Analyzing and Optimizing Adaptive Modulation Coding Jointly With ARQ for QoS-Guaranteed Traffic

Xin Wang, *Member, IEEE*, Qingwen Liu, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—A cross-layer design is developed for quality-of-service (QoS)-guaranteed traffic. The novel design jointly exploits the error-correcting capability of the truncated automatic repeat request (ARQ) protocol at the data link layer and the adaptation ability of the adaptive modulation and coding (AMC) scheme at the physical layer to optimize system performance for QoS-guaranteed traffic. The queuing behavior induced by both the truncated ARQ protocol and the AMC scheme is analyzed with an embedded Markov chain. Analytical expressions for performance metrics such as packet loss rate, throughput, and average packet delay are derived. Using these expressions, a constrained optimization problem is solved numerically to maximize the overall system throughput under the specified QoS constraints.**

***Index Terms*—Adaptive modulation and coding (AMC), automatic repeat request (ARQ), cross-layer design, embedded Markov chain, quality-of-service (QoS).**

I. Introduction

RECENTLY, there has been much interest in cross-layer designs, where one allows the physical layer to interact and share information with higher layers (e.g., the data link with the network layer) to achieve significant performance gains; see, e.g., [13]–[15] and references therein. Especially, many recent works focus on cross-layer combining adaptive modulation and coding (AMC) at the physical layer with automatic repeat request (ARQ) protocol at the data link layer [9]–[12], [16]. The main objective behind these designs is to improve the spectral efficiency by jointly incorporating the adaptation ability of the AMC and the error-correcting capability of ARQ.

Among the various cross-layer designs, Liu *et al.* [11] put forth an interesting scheme combining AMC with the truncated ARQ protocol in order to enhance throughput while fulfilling packet loss and delay constraints for delay-sensitive traffic. Having lowered the physical layer packet error rate (PER) requirement by the error-correcting capability of ARQ, modulation and coding modes with higher transmission rate can be chosen at the physical layer. As a result, by optimizing across the AMC and truncated ARQ modules, it becomes possible to improve the overall spectral efficiency. However, queuing delay of the packets was not considered in [11]. The same authors further proposed cross-layer combining of queuing with AMC and derived analytical expressions for the packet loss rate and throughput using a finite-state Markov chain analysis [12]. Based on the latter, they provided a cross-layer design to minimize the packet loss rate and maximize the average throughput, but the possible performance improvement from the ARQ protocol at the data link layer was not taken into account.

In [16], a joint design accounting for the code distribution, truncated ARQ protocol, and an average SNR-based AMC scheme was proposed for the multicode code-division multiple-access (CDMA) uplink setup. Based on analytical expressions derived for the packet loss rate, throughput, average packet delay, and the quality-of-service (QoS) constraints, a cross-layer design was also formulated. Although different users in [16] can choose different Transmission modes (TMs) based on their average SNRs, each user is only allowed to select a fixed TM throughout in its transmission. As a result, the adaptation capability of the AMC scheme is not fully exploited in [16] since the TM of a single user is not adapted according to the "instantaneous" variation of its received SNR.

This paper fills the gap between [12] and [16] by designing jointly the truncated ARQ protocol and the AMC scheme based on instantaneous rather than average SNR. The queuing process induced by both the truncated ARQ protocol and the AMC scheme is analyzed using an embedded Markov chain (Section III). Guided by the queuing analysis, we then design jointly the truncated ARQ protocol and the AMC scheme to ensure QoS-guaranteed traffic (Section IV). Although our focus is on a point-to-point link, the proposed cross-layer design can be applied to multiple links and can also be coupled with a multicode CDMA uplink cross-layer design as in [16]. The main contributions of this paper are the following: 1) generalization of the cross-layer combining of queuing with AMC in [12] and the cross-layer combining of queuing with truncated ARQ and average SNR-based AMC in [16]; 2) judicious construction of an embedded Markov chain to capture the joint queuing process induced by the truncated ARQ protocol and the AMC scheme and derivation of analytical expressions for packet loss rate,

Manuscript received January 6, 2006; revised March 30, 2006 and April 20, 2006. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The review of this paper was coordinated by Prof. D. O. Wu.

X. Wang was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with the Department of Electrical Engineering, Florida Atlantic University, Boca Raton, FL 33431 USA (e-mail: xin.wang@ece.umn.edu).

Q. Liu was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with Navini Networks, Richardson, TX 75082 USA (e-mail: Qingwen.liu@gmail.com).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (email: georgios@ece.umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2007.891465

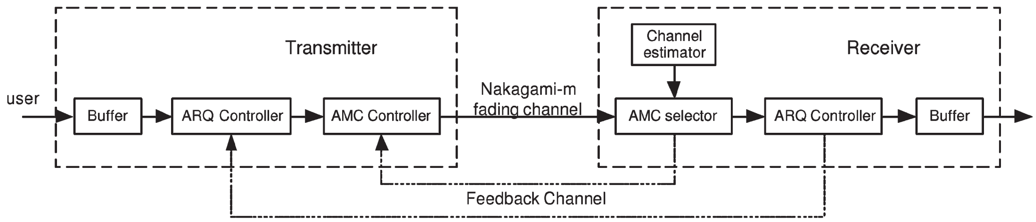


Fig. 1. Single link between a single-antenna transmitter (user) and a single-antenna receiver (BS).

throughput, and average packet delay; and 3) formulation of a cross-layer design as a constrained optimization problem over a finite set to fully exploit the error-correcting capability of the ARQ protocol and the adaptation ability of the AMC scheme. Section V provides some numerical results, followed by the conclusions of this paper.

II. Modeling

As shown in Fig. 1, we consider a point-to-point wireless packet communication link between a single-antenna transmitter (user) and a single-antenna receiver (base station, BS). This link is to support QoS-guaranteed traffic, which is characterized by a maximum average packet delay δ and a maximum packet loss rate ρ. At both transmitter and receiver, ARQ controllers are used to regulate the operation of the truncated ARQ protocol at the data link layer. Following the ARQ controller at the transmitter end, the packets go through an AMC controller, which updates the modulation and code pair (i.e., the TM) according to the feedback received from the AMC selector at the receiver end. The latter selects the TM based on the estimated received SNR. The processing unit at the data link layer is a packet consisting of information bits, while the processing unit at the physical layer is a frame consisting of transmitted symbols. Similar to [12] and [16], we suppose that each user's packets are generated according to a Markovian process, i.e., the packet arrival process is memoryless. Each packet contains N_p bits. At both transmitter and receiver ends, there is a buffer (queue) that operates in a first-in–first-out (FIFO) mode and can store as many as B packets.

Our system model adheres to the following assumptions.

A1) Time is slotted as in [11], [12], and [16], and one frame is transmitted per slot. Each frame at the physical layer contains at most one packet from the data link layer. This assumption facilitates the queuing process since it ensures that the transmit buffer operates in FIFO mode under the ARQ protocol. The data link layer and physical layer overhead consumes negligible bandwidth, and the propagation delay is also negligible, as in [16].

A2) A Nakagami-m block-frequency flat-fading model [7], [8] is adopted for the propagation channel, according to which the channel remains time invariant during the *coherence time interval (CTI)* of T_f seconds, but is allowed to vary across successive CTIs of T_f seconds. This model also describes frequency-selective fading channels when transmitters rely on orthogonal frequency-division multiplexing [20].

TABLE I
TRANSMISSION MODES WITH CONVOLUTIONALLY CODED MODULATION

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Modulation	BPSK	QPSK	QPSK	16-QAM	64-QAM
Code rate R_c	1/2	1/2	3/4	3/4	3/4
R_n (bits/sym.)	0.50	1.00	1.50	3.00	4.50
a_n	274.723	90.251	67.618	53.399	35.351
g_n	7.993	3.500	1.688	0.376	0.090
γ_{pn} (dB)	-1.533	1.094	3.972	10.249	15.978

A3) Perfect channel state information (CSI) is available at the receiver using training-based channel estimation, and the resultant TM is fed back from the ARQ selector at the receiver without error and latency, as in [5], [6], and [12]. The assumption that the feedback channel is error free and has no latency could be at least approximately satisfied by using a fast feedback link with powerful error control coding. Further considerations on system design with, e.g., delayed or noisy CSI, will be left for future investigation. A "pure" ARQ protocol is employed to coordinate retransmissions of the erroneous packets. Generalization to hybrid ARQ techniques [18], [19] is possible but is left for future research.

A4) Error detection is perfect at the receiver provided that sufficiently reliable error detection cyclic redundancy check codes are used [2], [11], [12], [16]. Packets with detected errors are dropped after N_r retransmissions [11], [16]. When the buffer of a transmitter is full, subsequently arriving packets are also dropped, as in [11], [12], and [16].

The wireless link supports different bit rates via AMC with a fixed number of TMs. Convolutionally coded M_n-ary rectangular or square quadratic amplitude modulation, adopted from IEEE 802.11a standard [4], is used in the AMC pool. All possible TMs are listed in Table I in a rate ascending order. In Table I, a_n, g_n, and γ_{pn} are the fitting parameters for TMs with packet length $N_p = 1080$ bits, which we will use later on to calculate the PER. As per A1, we let each frame transmitted over one slot to contain one packet. The packet and frame structures are depicted in Fig. 2.

A. Channel Modeling and AMC

In this paper, we adopt the channel model and structure of the AMC scheme in [12], which we review briefly in this section for completeness and motivational purposes.

As described in [12], the quality of a flat-fading channel can be simply captured by the received SNR γ. For the

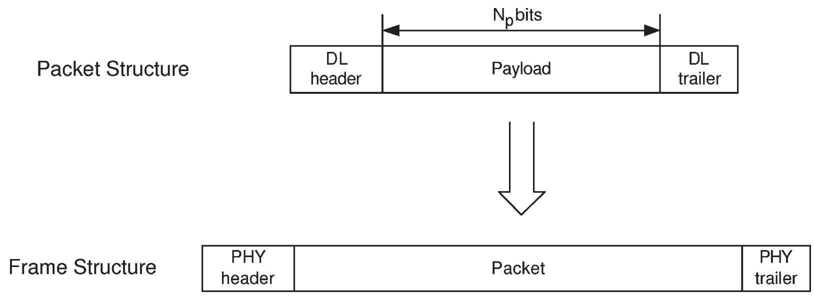


Fig. 2. Packet and frame structures.

block-fading channel model in A2, γ is described by the general Nakagami-m model that prescribes a Gamma probability density function (pdf) [3]

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

where $\bar{\gamma} := E\{\gamma\}$ is the average received SNR, $\Gamma(m) := \int_0^\infty t^{m-1} \exp(-t)dt$ is the Gamma function, and m is the Nakagami fading parameter ($m \geq 1/2$). Given γ in this channel model, the objective of AMC is to select a suitable TM to maximize the data rate while maintaining a prescribed PER P_0. Let N denote the total number of available TMs. In addition to the N modes, the user can also choose a TM 0, i.e., to stay silent and to avoid deep channel fading. For these $N+1$ choices available of the AMC selector, the entire SNR range is partitioned into $N+1$ nonoverlapping consecutive intervals with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$ such that TM n is chosen when $\gamma \in [\gamma_n, \gamma_{n+1})$. In the presence of additive white Gaussian noise, the PER can be approximated as [12]

$$\text{PER}_n(\gamma) \approx \begin{cases} 1, & 0 < \gamma < \gamma_{pn} \\ a_n \exp(-g_n \gamma), & \gamma \geq \gamma_{pn} \end{cases} \tag{2}$$

where n is the mode index, and a_n, g_n, and γ_{pn} are mode-dependent parameters, which are listed in Table I for packet length $N_p = 1080$ bits. These parameters are obtained by fitting (2) to the exact PER, as explained in [11, App.]. Given the mode selection scheme and the pdf of γ in (1), the probability of TM n being chosen is given by

$$\Pr(n) = \int_{\gamma_n}^{\gamma_{n+1}} p_\gamma(\gamma) d\gamma = \frac{\Gamma\left(m, \frac{m\gamma_n}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{m\gamma_{n+1}}{\bar{\gamma}}\right)}{\Gamma(m)} \tag{3}$$

where $\Gamma(m, x) := \int_x^\infty t^{m-1} \exp(-t)dt$ is the complementary Gamma function. If, in practice, we have $\gamma \geq \gamma_n$, then the average PER corresponding to mode n, known as $\overline{\text{PER}}_n$, is given by

$$\begin{aligned} \overline{\text{PER}}_n &= \frac{1}{\Pr(n)} \int_{\gamma_n}^{\gamma_{n+1}} a_n \exp(-g_n \gamma) p_\gamma(\gamma) d\gamma \\ &= \frac{a_n m^m \left(\Gamma(m, b_n \gamma_n) - \Gamma(m, b_n \gamma_{n+1})\right)}{\Pr(n) \Gamma(m) \bar{\gamma}^m b_n^m} \end{aligned} \tag{4}$$

where $b_n := m/\bar{\gamma} + g_n$, $n \in [1, N]$.

The algorithm searching for the thresholds $\{\gamma_n\}_{n=0}^{N+1}$ to achieve the prescribed P_0 per mode operates as follows [12]: 1) Set $n = N$, and $\gamma_{N+1} = +\infty$. 2) For each n, search for the unique $\gamma_n \in [\gamma_{pn}, \gamma_{n+1}]$ that satisfies $\overline{\text{PER}}_n = P_0$, or if there is no such γ_n, pick $\gamma_n = \gamma_{pn}$. 3) If $n > 1$, set $n = n - 1$, and go to Step 2; otherwise, set $\gamma_0 = 0$ and stop. Given a prescribed P_0, this algorithm guarantees that $\overline{\text{PER}}_n \leq P_0$ for all $n \in [1, N]$.

For a given P_0, let C_n denote the channel state corresponding to the SNR region $[\gamma_n, \gamma_{n+1})$, $n \in [0, N]$, in which TM n is chosen. By the slow-fading condition of the block-fading channel model, transition happens only between adjacent states at the edge of two CTIs. For this channel model, the following has been established [13].

Lemma 1: The channel can be modeled as a Markov chain with $(N+1) \times (N+1)$ state transition matrix given by

$$\mathbf{P}_C = \begin{bmatrix} P_{0,0} & P_{0,1} & 0 & \cdots & 0 \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & P_{N-1,N-2} & P_{N-1,N-1} & P_{N-1,N} \\ 0 & \cdots & 0 & P_{N,N-1} & P_{N,N} \end{bmatrix}. \tag{5}$$

The associated state transition probability is given by

$$P_{n,l} = 0, \quad |l - n| \geq 2 \tag{6}$$

$$P_{n,n+1} = \frac{N_{n+1} T_f}{\Pr(n)}, \quad P_{n,n-1} = \frac{N_n T_f}{\Pr(n)} \tag{7}$$

$$P_{n,n} = \begin{cases} 1 - P_{n,n+1} - P_{n,n-1}, & 0 < n < N \\ 1 - P_{n_0,n_1}, & n = 0 \\ 1 - P_{N,N-1}, & n = N \end{cases} \tag{8}$$

where N_n is the cross rate of TM n. With f_d denoting the maximum Doppler shift, N_n can be estimated as [17, eq. (17)]

$$N_n = \frac{\sqrt{2\pi} f_d}{\Gamma(m)} \left(\frac{m\gamma_n}{\bar{\gamma}}\right)^{m-0.5} \exp\left(-\frac{m\gamma_n}{\bar{\gamma}}\right). \tag{9}$$

B. Slot Configuration

Depending on the channel state, different TMs are chosen. Since a packet contains a fixed number of bits, its duration varies for different TMs. As per A1, the wireless link is slotted, and each slot contains one frame, whereas each frame contains at most one packet. As a result, each slot's duration varies, depending on the underlying channel state.

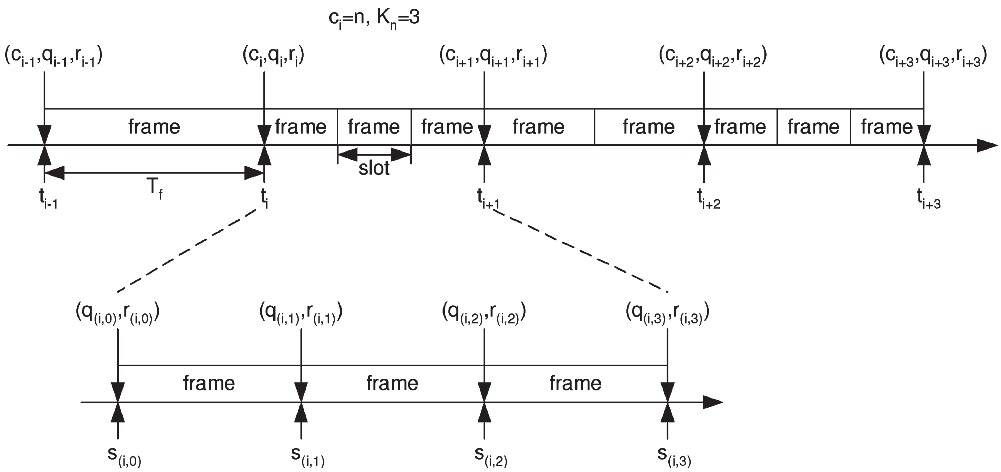


Fig. 3. State and substate transition.

Given channel state C_n (i.e., the TM n is chosen by the AMC controller), let L_n denote the frame duration in seconds, which is equal to the slot duration since the propagation delay is negligible by A1. For $n = 1, \ldots, N$, we have

$$L_n = \frac{N_p + N_{\text{ohd}}}{R_n R_s} \approx \frac{N_p}{R_n R_s} \tag{10}$$

where R_n denotes the number of bits carried per symbol for TM n (refer to Table I for the specific value), R_s is the symbol rate in symbols per second, N_p stands for the number of information bits carried by a packet, and N_{ohd} denotes the number of bits for the physical and data link layer overhead in a frame. The approximation in (11) is due to the fact that $N_p \gg N_{\text{ohd}}$ by A1. When in channel state C_0, we let the slot duration be identical to that under C_1, i.e., $L_0 = L_1$. From the truncated ARQ protocol viewpoint, we assume that $\overline{\text{PER}}_0 = 1$ under channel state C_0 since the packet will surely fail if it is transmitted.

Referring to Table I, the number of bits R_n carried per symbol for TM $n \in [2, N]$ is an integer multiple of R_1. Therefore, by (10), the frame duration L_1 is R_n/R_1 times larger than L_n for $n \in [2, N]$, given the unchanged symbol rate R_s. We assume henceforth that $L_1 = T_f$, where T_f is the duration of a CTI defined by the Nakagami-m block-fading model in A2. This implies that we should choose a packet length such that the frame duration is less than the channel's coherence time. Under the block-fading model, this in turn implies that over a CTI (with the same fading state), only one slot is present under channel states C_0 and C_1, but as many as R_n/R_1 slots of length L_n are present under channel state C_n for $n \in [2, N]$. As a result, given the channel state C_n, the number K_n of slots per T_f block is given by

$$K_n = \begin{cases} 1, & n = 0, 1 \\ R_n/R_1, & n = 2, \ldots, N \end{cases}. \tag{11}$$

Note that (11) relies on the fact that the overhead is negligible, as per A1, but even when the overhead is nonnegligible, we can also adjust the overhead of different TMs to reach (11).

III. Queuing Analysis of AMC With ARQ

Given the system configuration outlined in the previous section, our proposed cross-layer approach is to jointly design the truncated ARQ protocol at the data link layer and the AMC scheme at the physical layer. To this end, we need to analyze the queuing process induced by the truncated ARQ protocol and the AMC scheme. In the ensuing queuing analysis, we assume for simplicity that the packet generation adheres to a Poisson process with intensity λ (in packets per second),[1] but our analysis can also be readily applied to any other Markov packet arrival process. Each packet contains N_p bits. The buffer at the transmitter can store as many as B packets. Here, we let $B < \infty$, which amounts to a finite buffering system. At the physical layer, the symbol rate of the wireless link is fixed at R_s (in symbols per second). There are N available TMs and thus $N + 1$ choices available to the AMC selector. In the AMC operation, a prescribed PER P_0 must be ensured for the packet transmissions. The thresholds $\{\gamma_n\}_{n=0}^{N+1}$ required by the AMC selector to achieve the prescribed P_0 are determined through the algorithm given in Section II-A. At the data link layer, the truncated ARQ protocol uses a retry limit N_r. Given those system parameters, especially P_0 and N_r, the next subsection describes our novel queuing analysis. Note that our framework provides the analysis for a queuing process induced by *both* the ARQ protocol and the AMC scheme, which is significantly different from [12], where the ARQ protocol was not present to simplify the analysis for queuing with AMC.

A. Embedded Markov Chain

As in Fig. 3, we divide the time axis into CTIs and let t_i denote the starting point of the ith such interval. Let (c_i, q_i, r_i) denote the channel, queue, and ARQ protocol state indices at t_i, where c_i, q_i, and r_i are integers with $c_i \in [0, N]$, $q_i \in [0, B]$, and $r_i \in [0, N_r]$. The state triplet (c_i, q_i, r_i) indicates that when

[1]In practice, the arrival packet rate λ is usually measured or estimated based on previous experience or desired system capacity. This Poisson arrival assumption is widely used in performance analysis, e.g., in [12]–[15]. Although this assumption may not always be realistic, the analysis based on it provides an initial basis for design.

the channel state lies in C_{c_i}, there are q_i packets left in the buffer, and r_i transmission tries have been completed for the first packet in the buffer. Note that when $q_i = 0$, we have $r_i = 0$ since there is no packet in the queue at all, let alone the trying history of the first packet. If we just look at the set of t_i time points instead of the time axis, the transitions of (c_i, q_i, r_i) is Markovian. Therefore, we can use an embedded Markov chain to describe the underlying queuing process.

Let us first investigate the state transition between slots. As shown in Fig. 3, given the channel state index $c_i = n$, there will be K_n slots in the ith CTI. Let $s_{(i,j)}$, $j \in [1, K_n - 1]$ denote the ending instant of the jth slot in the ith CTI, and define $s_{(i,0)} \equiv t_i$ and $s_{(i,K_n)} \equiv t_{i+1}$. Moreover, noticing that the channel state c_i is unchanged during the whole CTI, we let $(q_{(i,j)}, r_{(i,j)})$ denote the queue and truncated ARQ states at $s_{(i,j)}$ and define the stationary distribution vector $\boldsymbol{\chi}^{(j,n)}$ of the substates $(q_{(i,j)}, r_{(i,j)})$ at $s_{(i,j)}$, $j = 0, \ldots, K_n$, as

$$\boldsymbol{\chi}^{(j,n)} := \left[\chi^{(j,n)}_{(0,0)}, \chi^{(j,n)}_{(1,0)}, \ldots, \chi^{(j,n)}_{(1,N_r)}, \ldots, \chi^{(j,n)}_{(B,0)}, \ldots, \chi^{(j,n)}_{(B,N_r)}\right] \tag{12}$$

where $\chi^{(j,n)}_{(q,r)}$ denotes the stationary probability of the queue and the truncated ARQ state indices (q, r) at $s_{(i,j)}$ under channel state C_n. Then, we can establish the following result.

Lemma 2: When the system is stable, we have

$$\boldsymbol{\chi}^{(K_n,n)} = \boldsymbol{\chi}^{(0,n)} \mathbf{T}_n^{K_n} \tag{13}$$

where the state transition probability matrix $\mathbf{T}_n$ is defined as

$$\mathbf{T}_n := \begin{bmatrix} T^{(n)}_{(0,0),(0,0)} & T^{(n)}_{(0,0),(1,0)} & \cdots & T^{(n)}_{(0,0),(B,N_r)} \\ T^{(n)}_{(1,0),(0,0)} & T^{(n)}_{(1,0),(1,0)} & \cdots & T^{(n)}_{(1,0),(B,N_r)} \\ \vdots & \vdots & \ddots & \vdots \\ T^{(n)}_{(B,N_r),(0,0)} & T^{(n)}_{(B,N_r),(1,0)} & \cdots & T^{(n)}_{(B,N_r),(B,N_r)} \end{bmatrix} \tag{14}$$

where $T^{(n)}_{(x,y),(v,w)}$ denotes the transition probability from substate (x, y) at $s_{(i,j-1)}$ to substate (v, w) at $s_{(i,j)}$ under channel state C_n.

Proof: By the fact that the substate transition from $s_{(i,j-1)}$ to $s_{(i,j)}$ is Markovian for all $j \in [1, K_n]$ and that the state transition probability matrix $\mathbf{T}_n$ is the same for all the slots of a CTI when the system is stable, we have

$$\begin{aligned} \boldsymbol{\chi}^{(j,n)} &= \boldsymbol{\chi}^{(j-1,n)} \mathbf{T}_n, \quad j \in [1, K_n] \\ \Leftrightarrow \quad \boldsymbol{\chi}^{(K_n,n)} &= \boldsymbol{\chi}^{(0,n)} \mathbf{T}_n^{K_n}. \end{aligned} \tag{15}$$

■

Given Poisson-distributed packet arrivals per user, the substate transition between two slots can be derived, and the nonzero entries of $\mathbf{T}_n$ are determined by the following rules:

1) If $x = 0$, then

$$T^{(n)}_{(0,0),(v,0)} = \begin{cases} P_{A|(j,n)}(v), & v \in [0, B-1] \\ 1 - \sum_{k=0}^{B-1} P_{A|(j,n)}(k), & v = B \end{cases}. \tag{16}$$

2) If $1 \le x \le B - 1$ and $0 \le y \le N_r - 1$, then

$$T^{(n)}_{(x,y),(v,y+1)} \approx \begin{cases} \overline{\text{PER}}_n P_{A|(j,n)}(v - x), & v \in [1, B-1] \\ \overline{\text{PER}}_n \left[1 - \sum_{k=0}^{B-1-x} P_{A|(j,n)}(k)\right], & v = B \end{cases} \tag{17}$$

$$T^{(n)}_{(x,y),(v,0)} \approx \begin{cases} (1-\overline{\text{PER}}_n) P_{A|(j,n)}(v-x+1), & v \in [x-1, B-2] \\ (1-\overline{\text{PER}}_n)\left[1-\sum_{k=0}^{B-1-x} P_{A|(j,n)}(k)\right], & v = B-1 \end{cases}. \tag{18}$$

3) If $1 \le x \le B - 1$ and $y = N_r$, then

$$T^{(n)}_{(x,N_r),(v,0)} = \begin{cases} P_{A|(j,n)}(v-x+1), & v \in [x-1, B-2] \\ 1-\sum_{k=0}^{B-1-x} P_{A|(j,n)}(k), & v = B-1 \end{cases}. \tag{19}$$

4) If $x = B$ and $0 \le y \le N_r - 1$, then

$$T^{(n)}_{(B,y),(B,y+1)} = \overline{\text{PER}}_n \tag{20}$$

$$T^{(n)}_{(B,y),(B-1,0)} = 1 - \overline{\text{PER}}_n. \tag{21}$$

5) If $x = B$ and $y = N_r$, then

$$T^{(n)}_{(B,N_r),(B-1,0)} = 1. \tag{22}$$

In (16)–(22), $\overline{\text{PER}}_n$ is given by (4) for $n \in [1, N]$ ($\overline{\text{PER}}_0 = 1$), and $P_{A|(j,n)}(a)$ denotes the probability of a packets arriving during the jth slot, which for Poisson arrivals is

$$P_{A|(j,n)}(a) = \frac{(\lambda L_n)^a}{a!} \exp(-\lambda L_n), \quad a \ge 0. \tag{23}$$

We should remark that for any Markov arrival process with distribution $\tilde{P}_{A|(j,n)}(a)$, we can apply a similar analysis with $\tilde{P}_{A|(j,n)}(a)$ playing the role of $P_{A|(j,n)}(a)$.

Considering the overall queuing process, let us now define the stationary probability vector of the states (c_i, q_i, r_i) at t_i as

$$\boldsymbol{\pi} := [\boldsymbol{\pi}_0, \ldots, \boldsymbol{\pi}_N] \tag{24}$$

where

$$\boldsymbol{\pi}_n := \left[\pi_{(n,0,0)}, \pi_{(n,1,0)}, \ldots, \pi_{(n,1,N_r)}, \ldots \right. \\ \left. \pi_{(n,B,0)}, \ldots, \pi_{(n,B,N_r)}\right] \tag{25}$$

with $\pi_{(n,q,r)}$ denoting the stationary probability of the channel, queue, and ARQ protocol states being (n, q, r) at t_i. Using Lemmas 1 and 2, we arrive at the main result of our embedded Markov chain modeling.

Proposition 1: The stationary state distribution vector $\boldsymbol{\pi}$ can be computed from

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}, \quad \sum_{n=0}^{N} \left[\sum_{\pi_{(n,q,r)} \in \boldsymbol{\pi}_n} \pi_{(n,q,r)}\right] = 1 \tag{26}$$

where the overall transition probability matrix can be organized in block form as

$$\mathbf{P} := \begin{bmatrix} \mathbf{P}_{0,0} & \cdots & \mathbf{P}_{0,N} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{N,0} & \cdots & \mathbf{P}_{N,N} \end{bmatrix} \tag{27}$$

and the submatrix $\mathbf{P}_{n,l}$ is defined as

$$\mathbf{P}_{n,l} = P_{n,l}\mathbf{A}_n \tag{28}$$

with $P_{n,l}$ defined in (5), and $\mathbf{A}_n := \mathbf{T}_n^{K_n}$.

Proof: Again, we first suppose that the system is always stable. For analyzing the resultant stable Markov chain of states (c_i, q_i, r_i), we need to study the transition probability from state (c_i, q_i, r_i) to state $(c_{i+1}, q_{i+1}, r_{i+1})$. From Lemma 1, we know that c_{i+1} only depends on c_i. Letting $\Pr(a|b)$ denote the transition probability from state b to state a, we have

$$\Pr\left((c_{i+1}, q_{i+1}, r_{i+1})|(c_i, q_i, r_i)\right) = \Pr\left(c_{i+1}|c_i\right)\Pr\left((q_{i+1}, r_{i+1})|(c_i, q_i, r_i)\right) \tag{29}$$

where $\Pr(c_{i+1}|c_i)$ is available through the entries of matrix $\mathbf{P}_C$ in (5).

Given the channel state index $c_i = n$, we can use Lemma 2 and the fact that $s_{(i,0)} \equiv t_i$ and $s_{(i,K_n)} \equiv t_{i+1}$ to obtain

$$\begin{aligned}\Pr\left((q_{i+1}, r_{i+1})|(n, q_i, r_i)\right) &:= \Pr\left(q_{(i,K_n)}, r_{(i,K_n)}|q_{(i,0)}, r_{(i,0)}\right) \\ &= \sum_{\forall x_j, y_j} T_{\left(q_{(i,0)}^{(n)}, r_{(i,0)}\right),(x_1,y_1)} \left(\prod_{j=1}^{K_n-2} T^{(n)}_{(x_j,y_j),(x_{j+1},y_{j+1})}\right) \\ &\quad \times T^{(n)}_{(x_{K_n-1},y_{K_n-1}),(q_{(i,K_n)},r_{(i,K_n)})}\end{aligned} \tag{30}$$

where $T^{(n)}_{(x,y),(v,w)}$ is defined in (14).

With (29) and (30), we can readily organize the overall transition probability matrix as in (27), and the stationary state distribution vector $\boldsymbol{\pi}$ can then be computed from (26). It is easy to show that the Markov chain characterized by the transition probability matrix $\mathbf{P}$ is irreducible, homogeneous, and positive recurrent, which in turn establishes that a stationary distribution $\boldsymbol{\pi}$ always exists and is unique. This justifies our initial assumption that the stability of this Markov chain is guaranteed, and the proof is complete. ■

Analyzing this embedded Markov chain is the core of our queuing approach. Equation (26) implies that $\boldsymbol{\pi}$ is the left eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 and can be computed by standard techniques. With $\boldsymbol{\pi}$ computed by (26), we are ready to derive the steady-state performance metrics of interest, namely throughput, average packet delay, and packet loss rate for the wireless link under consideration.

B. Packet Loss Rate and Throughput

Let $\bar{S}$ and $\bar{\xi}$ denote the average throughput (in bits per second) and packet loss rate, respectively. Packet loss in our finite buffering system comes from both packet failures at the data link layer after $N_r + 1$ transmission tries and blockage due to buffer overflow. Let $\bar{N}_b$ and $\bar{N}_f$ denote the expected number of blocked packets and failed packets during a CTI of T_f seconds, respectively. Next, we derive $\bar{N}_b$ and $\bar{N}_f$ given the stationary probability vector $\boldsymbol{\pi}$.

To estimate $\bar{N}_b$, let us first look at the expected number of blocked packets $\bar{N}_b(j, n)$, $j \in [1, K_n]$, during the jth slot of the ith CTI under channel state C_n. Using the notation $\boldsymbol{\chi}^{(j,n)}$ defined in (12), we have

$$\bar{N}_b(j, n) = \sum_{(q,r)} \chi^{(j-1,n)}_{(q,r)} \bar{N}_b(j, n, q, r), \quad j \in [1, K_n] \tag{31}$$

where $\bar{N}_b(j, n, q, r)$ denotes the expected number of blocked packets at $s_{(i,j)}$ given the state (n, q, r) at $s_{(i,j-1)}$. Clearly, if a packet arrives when there are already B packets in the buffer, it is blocked and dropped. If there are already $q (0 \le q \le B)$ packets in the transmitter's buffer at $s_{(i,j-1)}$, at most the first $B - q$ incoming packets can be kept during the jth slot, whereas the remaining packets (if in existence) are blocked. Therefore, $\bar{N}_b(j, n, q, r)$ is given by

$$\begin{aligned}\bar{N}_b(j, n, q, r) &= \sum_{a=B-q+1}^{\infty} \left[a - (B - q)\right] P_{A|(j,n)}(a) \\ &= \sum_{a=B-q+1}^{\infty} a P_{A|(j,n)}(a)(B - q) \sum_{a=B-q+1}^{\infty} P_{A|(j,n)}(a) \\ &= \lambda L_n \left[1 - \sum_{a=0}^{B-q-1} P_{A|(j,n)}(a)\right] \\ &\quad - (B - q)\left[1 - \sum_{a=0}^{B-q} P_{A|(j,n)}(a)\right].\end{aligned} \tag{32}$$

Note that we define $\sum_{a=0}^{-1} P_{A|(j,n)}(a) = 0$ when $q = B$, i.e., $\bar{N}_b(j, n, B, r) = \lambda L_n$.

Using $\boldsymbol{\pi}$, we can calculate the stationary substate distribution vector $\boldsymbol{\chi}^{(j,n)}$. First, by $t_i \equiv s_{(i,0)}$, we have

$$\boldsymbol{\chi}^{(0,n)} = \frac{1}{\sum_{\pi_{(n,q,r)} \in \boldsymbol{\pi}_n} \pi_{(n,q,r)}} \boldsymbol{\pi}_n = \frac{1}{\Pr(n)} \boldsymbol{\pi}_n \tag{33}$$

where $\boldsymbol{\pi}_n$ is defined in (24), and $\Pr(n)$ denotes the probability of the channel state being C_n, as given by (3). The second equality follows from the fact that the channel state c_i is independent of the queue and ARQ protocol states q_i and r_i. Starting from $\boldsymbol{\chi}^{(0,n)}$ and using (15), we can calculate $\boldsymbol{\chi}^{(j,n)}$ for $j \in [1, K_n - 1]$. Having calculated $\boldsymbol{\chi}^{(j,n)}$, $j \in [1, K_n - 1]$, and with $\bar{N}_b(j, n, q, r)$ given by (32), we can obtain $\bar{N}_b(j, n)$ for $n \in [0, N]$ and $j \in [1, K_n]$ by (31). Finally, we can obtain the expected number $\bar{N}_b$ of blocked packets during a CTI as

$$\bar{N}_b = \sum_{n=0}^{N} \Pr(n) \left[\sum_{j=1}^{K_n} \bar{N}_b(j, n)\right]. \tag{34}$$

Similarly, we let $\bar{N}_f(j,n)$ be the expected number of failed packets during the jth slot of the ith CTI given channel state C_n. A packet fails and is discarded at the data link layer with probability $\overline{\text{PER}}_n$ at $s_{(i,j)}$ if the ARQ state is N_r at instant $s_{(i,j-1)}$. Hence, we have

$$\bar{N}_f(j,n) = \sum_{q=1}^{B} \chi_{(q,N_r)}^{(j-1,n)} \overline{\text{PER}}_n \tag{35}$$

where $\chi_{(q,N_r)}^{(j,n)}$ is defined in (12) and can be obtained through (33) and (15), and $\overline{\text{PER}}_n$ is given by (4) for $n \in [1, N]$ (and $\overline{\text{PER}}_0 = 1$). Similarly, we have

$$\bar{N}_f = \sum_{n=0}^{N} \Pr(n) \left[\sum_{j=1}^{K_n} \bar{N}_f(j,n) \right]. \tag{36}$$

By the fact that the packet loss is contributed by both $\bar{N}_b$ and $\bar{N}_f$, and that λT_f is the average number of arriving packets during a CTI, we have following result.

Proposition 2: Given the packet arrival rate λ (in packets per second), the packet loss rate $\bar{\xi}$ is evaluated as

$$\bar{\xi} = \frac{\bar{N}_b + \bar{N}_f}{\lambda T_f} \tag{37}$$

where $\bar{N}_b$ and $\bar{N}_f$ are given by (34) and (36), respectively. Then, given the packet loss rate $\bar{\xi}$, we can obtain the throughput $\bar{S}$ as

$$\bar{S} = \lambda N_p (1 - \bar{\xi}) \tag{38}$$

where λN_p is the average arrival information bits per second.

C. Average Packet Delay

The total average delay $\bar{D}$ for a packet in the slotted system can be decomposed into two parts, namely 1) the average service time $\bar{D}_E$ for the enable transmission interval (ETI), which stands for the time duration each packet has to wait from the time it arrives until the beginning of the next slot, and 2) the average delay $\bar{D}_Q$ in the embedded Markov chain of Section III-A.

In the embedded Markov chain analysis, we implicitly assumed that all newly arriving packets during a slot enter the system at the end of the slot since we only look at t_i (and $s_{(i,j)}$) instead of the whole time axis. This compensates for $\bar{D}_E$ in true average packet delay calculation. Actually, if infinite buffering is allowed, the underlying queuing process can be approximated by an M/G/1 queue [1, Ch. 5]. Due to the nature of a slotted system, this M/G/1 queue would take vocation. Then, when calculating the average packet delay, an extra delay should be added for the vocations. In our finite buffering system, the M/G/1 queue approximation is not valid, but an extra delay $\bar{D}_E$ is still present, which plays the role of the extra vocation delay in the M/G/1 queue. Let $\bar{D}_E(n)$ denote the average service time of the ETI for the slots contained in CTI under channel state C_n. Similar to [16], we further assume that $\bar{D}_E(n)$ is given by

$$\bar{D}_E(n) \approx L_n/2 \tag{39}$$

where L_n denotes the slot duration given by (10) for $n \in [1, N]$ and $L_0 = L_1$. Our simulations in the ensuing section confirm that this approximation is reasonable. Because there are K_n slots in each CTI under channel state C_n, we have

$$\bar{D}_E = \frac{\sum_{n=0}^{N} K_n \Pr(n) \bar{D}_E(n)}{\sum_{n=0}^{N} K_n \Pr(n)} = \frac{\sum_{n=0}^{N} K_n \Pr(n) L_n/2}{\sum_{n=0}^{N} K_n \Pr(n)} \tag{40}$$

where $\Pr(n)$ denotes the probability of mode n being chosen, which is given by (3).

By Kleinrock's result [1, Ch. 2], the average delay for our embedded Markov chain is

$$\bar{D}_Q = \frac{\bar{Q}}{\lambda(1 - P_b)} \tag{41}$$

where $\bar{Q}$ denotes the average number of packets in the transmit queue at t_i, P_b denotes the probability of having a packet blocked, and thus, $\lambda(1 - P_b)$ is the effective packet arrival rate. With the stationary distribution $\boldsymbol{\pi}$ computed from (26), we can calculate $\bar{Q}$ as

$$\bar{Q} = \sum_{n=0}^{N} \left\{ \sum_{q=1}^{B} q \left[\sum_{r=0}^{N_r} \pi_{(n,q,r)} \right] \right\}. \tag{42}$$

With the expected number of blocked packets $\bar{N}_b$, which is given by (34), P_b is simply given by

$$P_b = \frac{\bar{N}_b}{\lambda T_f}. \tag{43}$$

Note that in our system, the packet being served is not immediately removed from the transmit buffer since it may need to be retransmitted with the truncated ARQ protocol. A packet is removed from the buffer only at the end of a slot when it is successfully received or discarded when the retry limit is exceeded. Therefore, $\bar{D}_Q$ in (41) indicates the packet delay from the beginning of the slot following its arrival until it is successfully received.

Overall, the final result of our queuing analysis is summarized in the following proposition.

Proposition 3: The average packet delay in our system can be evaluated as

$$\bar{D} = \bar{D}_E + \bar{D}_Q \tag{44}$$

where $\bar{D}_E$ and $\bar{D}_Q$ are given by (40) and (41), respectively.

IV. Cross-Layer Design

With the analytical expressions derived in Section III, we are now ready to optimize system performance using a novel cross-layer design. For the truncated ARQ protocol, the retry limit N_r can be any positive integer. However, only a finite retry limit can be afforded in practice. Thus, $N_r \in \mathbf{\Omega}$, where $\mathbf{\Omega}$ is a finite positive integer set. From Section II-A, we know that the operation of AMC at the physical layer only depends on the prescribed PER P_0, which is a real number in the range $\mathbf{\Phi} = (0, 1)$. Therefore, given the measured or estimated arrival packet rate λ (in packets per second), packet length N_p (in bits), user buffer size B (in packets), symbol rate R_s (in symbols per second), number of available TMs N, and QoS requirements, namely maximum average packet delay δ and maximum packet loss rate ρ, the proposed cross-layer design aims to optimally determine the retry limit N_r at the data link layer and the prescribed PER P_0 at the physical layer. Recall that in Propositions 2 and 3, we derived analytical expressions for the average throughput, packet loss rate, and packet delay that depend on N_r and P_0. Let $\bar{S}(N_r, P_0)$, $\bar{\xi}(N_r, P_0)$, and $\bar{D}(N_r, P_0)$ denote the average throughput, packet loss rate, and packet delay, respectively, given the specific N_r and P_0 parameters. Then, our cross-layer design can be formulated as searching for the optimal N_r and P_0, i.e.,

$$(N_r^{\text{opt}},\ P_0^{\text{opt}}) = \arg \max_{N_r \in \mathbf{\Omega};\, P_0 \in \mathbf{\Phi}} \bar{S}(N_r, P_0) \qquad (45)$$

$$\text{s.t.} \qquad \bar{\xi}(N_r, P_0) \le \rho \qquad (46)$$

$$\bar{D}(N_r, P_0) \le \delta. \qquad (47)$$

Inequalities (46) and (47) represent the packet loss rate and average delay constraints, respectively. Since the expressions for $\bar{S}(N_r, P_0)$, $\bar{\xi}(N_r, P_0)$, and $\bar{D}(N_r, P_0)$ are complicated in general and do not have a closed form, there is not much room for developing efficient algorithms in solving (45). However, because the pair (N_r, P_0) lies in a bounded space $\mathbf{\Omega} \times \mathbf{\Phi}$, we can resort to a 2-D exhaustive search to solve (45) numerically and obtain N_r^{opt} and P_0^{opt}.

V. Numerical Results

In this section, we resort to computer simulations to verify the performance analysis in Section III and provide a numerical example to illustrate the cross-layer design in Section IV.

A. Verification of Performance Analysis

Consider a point-to-point packet communication system with total bandwidth $R_s = 1.08M$ (in symbols per second). From (10), the frame duration is $L_n \approx N_p/R_n R_s$ under channel state C_n with $N_p = 1080$ bits. We assume that the Nakagami fading parameter $m = 1$ for the propagation channel (this corresponds to Rayleigh fading) with coherence interval $T_f = 2$ ms and Doppler frequency $f_d = 10$ Hz, i.e., $f_d T_f = 0.02$. We carried out simulations under three different system parameter settings. The first setting corresponds to the average received SNR $\bar{\gamma} = 15$ dB, buffer size $B = 10$ packets, prescribed PER $P_0 = 0.05$ for AMC, and retry limit $N_r = 3$ for the ARQ, while the same parameters for the second and third settings are $\bar{\gamma} = 10$ dB, $B = 15$ packets, $P_0 = 0.01$, $N_r = 3$, and $\bar{\gamma} = 10$ dB, $B = 10$ packets, $P_0 = 0.02$, and $N_r = 5$, respectively. In the simulations, the transmitter's buffer was fed with a Poisson source having intensity λ (in packets per second). Under each

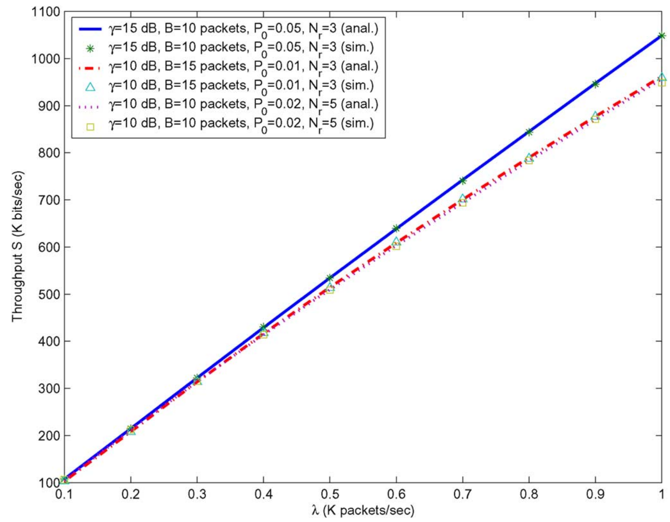


Fig. 4. Comparison between analytical and simulated throughput.

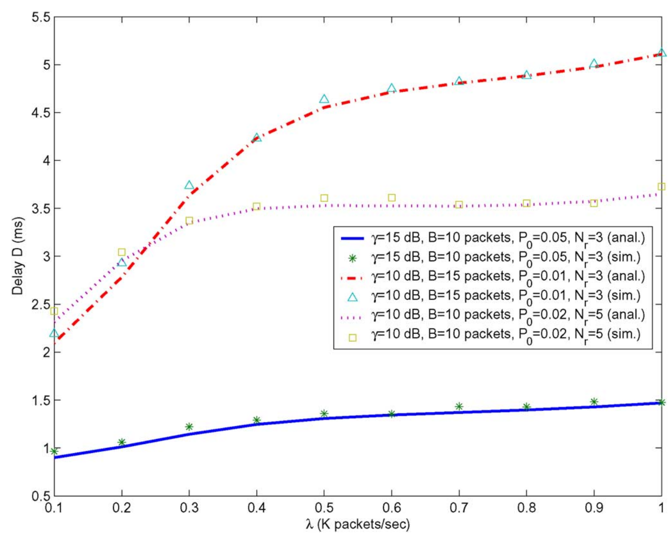


Fig. 5. Comparison between analytical and simulated average delay.

parameter setting, ten cases were carried out, where in each case, the transmitter's buffer was fed with a different λ. For each case, the result was obtained as the average of ten independent runs, where in each run, the system was simulated for a time period equivalent to 100 000 ms. Figs. 4–6 compare analytical with simulation results for throughput, average delay, and packet loss rate, respectively. In the figures, "lines" correspond to analytical expressions, while each point signifies the corresponding simulation-based results. As corroborated by Figs. 4–6, simulations validate the analytical expressions in Section III, which are the basis of our cross-layer design.

B. Cross-Layer Design Examples

To illustrate the proposed cross-layer design, let us consider the same point-to-point system with bandwidth $R_s = 1.08M$ (in symbols per second). Let the transmitter be fed with a Poisson source with intensity $\lambda = 0.1\,K$ (in packets per second) and buffer size $B = 10$ packets, and let the average received

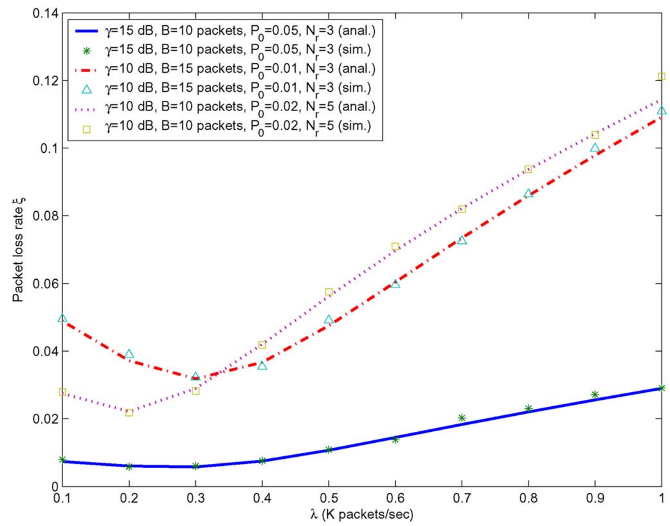


Fig. 6. Comparison between analytical and simulated packet loss rate.

SNR be $\bar{\gamma} = 15$ dB. We suppose that the truncated ARQ protocol can afford a maximum retry limit equal to 10. The QoS-guaranteed traffic is characterized by maximum packet loss rate $\rho = 0.01$ and maximum average packet delay $\delta = 1$ ms. Note that the well-known QoS requirements for voice packet traffic correspond to a maximum packet loss rate of 0.01 and a maximum (not average) packet delay of 200 ms. We retain the same packet loss rate constraint but use a much stricter packet delay constraint since our system can afford a very high symbol rate. We tested three designs for this system. Design 1 is the proposed cross-layer design obtained via (45). Design 2 is the cross-layer combining of queuing with ARQ and average SNR-based AMC for the single link in [16]. Design 3 is the cross-layer combining of queuing with AMC in [12], which is reformulated to a constrained optimization similar to (45) for the QoS-guaranteed traffic in this example.

Fig. 7 depicts the system performance by arbitrarily selecting different retry limits N_r and different prescribed PERs P_0 instead of judiciously selecting them as in the proposed cross-layer design. It turns out that through our joint design, one is capable of optimizing system performance under the specified QoS constraints. The results obtained by the three tested designs are summarized in Table II, where $\bar{S}$, $\bar{\xi}$, and $\bar{D}$ denote the expected throughput, packet loss rate, and average packet delay, respectively, while P_0^{opt}, N_r^{opt}, and n^{opt} denote the optimal prescribed PER, optimal retry limit, and optimal fixed transmission mode, respectively. Note that n^{opt} only exists in Design 2, and N_r^{opt} does not exist in Design 3. Clearly, the proposed cross-layer design (Design 1) satisfies all the QoS requirements and provides the optimal throughput $\bar{S} = 107.95$ K bits/s. Design 2 [16] also provides a good solution satisfying all the QoS requirements, but the achieved throughput ($\bar{S} = 107.08$ K bits/s) is smaller than that achieved by Design 1, which takes full advantage of the adaptation capability of the AMC scheme. Design 3 fails to yield a solution satisfying the QoS requirements. For illustration purposes, a solution "close" to Design 3 (in which the QoS requirements are not exceeded by much) is also listed in Table II. In this solution, both the resultant packet loss rate (0.01001) and the

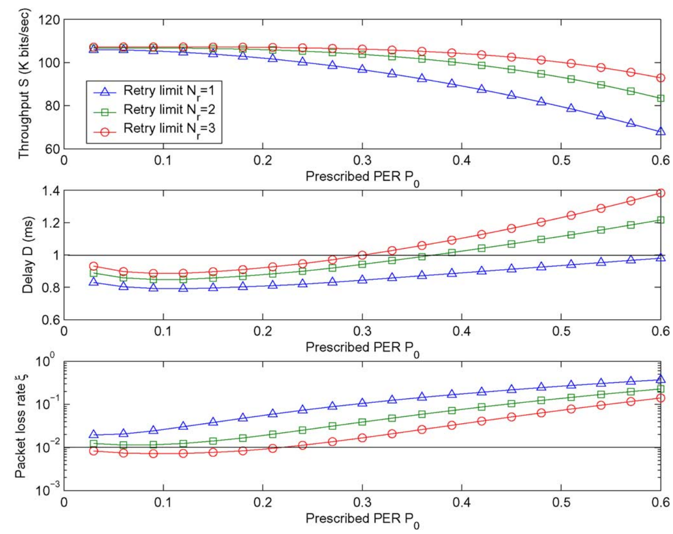


Fig. 7. System performance for different retry limits N_r and different prescribed PERs P_0.

TABLE II
CROSS-LAYER DESIGN EXAMPLE 1: POISSON ARRIVAL (DESIGN 1: PROPOSED CROSS-LAYER DESIGN (45); DESIGN 2: CROSS-LAYER COMBINING OF QUEUING WITH ARQ AND AVERAGE SNR AMC FOR SINGLE LINK AS IN [16]; DESIGN 3: CROSS-LAYER COMBINING OF QUEUING WITH AMC AS IN [12]; QoS REQUIREMENTS: $\rho = 0.01$, $\delta = 1$ ms)

	Design 1	Design 2	Design 3
Throughput $\bar{S}$ (K bits/sec)	107.95	107.08	106.92
Packet loss rate $\bar{\xi}$	0.00049208	0.0085523	0.01001
Average delay $\bar{D}$ (ms)	0.9829	0.97109	5.3507
Prescribed SNR P_0^{opt}	0.05	–	0.01
Retry limit N_r^{opt}	9	10	–
Transmission mode n^{opt}	–	4	–

average packet delay (5.3507 ms) do not satisfy the QoS requirements. The proposed Design 1 outperforms Design 3 [12] for the QoS-guaranteed traffic in this example mainly because it capitalizes on the error-correcting capability of the truncated ARQ protocol at the data link layer. Note that the good delay performance of the proposed cross-layer design also benefits from the employed single-packet-per-frame structure. However, the multiple-packet-per-frame, i.e., packet-packing, structure considered in Design 3 can save the overhead (when it is nonnegligible) and simplify the overall system implementation.

As stated in Section III, our queuing analysis (and thus the proposed cross-layer design) applies to any Markovian packet arrival process. In another example, we assume that the arrival process to the queue is Bernoulli distributed with a given average rate $\lambda = 0.1K$ packets/s and parameter $p \in (0, 1)$. As a result, the instantaneous arriving rate at time t can be expressed as

$$A(t) = \begin{cases} 0, & \text{with probability } p \\ \lambda/(1-p), & \text{with probability } 1-p. \end{cases} \quad (48)$$

With the other system parameters remaining the same as the last example, Table III shows the comparison of the three designs with Bernoulli arrivals. It is clear that similar trends are observed.

TABLE III
CROSS-LAYER DESIGN EXAMPLE 2: BERNOULLI ARRIVAL (DESIGN 1: PROPOSED CROSS-LAYER DESIGN (45); DESIGN 2: CROSS-LAYER COMBINING OF QUEUING WITH ARQ AND AVERAGE SNR AMC FOR SINGLE LINK AS IN [16]; DESIGN 3: CROSS-LAYER COMBINING OF QUEUING WITH AMC AS IN [12]; QoS REQUIREMENTS: $\rho = 0.01$, $\delta = 1$ ms)

	Design 1	Design 2	Design 3
Throughput $\bar{S}$ (K bits/sec)	107.95	107.05	106.92
Packet loss rate $\bar{\xi}$	0.0004931	0.008763	0.010004
Average delay $\bar{D}$ (ms)	0.9802	0.9822	5.3231
Prescribed SNR P_0^{opt}	0.05	–	0.01
Retry limit N_r^{opt}	9	10	–
Transmission mode n^{opt}	–	4	–

VI. CONCLUSION

In this paper, we derived a cross-layer design across the data link and physical layers. The key behind the novel design is to jointly exploit the error-correcting capability of the truncated ARQ protocol and the adaptation ability of the AMC scheme at the physical layer to optimize the system performance for QoS-guaranteed traffic. The queuing process induced by both the truncated ARQ protocol and the AMC scheme was analyzed using an embedded Markov chain. With the derived analytical expressions of pertinent performance metrics, we jointly specified the retry limit for the truncated ARQ protocol as well as the prescribed PER for AMC to optimize the system throughput for QoS-guaranteed traffic. Computer simulations were carried out to verify the performance analysis, and a numerical example was used to illustrate the novel cross-layer design, which outperformed existing alternatives.

ACKNOWLEDGMENT

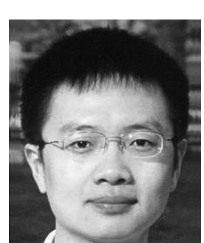
The views and conclusions in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

REFERENCES

[1] L. Kleinrock, *Queuing Systems*, vol. I. New York: Wiley, 1975.
[2] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[3] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer, 2001.
[4] IEEE, *802.11: Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. New York: IEEE, 1997.
[5] M.-S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer J. Wirel. Commun.*, vol. 13, no. 12, pp. 119–143, May 2000.
[6] A. J. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.
[7] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Oxford, U.K.: Pergamon, 1960, pp. 3–36.
[8] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of blockfading channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.
[9] M. Assaad and D. Zeghlache, "Cross-layer design in HSDPA system to reduce the TCP effect," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 614–625, Mar. 2006.
[10] L. Le, E. Hossain, and A. Alfa, "Service differentiation in multirate wireless networks with weighted round-robin scheduling and ARQ-based error control," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 208–215, Feb. 2006.
[11] Q. Liu, S. Zhou, and G. B. Giannakis, "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, Sep. 2004.
[12] ——, "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.
[13] C. Comaniciu and H. V. Poor, "Jointly optimal power and admission control for delay sensitive traffic in CDMA networks with LMMSE receivers," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2031–2042, Aug. 2003.
[14] A. Maharshi, L. Tong, and A. Swami, "Cross-layer designs of multichannel reservation MAC under Rayleigh fading," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2054–2067, Aug. 2003.
[15] X. Wang and J. K. Tugnait, "A bit-map-assisted dynamic queue protocol for multi-access wireless networks with multiple packet reception," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2068–2081, Aug. 2003.
[16] ——, "Joint design of channel distribution, truncated ARQ protocol and AMC scheme for multicode CDMA uplink," in *Proc. 38th Conf. Inf. Sci. and Syst.*, May 17–19, 2004.
[17] M. D. Yacoub, J. E. Vargas Bautista, and L. Guerra de Rezende Guedes, "On higher order statistics of the Nakagami-m distribution," *IEEE Trans. Veh. Technol.*, vol. 48, no. 3, pp. 790–794, May 1999.
[18] B. Lu, X. Wang, and J. Zhang, "Throughput of CDMA data networks with multiuser detection, ARQ, and packet combining," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1576–1589, Sep. 2004.
[19] H. Zheng and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 495–506, Mar. 2005.
[20] Z. Kang, K. Yao, and F. Lorenzelli, "Nakagami-m fading modeling in the frequency domain for OFDM system analysis," *IEEE Commun. Lett.*, vol. 7, no. 10, pp. 484–486, Oct. 2003.

Xin Wang (S'03–M'04) received the B.Sc. and M.Sc. degrees from Fudan University, Shanghai, China, in 1997 and 2000, respectively, and the Ph.D. degree from Auburn University, Auburn, AL, in 2004, all in electrical engineering.

From September 2004 to August 2006, he was a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. Since September 2006, he has been an Assistant Professor at the Department of Electrical Engineering, Florida Atlantic University, Boca Raton. His research interests include medium access control, cross-layer design, resource allocation, and signal processing for communication networks.



Qingwen Liu (S'04) received the B.S. degree in electrical engineering and information science from the University of Science and Technology of China, Hefei, China, in 2001 and the M.S. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, in 2003 and 2006, respectively.

He is currently with Navini Networks, Richardson, TX. His interests are in the areas of communications, signal processing, and networking, with emphasis on cross-layer design and radio resource management in wired-wireless networks.

Georgios B. Giannakis (S'84–M'86–SM'91–F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he now holds an ADC Chair in wireless telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 250 journal papers, 400 conference papers, and two edited books. Current research focuses on diversity techniques for fading channels, complex-field and space–time coding, multicarrier, ultrawideband wireless communication systems, cross-layer designs, and sensor networks.

Prof. Giannakis served as Editor-in-Chief for IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE SIGNAL PROCESSING LETTERS, as Secretary of the Signal Processing (SP) Conference Board, as member of the SP Publications Board, as member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a member of the IEEE Fellows Election Committee. He has also served as a member of the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the Steering Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is the (co)recipient of six paper awards from the IEEE SP and Communications Societies (1992, 1998, 2000, 2001, 2003, 2004). He also received Technical Achievement Awards from the SP Society in 2000 and from EURASIP in 2005.