* can be obtained from our framework easily and are omitted for brevity.

We also compare the performance results for an opportunistic scheduler (shown with legend 'OPP') to those for a round-robin scheduler (shown with legend 'RR'), where all the mobiles are scheduled in sequence regardless of their channel states. Due to the deterministic nature of the round-robin scheduling, $f_{\gamma_{sys}}(c) = \pi_c$, $E[\gamma_{sys}] = \sum_{\forall c} c \cdot \pi_c$, $E[\mathscr{D}_{acc}] = E[\mathscr{D}_{acc}|\boldsymbol{e}_i] = n$, $f_{\gamma_{sys}^{err}}(s) = \sum_{\forall c} \pi_c \cdot q(s|c)$, and the results for p_{rst}, p_{sc}, $E[\mathscr{D}_{rst}]$, and $E[\mathscr{D}_{sc}]$ can be obtained from Corollary 2.

[5]This particular case could be useful to estimate the statistics at the mobile, which is aware only of its own state. The next best assumption is to use $\boldsymbol{\pi}$ as the initial probabilities or to use RSC model for all other $(n-1)$ mobiles.

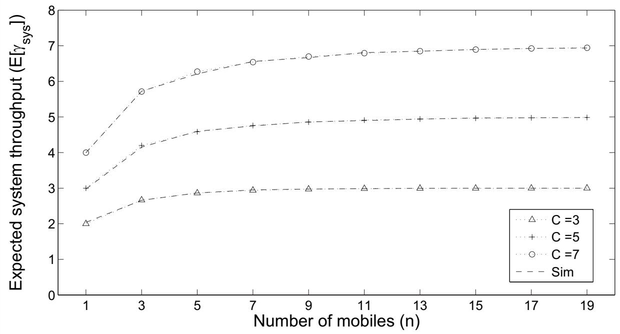

Fig. 2. Expected system throughput in **Case I** without ARQ.

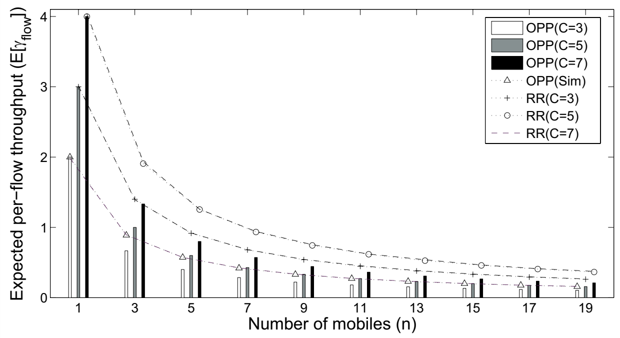

Fig. 3. Expected per-flow throughput for each mobile in **Case I** without ARQ.

B. Simulation Methodology

We validate our analytical results by means of simulations. We collect data samples over 10^4 time slots and find their averages for each performance metric. A sample for $E[\gamma_{sys}]$ is the transmission rate for the selected mobile in each time slot. A sample for $E[\mathscr{D}_{acc}]$ is the interval between two consecutive slots where the same mobile is selected. These samples are classified based on their initial states and are used to calculate $E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$. With non-zero p_{err}, a sample for $E[\mathscr{D}_{sc}]$ is the number of time slots from a connection reset or initiation (rst) or a transmission success (sc) to the next sc event. Again, the statistics for the number of time slots from an rst or an sc event to the next sc event is identical to those between two consecutive sc events because they both start with the retransmission counter set to zero.

C. AMC without ARQ: Results and Discussions

1) All-RSC: Figs. 2-3 plot the expected values of system throughput ($E[\gamma_{sys}]$) and per-flow throughput ($E[\gamma_{flow}]$) as functions of the number of channel states (C) and the number of mobiles (n). In both the figures, the simulation results (shown with legend 'Sim') follow the numerical results very

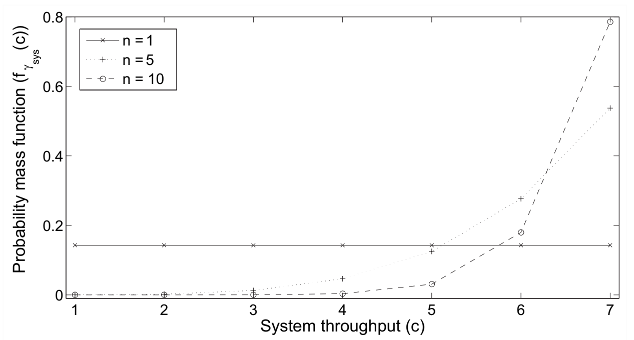

Fig. 4. Probability mass function of system throughput in **Case I** without ARQ for $C = 7$ (continuous plots are used for better readability).

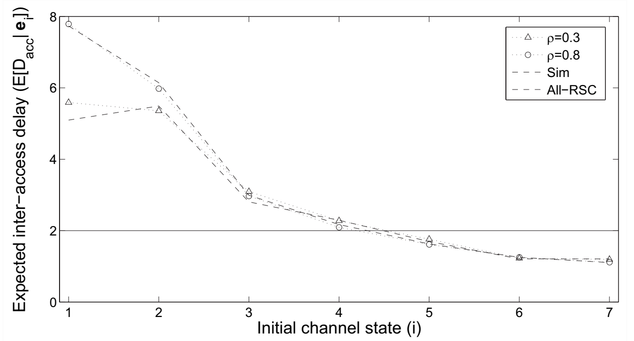

Fig. 5. Impact of initial channel state on expected inter-access delay.

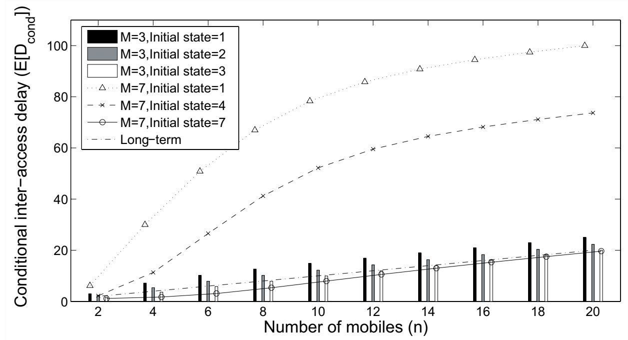

Fig. 6. Typical variations in expected inter-access delay with the number of mobiles for **Case II** without ARQ.

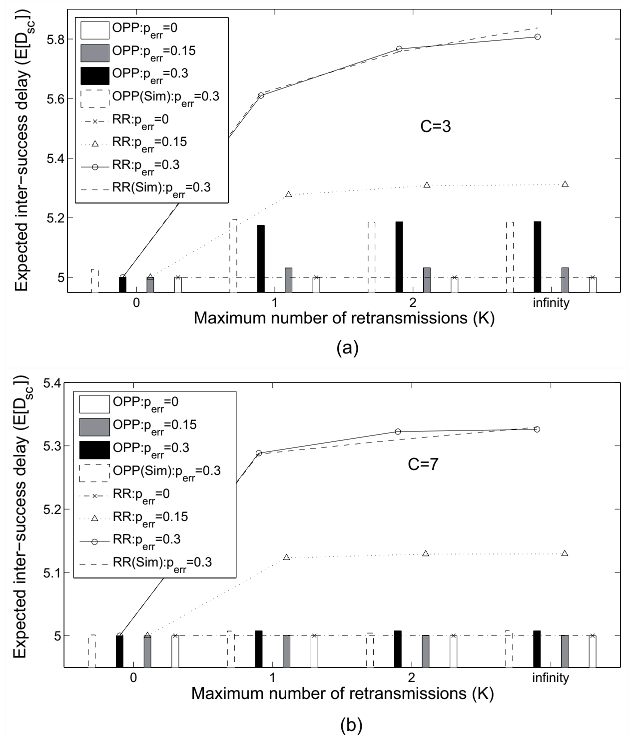

Fig. 7. Comparison of inter-success delay for opportunistic and round-robin scheduling with (a) $C = 3$ and (b) $C = 7$.

closely. As expected, in case of an equally likely wireless channel, increasing C increases $E[\gamma_{sys}]$ and $E[\gamma_{flow}]$.

Increasing the number of mobiles leads to increased $E[\gamma_{sys}]$ due to the multiuser diversity gain. When $n = 1$, the *pmf* of system throughput $f_{\gamma_{sys}}(c) = \pi_c$ (Fig. 4) and $E[\gamma_{flow}] = E[\gamma_{sys}]_{OPP} = E[\gamma_{sys}]_{RR}$ (Fig. 3), where $E[\gamma_{sys}]_{RR} = \sum_{\forall c} c \cdot \pi_c$ is the average system throughput under round-robin scheduling, which is 2, 3, and 4 for $C = 3$, 5, and 7, respectively. For $n > 1$, $f_{\gamma_{sys}}(c)$ is shifted from $\boldsymbol{\pi}$ towards the best state and $E[\gamma_{sys}]_{OPP} > E[\gamma_{sys}]_{RR}$. Despite increasing diversity gain, admitting more mobiles into the system always results in reduction in per-flow throughput as can be observed in Fig. 3 (the proof of this statement is also given in Appendix III).

2) FSMC-RSC: Fig. 5 plots the expected inter-access delay ($E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$) as a function of initial channel state (i) and average channel correlation (ρ) for $n_r = 1$ and $n_f = 1$. As a comparison, we also draw the delay ($= 2$ time slots) for the *All-RSC* case with $n = 2$. Again, this line acts as long-term inter-access delay for *Case II*.

From Fig. 5, we observe that the inter-access delay for the mobile with FSMC depends strongly on the initial channel state. When the initial state is good/bad, the mobile will experience shorter/longer inter-access delay ($E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$). Note that, the initial state i in $E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$ refers to the state (i) when the mobile last acquired the channel access. At steady state, the state i is distributed according to $f_{\gamma_{sys}}(i)$. From these statistics, we can calculate long-term (or steady state) inter-access delay by using the *total probability theorem*; $E[\mathscr{D}_{acc}] = \sum_{i=1}^{C} f_{\gamma_{sys}}(i) \cdot E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$ (which is equal to 2 time slots in Fig. 5). From Fig. 4, we observe that the probability that a mobile with bad (initial) state (e.g., 1) acquires a channel access ($f_{\gamma_{sys}}(i)$) is rather small. In such a case, it is more likely that the mobile will have to wait for a long period of time before it is granted another channel access (Fig. 5).

The delay variations due to initial state is augmented with increasing ρ, since the mobile is more likely to stay in the same state. We can observe in Fig. 5 that the range of the delay with $\rho = 0.8$ is larger than that with $\rho = 0.3$.

In Fig. 6, we set $\rho = 0.5$ and $n_r = 1$, and plot $E[\mathscr{D}_{acc}|\boldsymbol{e}_i]$ as a function of n. We also plot long-term delay ($= n$ time slots) for comparison. Intuitively, inter-access delay increases as n increases. We can observe stronger dependency of inter-access delay on initial states for larger number of mobiles. In this case, the mobile with FSMC starting from a bad initial state might not acquire the channel access after becoming eligible because of an increased number of total eligible mobiles. If not selected when being eligible, the mobile might experience

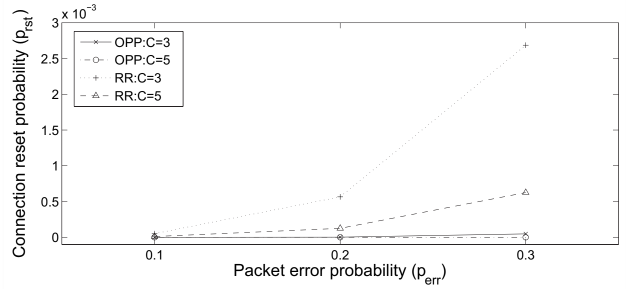

Fig. 8. Comparison of connection reset probability for opportunistic and round-robin scheduling.

bad channel states later in time. Again, it needs some time to regain the eligible status. As a result of these two effects, inter-access delay may increase significantly for bad initial states. This result is aggravated when C increases (e.g., $C = 7$), since the mobile might need longer time to claim the eligible status.

To summarize, larger values of both n and C lead to large delay variation, which can be interpreted as a measure of unfairness[6] among mobiles with different initial states. Here, we observe that an opportunistic scheduler offers the highest throughput but the lowest delay-fairness. On the other hand, a round-robin scheduler achieves the best delay-fairness at the expense of throughput.

D. AMC with ARQ: Results and Discussions

This section presents the results when ARQ is incorporated into the scheduler. In Fig. 7, we set $n = 5$, $p_{err} = \{0, 0.15, 0.3\}$, and maximum number of retransmissions $K = \{0, 1, 2, \infty\}$[7], and plot the expected inter-success delay ($E[\mathscr{D}_{sc}]$) for *Case I*. Fig. 7(a) and Fig. 7(b) represent the cases for $C = 3$ and $C = 7$, respectively. In Fig. 8, we fix $K = 2$ and plot connection reset probability p_{rst} for different values of p_{err}. Fig. 9 plots the joint cumulative distribution function (*cdf*) of cumulative delay d ($F_{\mathscr{S},\mathscr{D}}(s, d) = \sum_{i=1}^{d} f_{\mathscr{S},\mathscr{D}}(s, i)$) for $C = 7$ and $n = 10$. We plot the joint *cdf*s for $p_{err} = 0.05$ and $p_{err} = 0.1$ in Fig. 9(a) and Fig. 9(b), respectively. Finally, we show the *cdf* of inter-success delay ($F_{\mathscr{D}_{sc}}(d) = \sum_{i=1}^{d} f_{\mathscr{D}_{sc}}(i)$) with $C = 3$, $p_{err} = 0.15$, and $n = \{5, 10, 15\}$ in Fig. 10.

Consider Fig. 7 and 8 altogether. When $K = 0$, the transmission must be successful at the first channel access opportunity, otherwise the connection will be reset. Therefore, $E[\mathscr{D}_{sc}] = E[\mathscr{D}_{rst}] = E[\mathscr{D}_{acc}] = n$. For $K > 1$ and $p_{err} > 0$, each transmission might be unsuccessful and retransmission might be required. Therefore, both p_{rst} and $E[\mathscr{D}_{sc}]$ increase with increasing p_{err}. On the other hand, $E[\mathscr{D}_{rst}](= n(K + 1))$ is not affected by p_{err}, since it is conditioned on the occurrence of connection reset. Here, we observe the tradeoff between reliability ($p_{sc} = 1 - p_{rst}$) and latency ($\mathscr{D}_{sc}$) in that improving the successful transmission

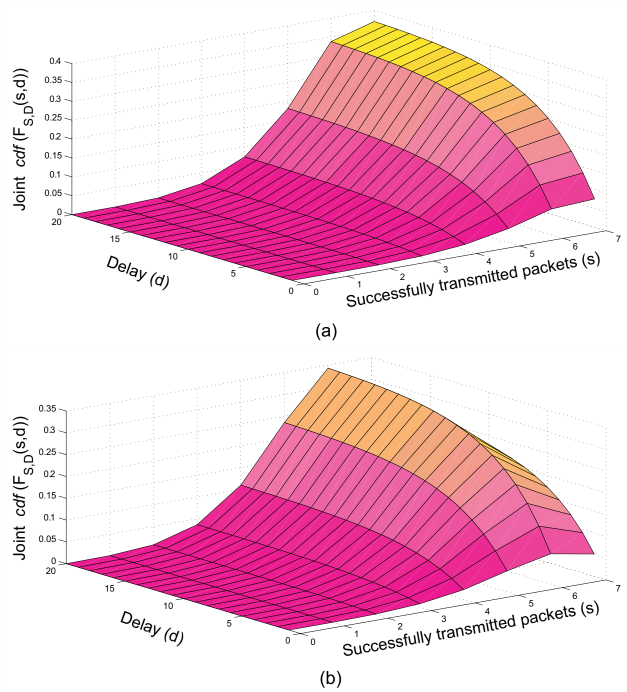

Fig. 9. Joint cumulative distribution function for **Case I** with ARQ when (a) $p_{err} = 0.05$ and (b) $p_{err} = 0.01$.

probability p_{sc} by means of retransmission could lead to an increase in $\mathscr{D}_{sc}$.

Since an opportunistic scheduler selects a mobile with good channel quality, the selected mobile is expected to transmit/receive several packets per time slot. With round-robin scheduling, on the other hand, the mobile tends to transmit/receive fewer number of packets per time slot. In Fig. 8, we observe that with opportunistic scheduling p_{rst} is always less than that with round-robin scheduling. For an equally-likely wireless channel, for both the schemes, increasing C results in an increase in the probability of transmitting more packets, and therefore, decreases p_{rst}. Despite increasing delay, for opportunistic scheduling, $E[\mathscr{D}_{sc}]$ becomes saturated very quickly. The values of $E[\mathscr{D}_{sc}]$ are very close to those for infinite persistence even with only two retransmissions, while the difference in case of round-robin scheduling is still perceptible for $K > 2$.

Fig. 9 presents the joint probability that s packets will be successfully transmitted within a certain threshold d. We observe that there exists a most likely point for the number of successfully transmitted packets (s) corresponding to each value of cumulative delay. This point is a decreasing function in p_{err} (as can be verified in (8)), but remains unchanged for increasing cumulative delay.

By integrating all states $s > 0$ and normalizing the result, we obtain $F_{\mathscr{D}_{sc}}(d)$, which is the probability that at least one packet is successfully transmitted within the limit d. The plot of $F_{\mathscr{D}_{sc}}(d)$ in Fig. 10 reveals that delay variation can be very wide, and the variation tends to increase with increasing n. Even with the best case ($n = 5$), after $E[\mathscr{D}_{sc}]$, the probability that a packet will be successfully transmitted is not more than 73.78%. This implies that the expected value would not be a true indicator of the instantaneous inter-success delay.

[6]Unfairness can be roughly estimated from the difference between the minimum and the maximum value of $E[\mathscr{D}_{acc}|\mathbf{e}_i]$.

[7]For $K = \infty$, we do not increase the retransmission counter and force the DTMC process to stay in state tx_0 for each failure.

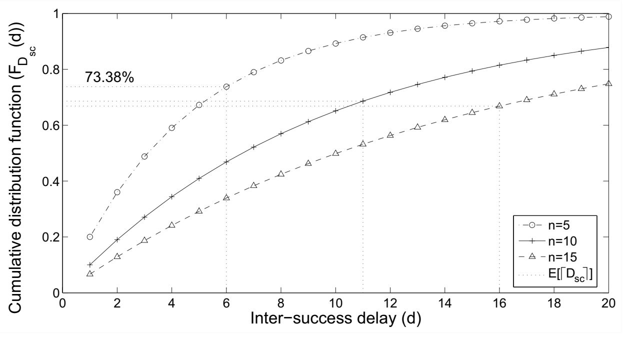

Fig. 10. Cumulative distribution function for inter-success delay in **Case I** with ARQ.

The above joint *pmf* and *cdf* would be useful in three different ways. First, we can utilize this statistics to obtain more accurate information about end-to-end round-trip time (RTT) in a wide-area wireless network, which in turn can be used to minimize the number of timeouts in the end-to-end flow control protocol (e.g., TCP timeouts). Secondly, we can also estimate the *pmf* of link layer bandwidth ($f_{\mathscr{R}}(r)$) from $\sum_{\forall s} \boldsymbol{f}_{\mathscr{S},\mathscr{D}}(s, s/r)$, and use this parameter to probabilistically set the transport layer transmission window size at the sender. Thirdly, for real-time data services, in which each data packet would be useless after some time τ, $F_{\mathscr{D}_{sc}}(\tau)$ is the probability to deliver packets within the delay limit, and can be used as a measure for *quality of service* (QoS).

V. Conclusions

We have presented an analytical model for radio link level channel-quality-based opportunistic scheduling with AMC and ARQ considering both uncorrelated and correlated wireless channels. We have derived complete statistics (in terms of probability mass functions) for the performance measures including system throughput, per-flow throughput, inter-access delay, connection reset delay, and inter-success delay. Performance results for an opportunistic scheduler have been compared to those of a round-robin scheduler under different AMC and channel parameters. Analytical results have been validated through simulations.

Although the multi-user diversity gain is an increasing function in number of mobiles and number of channel states, admitting more mobiles into the system always reduces per-flow throughput and increases delay variation (and hence delay-unfairness). Inter-access delay for a mobile with an FSMC strongly depends on initial channel states. The delay variation (among different initial states) is an increasing function of channel state correlation and diversity gain. The *pmf* and/or *cdf* for the delay reveal the tradeoff between the probability of successful transmission (i.e, reliability) and the corresponding delay for a packet. Since the proposed analytical framework derives the complete statistics for radio link level throughput and delay, it would be useful in modeling and optimizing higher layer protocol performance.

Appendix I
Proof of Propositions 1 and (29)

For *Case I*,

$$\begin{aligned} p_{tx}^{(t)}(c) &= Pr\{(\text{channel of mobile } i \text{ is in state } c) \text{ and} \\ &\qquad (\text{mobile } i \text{ is selected})\} \\ &= \pi_c \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{(\pi_c)^k (F_{c-1})^{n-k-1}}{k+1} \\ &= \frac{1}{n} \sum_{m=1}^{n} \binom{n}{m} (\pi_c)^m (F_{c-1})^{n-m} \qquad (31) \end{aligned}$$

which proves Proposition 1.

For *Case III*, when mobile 1 acquires a channel access, the number of eligible moibles with FSMC and RSC would be $mult_{\mathbf{c}}(c_1)$ and k, respectively. Similar to (31), the probability that k out of n_r mobiles will be eligible is $\binom{n_r}{k}(\pi_{c_1})^k (F_{c_1-1})^{n_r-k}$. By summing all possible values of k each with weight $1/(mult_{\mathbf{c}}(c_1)+k)$, (29) is proven.

By setting $n_r = n-1$,

$$\begin{aligned} g(c) &= \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{(\pi_c)^k (F_{c-1})^{n-k-1}}{k+1} \\ &= \frac{p_{tx}^{(t)}(c)}{\pi_c} = \frac{f_{\gamma_{sys}}(c)}{n\pi_c}. \qquad (32) \end{aligned}$$

When the channels for different mobiles are not i.i.d., $p_{tx}^{(t)}(c)$ becomes,

$$p_{tx}^{(t)}(c) = f_{\mathcal{C}_i}^{(t)}(c) \sum_{k=0}^{n-1} \frac{1}{k+1} \sum_{l=1}^{\binom{n-1}{k}} \left(\prod_{\forall j \in \mathbf{e}_l(n-1,k)} f_{\mathcal{C}_j}^{(t)}(c) \right) \cdot \left(\prod_{\forall j \notin \mathbf{e}_l(n-1,k)} F_{\mathcal{C}_j}^{(t)}(c) \right) \qquad (33)$$

where $\mathbf{e}_l(n,k), l \in \{1, 2, \cdots, \binom{n}{k}\}$ consists of $\binom{n}{k}$ sets of k eligible mobiles when the total number of mobiles is n.

Appendix II
Proof of Theorem 2

In this section, we use subscript ij to represent the (i,j) entry of each matrix.

Lemma 1 *Given that Ω_{ij} is a for $i=j$, b for $j=i+1$, and 0 elsewhere, entries (i,j) of $\mathbf{\Omega}^t$ and $\mathbf{\Gamma} = (\mathbf{I}-\mathbf{\Omega})^{-1}$ denoted by $\Omega_{ij}^{(t)}$ and Γ_{ij}, respectively, can be calculated from (34) below*

$$\begin{aligned} \Omega_{ij}^{(t)} &= \begin{cases} \binom{t}{j-i} a^{k-(j-i)} b^{j-i}, & t \geqslant j-i \\ 0, & otherwise. \end{cases}, \\ \Gamma_{ij} &= \begin{cases} \frac{1}{1-a}, & i=j \\ \frac{1}{1-a}\left(\frac{b}{1-a}\right)^{(j-i)}, & i<j \\ 0, & otherwise. \end{cases} \qquad (34) \end{aligned}$$

□

Proof: We use the Chapman-Kolmogorov equation in (35) below and prove Lemma 1 by induction.

$$\Omega_{ij}^{(t)} = \begin{cases} a \cdot \Omega_{ij}^{(t-1)} + b \cdot \Omega_{i+1,j}^{(t-1)}, & i<j \\ a \cdot \Omega_{ij}^{(t-1)}, & i=j \\ 0, & \text{otherwise.} \end{cases} \qquad (35)$$

For brevity, we skip the proof that Lemma 1 is true for $t = 1$ and prove that Lemma 1 is true for $t+1$, assuming that (34) is true for t. From (35),

$$\Omega_{ij}^{(t+1)} = \begin{cases} \binom{t+1}{j-i} a^{(k+1)-(j-i)} b^{j-i}, & i < j \\ a^{k+1}, & i = j. \end{cases} \tag{36}$$

We observe that (36) is equivalent to that obtained from (34), which proves the first part of Lemma 1. Similarly, we can also prove the second part with relationship $\mathbf{\Gamma}(\mathbf{I} - \mathbf{\Omega}) = \mathbf{I}$. The proof includes the cases for $\Gamma_{i,j+1}$, $\Gamma_{i+1,j}$, and $\Gamma_{i+1,j+1}$ for all values of i and j, which is omitted for brevity. ■

Consider an absorbing DTMC with initial probability matrix $\boldsymbol{\alpha}$ and TPM $\mathbf{W}$ given in (9). The joint probability that the process will be absorbed to state s at time d, absorbing probability to state s, and the expected time to absorption to state s can be calculated from $(\boldsymbol{\alpha}\mathbf{\Omega}^{(d-1)}\boldsymbol{\omega})_{1,s}$, $(\boldsymbol{\alpha}(\mathbf{I}-\mathbf{\Omega})^{-1}\boldsymbol{\omega})_{1,s}$, and $(\boldsymbol{\alpha}(\mathbf{I}-\mathbf{\Omega})^{-2}\boldsymbol{\omega})_{1,s}$, respectively [21]. For the DTMC in THEOREM 2, $a = \frac{n-1}{n}$ and $b = \overline{q_0}$. Therefore, all the entries of $\mathbf{\Omega}^{(d-1)}$, $(\mathbf{I}-\mathbf{\Omega})^{-1}$, and $(\mathbf{I}-\mathbf{\Omega})^{-2}$ can be obtained by replacing a and b in Lemma 1 with $\frac{n-1}{n}$ and $\overline{q_0}$, respectively.

APPENDIX III

By differentiating (7),

$$\frac{\partial}{\partial n} E[\gamma_{flow}] = \sum_{c=1}^{C-1} \left[(F_c)^{n-1} \cdot \left(\frac{F_c}{n^2} - 1 \right) \right] - \frac{C}{n^2}.$$

Since $\frac{F_c}{n^2} - 1 \leqslant 0$ with the equality when $n = 1$, $\frac{\partial}{\partial n} E[\gamma_{flow}] \leqslant 0$, and the per-flow throughput is a decreasing function of n.

ACKNOWLEDGEMENT

This work was supported in part by a scholarship from the TR*Labs*, Winnipeg, Canada, in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and in part by University of Manitoba Graduate Fellowship (UMGF). We sincerely appreciate anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *Proc. Int. Symp. Information Theory 1997*.

[2] R. Knopp and P. A. Humbelt, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC95*, 331-335.

[3] "White paper: Technical overview of 1xEV-DV version G1.4," Motorola Inc., 2002.

[4] X. Liu, E. K. P. Chong, and N. B. Shroff, "Transmission scheduling for efficient wireless utilization," in *Proc. IEEE INFOCOM01*, pp. 776-785.

[5] P. Bender, P. Black, *et al.*, "CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, no. 7, pp. 70-77, July 2000.

[6] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA/HDR: A high efficiency-high data rate personal communication wireless system," in *Proc. IEEE VTC00 Spring*.

[7] S. Shakkottai and A. L. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: the exponential rule," *Analytic Methods in Applied Probability*, vol. 207, no. 2, pp. 185-202, 2002.

[8] M. Andrews, K. Kumaran, *et al.*, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150-154, Feb. 2001.

[9] T. Issariyakul and E. Hossain, "ORCA-MRT: an optimization-based approach for fair scheduling in multi-rate TDMA wireless networks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2823–2835, Nov. 2005.

[10] Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proc. IEEE INFOCOM03*.

[11] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Opportunistic power scheduling for multi-server wireless systems with minimum performance constraints," in *Proc. IEEE INFOCOM04*.

[12] M. Andrews and L. Zhang, "Scheduling over non-stationary wireless channels with finite rate sets," in *Proc. IEEE INFOCOM04*.

[13] Q. Liu, S. Zhou, and G. B. Giannakis, "Queuing with adaptive modulation and coding over wireless links: Cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[14] , "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746-1755, Sept. 2004.

[15] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688-1692, Nov. 1999.

[16] Y. L. Guan and L. F. Turner, "Generalized FSMC model for radio channel with correlated fading," *IEE Proc. Commun.*, vol. 146, pp. 133-137, Apr. 1999.

[17] M. Hassan, M. M. Krunz, and I. Matta, "Markov-based channel characterization for tractable performance analysis in wireless packet networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 821-831, May 2004.

[18] M. Zorzi, "Data-link packet dropping models for wireless local communications," *IEEE Trans. Veh. Technol.*, vol. 51, no. 4, pp. 710-719, July 2002.

[19] M. S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Wireless Personal Commun.*, vol. 13, no. 1-2, pp. 119-143, May 2000.

[20] A. Doufexi, S. Armour, *et al.*, "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 172-180, May 2002.

[21] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. The John Hopkins University Press, 1981.

Teerawat Issariyakul (S'02) is currently a Post-Doctoral Fellow in the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, Canada. He received his bachelor in Electrical Engineering from Thammasat University, Thailand, in 1997. In 1999, he received M.Eng in telecommunications from Asian Institute of Technology (AIT), Thailand, where he had been awarded with Telephone Organization of Thailand Scholarship. He earned his Ph.D. from the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, Canada, in 2005. He was a researcher at TR*Labs*, Winnipeg, Canada. He held a scholarship from TRLabs as well as a University of Manitoba Graduate Fellowship (UMGF). Dr. Issariyakul's main research interests are in the area of modeling, analysis, and optimization of protocols for wireless networks.

Ekram Hossain (S'98-M'01-SM'06) is an Associate Professor (with tenure) in the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, Canada. He received his Ph.D. in Electrical Engineering from the University of Victoria, Canada, in 2000. He was a University of Victoria Fellow and also a recipient of the British Columbia Advanced Systems Institute (ASI) graduate student award. He received his B.Sc. and M.Sc., both in Computer Science and Engineering, from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1995 and 1997, respectively. Dr. Hossain's research interests include radio link control and transport layer protocol design and cross-layer optimization issues for the next-generation wireless networks. Currently he serves as an Editor for the *IEEE Transactions on Wireless Communications*, the *Wireless Communications and Mobile Computing* Journal (Wiley InterScience), the *IEEE/KICS Journal of Communications and Networks*, and the *International Journal of Sensor Networks (IJSNet)*.