Timely-Throughput Optimal Scheduling with Delayed CSIT for Chase Combining HARQ

Dony J. Muttath, *Student Member, IEEE*, M. Santhoshkumar, *Student Member, IEEE*, and K. Premkumar, *Member, IEEE*

***Abstract*—We consider a cross-layer packet scheduling problem with hybrid ARQ (HARQ) in fading channels in which the channel state information at the transmitter (CSIT) is known only after one slot delay. Packets arrive according to a Bernoulli process at the transmitter, and each packet is required to be timely-delivered at the receiver, within a delay of d time-slots, and is dropped, if the delay deadline is not met. Since the transmitter has only a delayed CSIT, a HARQ with Chase combining is employed for error recovery. *The problem is to decide the transmit-energy in each time-slot such that the timely-throughput is maximum for a given average transmit-energy constraint.* We pose this problem as a constrained Markov decision process, and provide an optimum policy based on Lagrangian relaxation. The optimum Lagrangian multiplier is obtained using a subgradient method. We obtain the structure of the optimum policy, based on which we propose a computationally simple policy READER that requires no CSIT. We show that for large Doppler spread (or mobility), the timely-throughput of READER is close to the optimum policy. We also provide two more policies: an optimum policy assuming perfect CSIT with zero delay, and a naive randomization policy, and compare the throughput performance of the proposed policies.**

***Index Terms*—Chase combining, constrained Markov decision process (CMDP), delayed CSIT, hybrid ARQ (HARQ), maximal ratio combining (MRC), packet delay deadline.**

I. Introduction

DESIGN of low latency systems with high throughput is desirable, as it is envisaged that nearly 79% of total wireless mobile traffic in 2022 would be video traffic [1]. This motivates us to study timely-throughput[1] optimal schedulers. In this work, we are interested in designing a timely-throughput optimum scheduler that efficiently chooses transmit-energy for each transmission opportunity.

Power allocation for fading channels mostly considers either perfect CSIT (with zero delay), or no CSIT [2]-[6], and a very few consider imperfect CSIT [7]. When there is no CSIT, or imperfect CSIT, to mitigate fading, diversity techniques and/or hybrid ARQ (HARQ) processes (if feedback is available) are used. In type II HARQ, where retransmissions of a packet are combined, there are mainly two methods: i) Chase combining HARQ (CC-HARQ) and ii) incremental redundancy (IR-ARQ) [2]. Type II HARQ is particularly suitable for short packet communications (e.g., Internet of Things).

A. Previous Work

Resource allocation with HARQ for fading channels has gained much attention in the recent past among the cross layer community [3]-[7]. For timely-throughput, only a few works [8]-[11] have been reported in the literature.

A downlink scheduling problem with delayed CSIT is studied in [3] with Markov ON/OFF channels. A Whittle's index policy for this problem has been studied and its asymptotic optimality properties in the limiting regime of many users has been shown. In [4], a delay optimal scheduling problem is studied for a multistate fading channel. [5] studies the tradeoff between energy efficiency and retransmission attempts in CC-HARQ. [6] proposes an HARQ scheme that outperforms the conventional HARQ schemes, but without any constraints on transmit-energy. [7] studies CC-HARQ with imperfect CSIT, and explores the tradeoff between energy efficiency and throughput. None of the work consider *delay deadlines, delayed CSIT, and transmit energy consumption collectively for throughput optimal scheduling,* which is the focus of our work.

Scheduling packets with delay deadlines has been explored in the following works. In [8], Collins and Cruz consider an average energy minimization problem over a finite horizon with an average packet delay deadline constraint. They show that the optimum policy is to transmit when the queue-length is larger than a channel-state dependent threshold. Fu *et al.* consider a finite horizon throughput optimal scheduling problem with individual packet deadlines and with a total energy constraint [9], and obtain optimal policies based on dynamic programming. An offline throughput optimal scheduler is proposed in [10], where packet arrivals are known a priori. Almost all the work described above assume causal CSIT, and provide information theoretic rate optimal solutions. In real-time systems, it is not possible to obtain CSIT without any delay. Also, rate maximization is not practical, as Shannon's capacity formula is valid only in the regime of infinite codeword length, and hence, is not suitable for systems with packet delay deadlines. For this reason, we consider a signal-to-noise ratio (SNR) model, which prescribes a target SNR to achieve a certain reliability [11]. In this work, we consider a cross layer scheduling problem with CC-HARQ in fading channels to maximize the timely-throughput with a limitation on the average transmit-energy, where only a one slot delayed CSIT is available.

B. Contributions of the Paper

In this work, we have the following contributions:

The authors are with the Communications Research (CoRe) Lab, Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Design and Manufacturing (IIITDM), Kancheepuram, Chennai, India (e-mail: kpk@iiitdm.ac.in).

[1] Timely-throughput is defined as the fraction of time-slots that deliver packets within the delay deadline requirement.

1) We propose a cross layer scheduling problem in which the sender has only one-slot delayed CSIT, and each packet has an individual delay deadline constraint. In the literature, the study of CC-HARQ is limited to a constraint on the number of retransmissions which does not take into account the delay between retransmissions. As the delay between retransmissions is of significance, it is important to consider timely throughput rather than the number of retransmissions.
2) The cross layer problem i) schedules timely packet transmissions (at the MAC layer), and ii) combines retransmissions (at the PHY layer) by maximal ratio combining (MRC) for optimum use of transmit-energy by CC-HARQ. Also, the sender has a time-average transmit-energy constraint, whereas much of the literature considers only an expected power constraint. We solve the problem of choosing an optimum transmit-energy that maximizes timely-throughput in a time-average transmit-energy constrained system.
3) We obtain an optimal policy, and show that *the optimum policy is a randomization of two deterministic stationary policies, both of which are optimal policies of a relaxed Lagrangian problem.* From the structure of the optimal policy, we propose a computationally less complex policy, READER, which requires no CSIT. We show that the timely-throughput of READER is close to the optimal policy for large Doppler spread.
4) We also quantify the throughput loss due to the delay in CSIT by evaluating the system having perfect CSIT with zero delay. For small Doppler spread, there is no performance loss due to delay in CSIT, whereas for large Doppler spread, there is a loss in throughput due to delayed CSIT. Also, we compare the optimum policy and READER with Blind, a policy that naively randomizes all possible fixed transmit-energy policies.

C. Organization of the Paper

The Paper is organised as follows. We present system model in Section II and pose an optimum scheduling problem in Section III. We present the optimum scheduler, and provide a computationally less complex algorithm in Section IV. In Section V, we provide a Q-learning policy to obtain optimal Lagrangian policy, and also provide a subgradient method to compute optimal Lagrangian multiplier λ^*. Numerical results are provided in Section VI and conclude in Section VII.

II. System Model

A. Network Model

We consider a point-to-point communication in which a sender sends packets across a fading link to a receiver. A discrete time system is considered in which time is measured in slots, indexed by $t \in \mathbb{Z}_+$, where $\mathbb{Z}_+ := \{0, 1, 2, \cdots\}$. The duration of a time-slot is the same as a packet transmission time.

At the beginning of each time-slot t, either a new packet arrives at the sender, which is denoted by $A[t] = 1$, or no new packet arrives, denoted by $A[t] = 0$. The arrival process $\{A[t], t \in \mathbb{Z}_+\}$ is independent and identically distributed (i.i.d.), and $A[t] \sim \text{Bernoulli}(p)$, for some $0 < p < 1$. A packet upon arrival is stored in a transmit-buffer until it is successfully delivered to the receiver, or dropped. Each packet has a delay deadline constraint of d time-slots, which is defined as follows. *A packet that arrives at time-slot t is required to be successfully received at the receiver before the end of time-slot $t + d$, i.e., within a delay of d time-slots.* If for some reason, the packet is not successfully delivered before $t + d$, the packet is dropped from the buffer, which is considered as packet loss.

Transmit Queue: At the beginning of time-slot t, let $Q[t]$ be the number of packets waiting for transmission in the transmit-buffer. The packets are numbered $1, 2, \cdots, Q[t]$ starting from the head of line (HOL) packet, and let the waiting time of packet i at time-slot t be $W_i[t]$. Define $\boldsymbol{W}[t] = [W_1[t], W_2[t], \cdots, W_{Q[t]}[t]]$. Note that the waiting time of a packet at its arrival epoch is zero. Thus, the queue is described by the tuple $(Q[t], \boldsymbol{W}[t])$.

The departure of the HOL packet during time-slot t is denoted by $X[t+1] = 1$ (and no departure by $X[t+1] = 0$). If $W_1[t] = d$ and $X[t+1] = 0$, at the beginning of time-slot $t + 1$, the HOL packet has been in transmit-buffer for $d + 1$ time-slots and hence, will be dropped at time-slot $t + 1$. Thus,

$$\begin{aligned} &Q[t+1] \\ &= \begin{cases} Q[t] - X[t+1] + A[t+1], & \text{if } W_1[t] < d, \\ Q[t] - 1 + A[t+1], & \text{if } W_1[t] = d, \end{cases} \\ &= Q[t] - D[t+1] + A[t+1], \end{aligned} \tag{1}$$

where $D[t+1] = 1$, if either $X[t+1] = 1$, or $W_1[t] = d$; otherwise, $D[t+1] = 0$. If $Q[t] - D[t+1] > 0$, the waiting time of packets $i = 1, 2, \cdots, Q[t] - D[t+1]$ is

$$\begin{aligned} &W_i[t+1] \\ &= \begin{cases} W_{i+1}[t] + 1, & \text{if } W_1[t] = d \text{ or } X[t+1] = 1, \\ W_i[t] + 1, & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

If there is a new arrival, i.e., $A[t+1] = 1$, then the waiting time of the packet that has just arrived, $W_{Q[t+1]}[t+1] = 0$. From (1) and (2), the evolution of queue length and waiting time can be expressed as

$$Q[t+1] = \Psi_Q(Q[t], \boldsymbol{W}[t], X[t+1], A[t+1]), \tag{3}$$
$$\boldsymbol{W}[t+1] = \Psi_W(Q[t], \boldsymbol{W}[t], X[t+1], A[t+1]). \tag{4}$$

B. Transmission Model

Channel model: Wireless channels vary with time and mobility of sender/receiver, and can be modelled as a finite state Markov chain (FSMC) [12]. In a K state FSMC, there are K channel states, where the channel state represents power gain. In each time-slot t, the channel state is denoted by $G[t] \in \mathbb{G} = \{g_0, g_1, g_2, \cdots, g_{K-1}\}$. Note that the states are ordered such that $0 = g_0 < g_1 < g_2 < \cdots < g_{K-1}$. The channel state (or gain) $\{G[t] : t \in \mathbb{Z}_+\}$ follows a Markov chain with the transition probability for each $g, g' \in \mathbb{G}$ being

$$\mathbb{T}(g, g') = \begin{cases} 0, & \text{if } |g - g'| > 1, \\ \alpha_{g,g'} f_D, & \text{if } |g - g'| = 1, \end{cases} \tag{5}$$

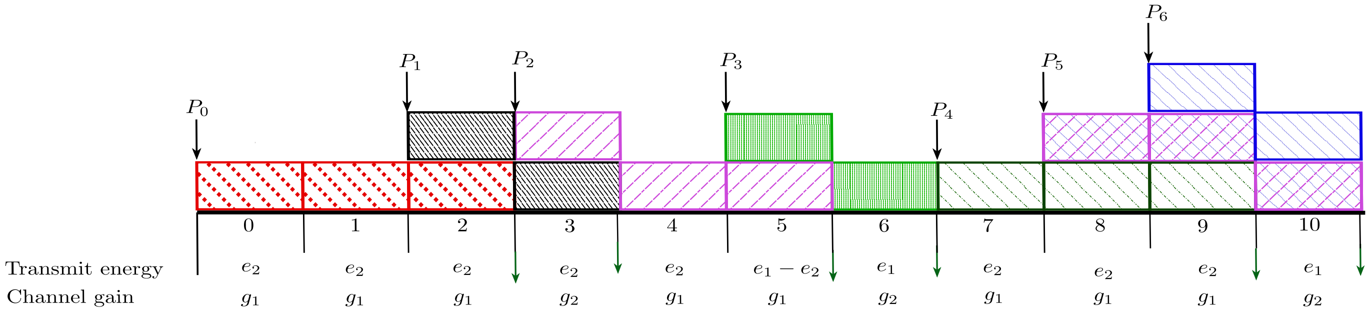

P_0
P_1
P_2
P_3
P_4
P_5
P_6
0
1
2
3
4
5
6
7
8
9
10
Transmit energy
e_2 e_2 e_2 e_2 e_2 $e_1 - e_2$ e_1 e_2 e_2 e_2 e_1
Channel gain
g_1 g_1 g_1 g_2 g_1 g_1 g_2 g_1 g_1 g_1 g_2

Fig. 1. An illustration of scheduling for a delay deadline of $d = 2$ slots (i.e., a packet can remain in the system for at most 3 slots), $\mathbb{G} = \{g_1, g_2\}$, and energy e_2 is such that $2e_2 < e_1$ and $3e_2 \geqslant e_1$. Packets P_0, P_1, P_2, P_3, P_4, and P_5 arrive at the beginning of time-slots 0, 2, 3, 5, 7, and 8, and depart at the end of time-slots 2, 3, 5, 6, 9, and 10, respectively. Multiple transmissions of packets P_0, P_2, P_4 are combined at the receiver by MRC to meet the target SNR. In time-slot 3, e_2 is just enough for channel gain g_2, whereas in time-slot 10, e_1 is more than required for channel gain g_2, and hence, packets P_1 and P_5 are received successfully. *Note that the channel gains are known at the sender only after one time-slot delay.*

where f_D is the Doppler frequency (that depends on the mobility), and $\alpha_{g,g'}$ is a parameter that depends on distribution of fading gain, level crossing rate of fades, and slot duration, [12, Eqns. (18)–(20)]. We consider a one-slot delayed CSIT, where the CSIT at time-slot t, denoted by $\tilde{G}[t]$, is the channel state at time-slot $t-1$, i.e., $\tilde{G}[t] = G[t-1]$.

Communication model: We consider an SNR model where the successful reception of a packet requires an SNR of least γ. In CC-HARQ scheme, the SNR of a packet after combining is the sum of SNRs of all transmissions of the packet (see Figure. 1 for an illustration). For a channel state $g_i > 0$, a transmit energy e_i is required to achieve a target SNR of γ in a single transmission attempt. Since, the sender has only one-slot delayed CSIT, it may happen that the chosen transmit energy may not achieve an SNR of γ at the combiner, requiring at least one more round of HARQ for a successful reception of the packet. For channel state $g_0 = 0$, no transmit-energy can achieve the target SNR of γ, and hence, to conserve energy, it is best not to transmit when $G[t] = g_0$.

If the transmit queue is non-empty, the controller decides either to transmit the HOL packet (by choosing a positive transmit energy), or not (by choosing 0 transmit energy). Let $U[t]$ be the transmit energy chosen by the controller during time-slot t. Note that $U[t] \in \{0, e_1, e_2, \cdots, e_{K-1}\} =: \mathbb{U}_0$, where $0 > e_1 > e_2 > \cdots > e_{K-1}$. For retransmission, we choose transmit-energy from $\mathbb{U}_1 = \mathbb{U}_0 \cup \{e_i - e_j : 1 \leqslant i < j \leqslant K-1\}$, as $e_i - e_j$s could possibly reduce average transmit-energy. Let $R[t]$ denote the sum of SNRs of all previous transmissions of the current HOL packet at the receiver before time-slot t.

Recall that $X[t+1]$ indicates whether or not the HOL packet is delivered successfully during time-slot t. We note that when the action $U[t] = 0$, $X[t+1] = 0$. For $U[t] > 0$,

$$X[t+1] = \begin{cases} 1, & \text{if } R[t] + \frac{U[t]\tilde{G}[t+1]}{N_0/2} \geqslant \gamma, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $N_0/2$ is the AWGN power spectral density. Note that

$$R[t+1] = \begin{cases} R[t] + \frac{U[t]\tilde{G}[t+1]}{N_0/2}, & \text{if } X[t+1] = 0, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

C. Timely-Throughput

In each time-slot t, at most one packet is successfully received, which is given by $X[t+1]$. We define timely-throughput as long term average number of packets per time-slot that are delivered within the delay deadline, i.e.,

$$\eta = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1} X[t]\right]. \tag{8}$$

A maximum throughput is achieved by choosing a maximum $U[t]$ in every time-slot. However, this strategy increases the average transmit-energy. In this work, we are interested in maximizing the throughput for a given average transmit-energy.

III. Optimal Scheduling Problem

In this Section, we describe the throughput optimum scheduling problem for a given average energy constraint, and provide an optimum solution.

At time-slot t, we define the state of the system by $\boldsymbol{S}[t] = [X[t], Q[t], \boldsymbol{W}[t], R[t], \tilde{G}[t]]$. The scheduling decision is to choose an energy $U[t] \in \mathbb{U}_0$ when $R[t] = 0$, and $U[t] \in \mathbb{U}_1$ when $R[t] \neq 0$ to transmit the HOL packet; $U[t] = 0$ is chosen to not transmit the HOL packet, or when $Q[t] = 0$. The scheduling problem that we consider is to choose $U[t]$ during each time-slot t such that the throughput is maximum. If there is no packet delay deadline constraint, one may always choose the smallest energy for transmission. Similarly, when there is no energy constraint, one can always choose e_1 (largest energy) for transmission. Thus, there is a tradeoff between the packet delay deadline constraint d and the average transmission energy. Let $\pi = (\mu_0, \mu_1, \mu_2, \cdots)$ be a policy. Under policy π, define the average reward

$$J^\pi(\boldsymbol{s}_0) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\left[\sum_{t=0}^{T-1} X[t] \middle| \boldsymbol{S}[0] = \boldsymbol{s}_0\right], \tag{9a}$$

and the average energy

$$C^\pi(\boldsymbol{s}_0) := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\left[\sum_{t=0}^{T-1} U[t] \middle| \boldsymbol{S}[0] = \boldsymbol{s}_0\right]. \tag{9b}$$

Let Π_{NA} denote the class of non-anticipated policies, i.e., any scheduling policy $\pi = (\mu_0, \mu_1, \mu_2, \cdots) \in \Pi_{\text{NA}}$ is a sequence of functions, where each function μ_t depends on the state sequence $\boldsymbol{s}_0, \boldsymbol{s}_1, \cdots, \boldsymbol{s}_t$ and the action sequence $u_0, u_1, \cdots, u_{t-1}$. In this Paper, we obtain an optimal policy $\pi^* \in \Pi_{\text{NA}}$ that maximizes the throughput, while the time average transmit-energy constraint is satisfied. The energy constrained scheduling problem is defined as follows.

Problem 1 (Energy Constrained Scheduling):

$$\max_{\pi \in \Pi_{\text{NA}}} \quad J^{\pi}(\boldsymbol{s}_0), \tag{10a}$$

$$\text{s.t.} \quad C^{\pi}(\boldsymbol{s}_0) \leqslant \bar{e}, \tag{10b}$$

where $\bar{e}$ is the average transmit-energy constraint. Let $\pi^* \in \Pi_{\text{NA}}$ be a solution to *Problem 1*. Since, *Problem 1* is a constrained MDP (CMDP) with a finite state space and a finite action space, the optimal policy π^* need not be stationary deterministic [13].

We relax the constraint in (10a), and propose a Lagrangian relaxed problem. For the Lagrangian with Lagrange multiplier $\lambda \geqslant 0$, the reward of a policy π is defined as,

$$J^{\pi}_{\lambda}(\boldsymbol{s}_0) := \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_{\pi} \left[\sum_{t=0}^{T-1} X[t] - \lambda U[t] \middle| \boldsymbol{S}[0] = \boldsymbol{s}_0 \right], \tag{11}$$

The unconstrained scheduling problem is thus formulated as,

Problem 2 (Unconstrained Scheduling):

$$J^*_{\lambda} = \max_{\pi \in \Pi_{\text{NA}}} J^{\pi}_{\lambda}(\boldsymbol{s}_0), \tag{12a}$$

$$\text{where,} \ \pi^*_{\lambda} \in \arg\max_{\pi \in \Pi_{\text{NA}}} J^{\pi}_{\lambda}(\boldsymbol{s}_0). \tag{12b}$$

A policy π is called λ-optimal if it achieves J^*_{λ}. Note that the one-stage reward of the *Unconstrained Scheduling Problem* for state $\boldsymbol{s} = [x, q, \boldsymbol{w}, r, \tilde{g}]$ and action u is

$$F(\boldsymbol{s}, u) = x - \lambda u, \tag{13}$$

where $\lambda \geqslant 0$ is the Lagrangian cost for using energy u.

In *Problem 2 (Unconstrained Scheduling)*, we have an average reward MDP for a finite state and action spaces, and hence, a stationary deterministic λ-optimal policy π^*_{λ} exists, which can be computed by a value iteration method Let $\pi^* = [\mu^*, \mu^*, \cdots]$ be the stationary optimal policy. The value iteration is given by

$$v_0(\boldsymbol{s}) = 0, \tag{14}$$

$$v_{i+1}(\boldsymbol{s}) = \max_{u \in \mathbb{U}} \left[F(\boldsymbol{s}, u) + \mathbb{E}\left[v_i(\boldsymbol{s}') \middle| \boldsymbol{s}, u \right] \right], \tag{15}$$

$$\mu_{i+1}(\boldsymbol{s}) = \arg\max_{u \in \mathbb{U}} \left[F(\boldsymbol{s}, u) + \mathbb{E}\left[v_i(\boldsymbol{s}') \middle| \boldsymbol{s}, u \right] \right], \tag{16}$$

$$\mu^*(\boldsymbol{s}) = \lim_{i\to\infty} \mu_i(\boldsymbol{s}). \tag{17}$$

where $\boldsymbol{s}'$ is the next state of the system when the current state is $\boldsymbol{s}$ and the current action is u.

IV. Optimal Scheduler

In this Section, we provide the optimal scheduling policy for *Problem 1* based on results from [13]. The existence and the structure of optimal policy for *Problem 1* is given as follows.

Theorem 1: For *Problem 1*, a stationary optimal policy π^* exists. π^* randomizes between stationary deterministic policies π^*_1 and π^*_2 with probabilities θ and $1-\theta$. The policies π^*_1 and π^*_2 are λ^*-optimal policies for *Problem 2* for some $\lambda = \lambda^*$.

Proof: Let C_{λ} be the average transmit-energy of λ-optimal policy. For a given $\bar{e}$, define λ^* as

$$\lambda^* = \inf\left\{\lambda > 0 : C_{\lambda} \leqslant \bar{e}\right\}. \tag{18}$$

Since there exists a policy $f(\boldsymbol{s}) := 0, \forall \boldsymbol{s}$ such that $C^f(\boldsymbol{s}_0) \leqslant \bar{e}$, we have from Lemma 3.3 of [13], $\lambda^* < \infty$, and Hypothesis 4.1 of [13] is true. Hence, from Theorem 4.4 of [13], there exists an optimal stationary randomized policy for *Problem 1*. ■

To find the optimal policy, we need λ^*, π^*_1, π^*_2 and θ. λ^* is given by (18). For a given λ^*, consider an increasing sequence $\{\lambda_n\} \to \lambda^*$, and a decreasing sequence $\{\lambda'_n\} \to \lambda^*$. Let

$$\lim_{\lambda_n \uparrow \lambda^*} C_{\lambda_n} = \alpha_1, \quad \lim_{\lambda'_n \downarrow \lambda^*} C_{\lambda'_n} = \alpha_2 \tag{19}$$

where $\alpha_2 \leqslant \bar{e} \leqslant \alpha_1$. It is shown in [13] that $\{\pi^*_{\lambda_n}\}$ converges to λ^*-optimal policy π^*_1 and $\{\pi^*_{\lambda'_n}\}$ converges to λ^*-optimal policy π^*_2. Hence, θ is chosen such that $C^{\pi^*}(\boldsymbol{s}_0) = \bar{e}$, i.e.,

$$\theta = \frac{\bar{e} - \alpha_2}{\alpha_1 - \alpha_2}. \tag{20}$$

Since, the optimal policy is computationally intensive, we propose a sub-optimal algorithm in the next Section.

A. Randomization of Energy Adjacent Decision Rules (READER)

Define $e_0 = 0$. For any state $\boldsymbol{s} = [x, q, \boldsymbol{w}, r, \tilde{g}]$, define the deterministic policies $f_0, f_1, f_2, \cdots, f_K$, as follows.

$$f_i(\boldsymbol{s}) = \begin{cases} 0, & \text{if } q = 0, \\ e_i, & \text{if } q > 0, \end{cases}$$

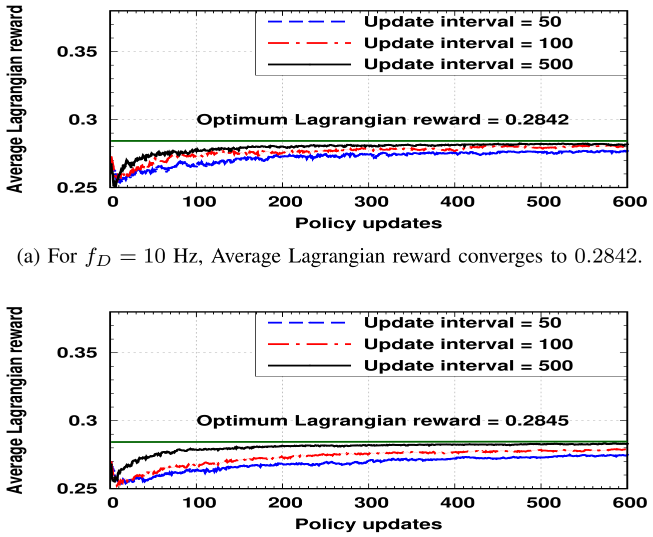
Let $C^{f_i}(\boldsymbol{s}_0) = E_i$. Note that $E_1 > E_2 > \cdots > E_{K-1} > E_0 = 0$. We propose a policy, Randomization of Energy Adjacent Decision Rules (READER) that computes action U_t for a given state $\boldsymbol{s}_t$ as follows.

Algorithm 1 READER

```
Input: $\boldsymbol{s}_t, \bar{e}, E_0, E_1, E_2, \cdots, E_{K-1}$
Output: $U_t$
if $\bar{e} \geqslant E_1$ then
    Choose action $U_t = f_1(\boldsymbol{s}_t)$
else if $E_{j+1} \leqslant \bar{e} < E_j$ for some $j = 1, 2, \cdots, K-2$ then
    Randomize between policies $f_{j+1}$ and $f_j$
    Choose action
        $$U_t = \begin{cases} f_{j+1}(\boldsymbol{s}_t), & \text{w.p. } \theta, \\ f_j(\boldsymbol{s}_t), & \text{w.p. } 1-\theta. \end{cases}$$
    where $\theta$ is chosen such that $\bar{e}$ is achieved.
end if
return $U_t$
```

V. Q-Learning Policy

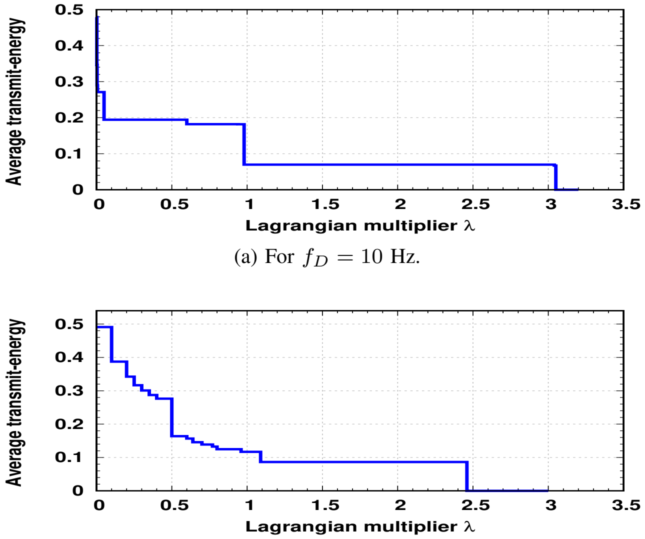
An optimum policy π^*_{λ} for a given Lagrangian multiplier λ can be computed using two time-scale Q-learning [14].



(a) For $f_D = 10$ Hz, Average Lagrangian reward converges to 0.2842.

(b) For $f_D = 100$ Hz, Average Lagrangian reward converges to 0.2845.

Fig. 2. Comparison of Q-learning with optimum policy for $\hat{e}_1 = 1$, $\hat{e}_2 = 0.3219$, $P = [0.2, 0.3, 0.5]$, $p = 0.4$, $d = 2$, $\lambda = 0.2$, and 20 trials.



(a) For $f_D = 10$ Hz.

(b) For $f_D = 100$ Hz.

Fig. 3. Average transmit-energy vs Lagrangian multiplier λ, for $\hat{e}_1 = 1$, $\hat{e}_2 = 0.3219$, $P = [0.2, 0.3, 0.5]$, $p = 0.4$, and $d = 2$.

Algorithm 2 Q-learning

```
Require: Update interval $I$, $\hat{\boldsymbol{s}}, \hat{u}$, $\delta_t(\boldsymbol{s}, u) = \frac{\mu}{\text{Visit}_t(\boldsymbol{s}, u)}$
Initialize $Q_0(\boldsymbol{s}, u)$ arbitrarily $\forall \boldsymbol{s}, u$
Choose initial state $\boldsymbol{s}_0$ arbitrarily
for $n = 0, 1, 2, \cdots$ do
    Update policy $\mu_n(\boldsymbol{s}) = \arg\max_u Q_{nI}(\boldsymbol{s}, u), \forall \boldsymbol{s} \in \mathbb{S}$
    for $t = nI, nI + 1, \cdots, (n+1)I - 1$ do
        For $\boldsymbol{s}_t$, choose $u_t$ from $\mu_n(\boldsymbol{s}_t)$ by $\epsilon$-greedy policy
        Take action $u_t$, observe reward $r_t = F(\boldsymbol{s}_t, u_t)$, and observe the next state $\boldsymbol{s}_{t+1}$
        Set $Q_{t+1}(\boldsymbol{s}_t, u_t; \lambda) = (1-\delta_t)Q_t(\boldsymbol{s}_t, u_t; \lambda) + \delta_t\big[r_t + \max_{u'} Q_k(\boldsymbol{s}_{t+1}, u'; \lambda) - Q_t(\hat{\boldsymbol{s}}, \hat{u}; \lambda)\big]$
    end for
end for
```

In the Q-learning algorithm, $\hat{\boldsymbol{s}}, \hat{u}$ is an arbitrary state-action pair, $\mu > 0$ is a constant, and $\text{Visit}_t(\boldsymbol{s}, u)$ is the number of times the state-action pair $(\boldsymbol{s}, u)$ is visited in the first t iterations.

We recall that for Lagrangian multiplier λ, the average transmit-energy of the λ-optimal policy is given by C_λ. Also, from (18), for a given $\bar{e}$, we are interested in obtaining the optimal Lagrangian multiplier, $\lambda^* = \inf\{\lambda > 0 : C_\lambda \leqslant \bar{e}\}$. We adapt the subgradient method shown in [14, (43)] to iteratively find λ^* as follows.

$$\lambda_{k+1} = \lambda_k + \epsilon_k\big(C_{\lambda_k} - \bar{e}\big), \quad k = 1, 2, 3, \cdots, \qquad (21)$$

where $\epsilon_k = 1/k$ and $\lambda_1 \geqslant 0$ is chosen arbitrarily.

VI. Numerical Results

We evaluate timely-throughput, defined in (8), as a function of average transmit-energy constraint $\bar{e}$. A 3-state Rayleigh fading channel is considered, with carrier frequency of 900 MHz, symbol transmission rate of 1 Mb/s, and packet size of 1000 bits. A medium/low mobility with a speed of 3.33 m/s, and a high mobility with a speed of 33.3 m/s are considered. The corresponding Doppler spreads are $f_D = 10$ Hz and 100 Hz, respectively. The probability mass function (pmf) of channel states considered is $P_G = [0.2, 0.3, 0.5]$, from which (using [12, Eqn. (14)]) we compute channel gains $g_1 = 0.2231$ and $g_2 = 0.6931$. The transition probabilities $\mathbb{T}(g, g')$ are computed using [12, Eqns. (18)–(20)] for $f_D = 10$ Hz and 100 Hz. A target SNR of $\gamma = 6.7895$ dB or 4.7748 is considered[2]. From γ and g_is, we get $e_1 = 21.3978 N_0$[3], and $e_2 = 6.8885 N_0$. All energies are normalized with respect to e_1, i.e., $\hat{e}_1 = e_1/e_1 = 1$, and $\hat{e}_2 = e_2/e_1 = 0.3219$. Thus, the energy constraint $\bar{e}$ is also normalized with respect to e_1. We consider Bernoulli(p) packet arrivals with $p = 0.4$.

In Figure 2, for $\lambda = 0.2$, we run the Q-learning algorithm for 20 trials, and plot the trial averaged Lagrangian reward versus the iteration index. We observe that the trial averaged Lagrangian reward converges to $J^*_{0.2}$ (see (12a)).

In Figure 3, we plot the average transmit-energy C_λ as a function of λ. λ^* for each $\bar{e}$ can be computed from Figure 3. At any point of discontinuity λ, the optimal policy π^*_λ is a mixture of policies computed from the Lagrangian multipliers $\lambda - \nu$ and $\lambda + \nu$, where $\nu > 0$ is arbitrarily close to 0.

In Figures 4 and 5, we plot the timely-throughput of various policies that we propose as a function of the average transmit-energy, $\bar{e}$. In order to quantify the effect of delayed CSIT, we consider the optimal policy for the case of perfect CSIT with zero-delay, which we call Bound (motivated by [9], and is a finite state fading version of [11]). Also, we consider a Blind randomized policy in which for any state with non-zero queue, the action is randomized between all energy levels such that the average energy constraint is met. We plot the timely-throughput of optimal policy, READER, Bound, and Blind in Figures 4 and 5, and compare their performance.

From Figures 4 and 5, we see that the optimal policy performs almost the same as that of Bound for $f_D = 10$ Hz,

[2] For a BER target of 10^{-3}, and for BPSK, γ is given by $Q(\sqrt{2\gamma}) = 10^{-3}$.

[3] Recall that N_0 is the AWGN power spectral density.

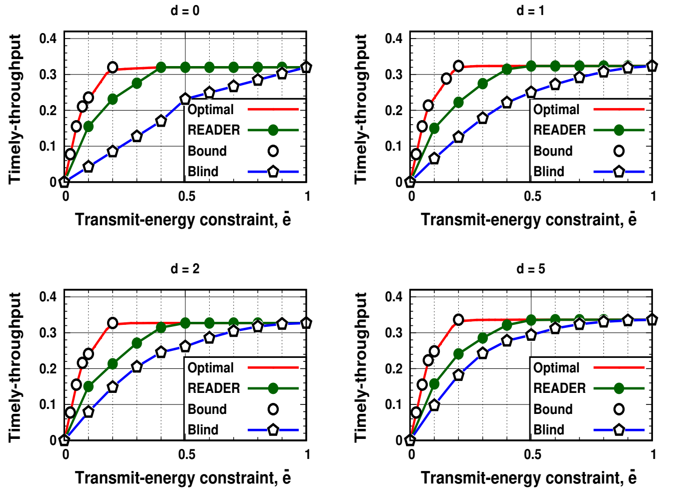


Fig. 4. Timely-throughput vs transmit-energy constraint, $\overline{e}$ for Doppler spread $f_D = 10$ Hz with $\hat{e}_1 = 1, \hat{e}_2 = 0.3219$, $P = [0.2, 0.3, 0.5]$, $p = 0.4$.

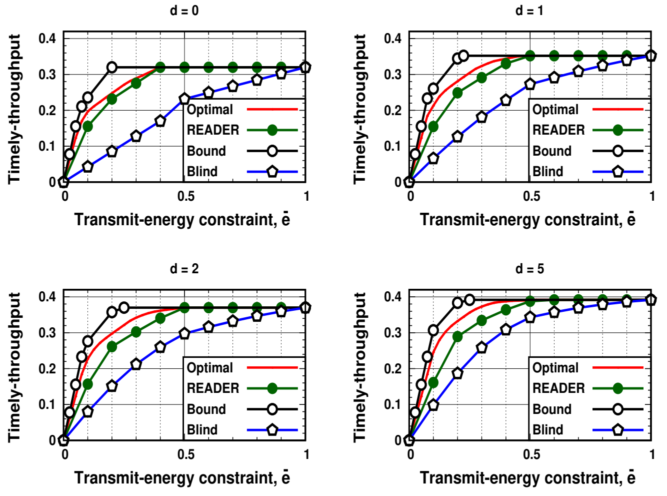


Fig. 5. Timely-throughput vs transmit-energy constraint, $\overline{e}$ for Doppler spread $f_D = 100$ Hz with $\hat{e}_1 = 1, \hat{e}_2 = 0.3219$, $P = [0.2, 0.3, 0.5]$, $p = 0.4$.

as the channel has more memory in this case. For each policy, for a given $\overline{e}$, as the delay deadline constraint d increases, throughput also increases. This is because, for a large d, the probability of finding a slot with a better channel state is large. For $f_D = 100$ Hz, the channel changes fast, and hence, READER (requiring a no CSIT) performs close to the optimal policy. Also, $f_D = 100$ Hz achieves more throughput than $f_D = 10$ Hz, due to inherent time-diversity.

VII. Conclusions

We have investigated a packet scheduling problem with CC-HARQ in a fading link with delay deadlines, and average transmit-energy constraint. We have formulated the problem as a CMDP, and have obtained the optimum policy which is shown to be a randomization of two stationary deterministic policies. As the optimal policy is computationally intensive, we propose the following policies: i) READER, which is a randomization of adjacent energy policies, and ii) Blind that randomizes across policies having the same transmit-energy levels. We compare the throughput performance of the optimal and the heuristic policies. We show that for $f_D = 10$ Hz, the channel memory alleviates the delay in CSIT, and for $f_D = 100$ Hz, the performance of READER (which requires no CSIT) is close to the optimal policy, as for $f_D = 100$ Hz, the channel behaves more independently.

Our work can be extended for IR-HARQ by considering a state model that includes the number of coded bits that are successfully received in all previous transmissions of the current HOL packet. A packet is successfully delivered, if the total number of coded bits that are successfully received crosses a threshold. Thus, our system model can be modified for IR-HARQ, which can be explored as a future work.

References

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022," https://s3.amazonaws.com.

[2] T. V. K. Chaitanya and E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in i.i.d. Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1835–1846, 2013.

[3] W. Ouyang, A. Eryilmaz, and N. B. Shroff, "Low-complexity optimal scheduling over time-correlated fading channels with ARQ feedback," *IEEE Trans. Mobile Comput.*, vol. 15, no. 9, pp. 2275–2289, Sep. 2016.

[4] J. Liu, W. Chen, and K. B. Letaief, "Delay optimal scheduling for ARQ-aided power-constrained packet transmission over multi-state fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7123–7137, 2017.

[5] J. P. Battistella Nadas, O. Onireti, R. D. Souza, H. Alves, G. Brante, and M. A. Imran, "Performance analysis of hybrid ARQ for ultra-reliable low latency communications," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3521–3531, 2019.

[6] M. Shirvanimoghaddam, H. Khayami, Y. Li, and B. Vucetic, "Dynamic HARQ with guaranteed delay," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[7] A. Chelli, E. Zedini, M.-S. Alouini, M. Pätzold, and I. Balasingham, "Throughput and delay analysis of HARQ with code combining over double Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4233–4247, 2018.

[8] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *1999 Allerton Conf. on Commun., Control., and Comp.*, 1999, pp. 709–717.

[9] A. Fu, E. Modiano, and J. N. Tsitsiklis, "Optimal transmission scheduling over a fading channel with energy and deadline constraints," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 630–641, Mar. 2006.

[10] W. Chen, U. Mitra, and M. J. Neely, "Energy-efficient scheduling with individual packet delay constraints over a fading channel," *Wireless Networks*, vol. 15, no. 5, pp. 601–618, Jul. 2009.

[11] D. J. Muttath, M. Santhoshkumar, and K. Premkumar, "Energy optimal packet scheduling with individual packet delay constraints," in *2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, 2018, pp. 1–6.

[12] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-state markov modeling of fading channels - a survey of principles and applications," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 57–80, 2008.

[13] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint," *Journal of Mathematical Analysis and Applications*, vol. 112, no. 1, pp. 236–252, 1985.

[14] D. V. Djonin and V. Krishnamurthy, "Q-learning algorithms for constrained Markov decision processes with randomized monotone policies: Application to MIMO transmission control," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2170–2181, 2007.