Delay-Optimal Buffer-Aware Scheduling With Adaptive Transmission

Xiang Chen, *Student Member, IEEE*, Wei Chen, *Senior Member, IEEE*, Joohyun Lee, *Member, IEEE*, and Ness B. Shroff, *Fellow, IEEE*

Abstract— In this paper, we aim to obtain the optimal tradeoff between the average delay and the average power consumption in a communication system. In our system, the arrivals occur at each timeslot according to a Bernoulli arrival process, and are buffered at the transmitter waiting to be scheduled. We consider a finite buffer and allow the scheduling decision to depend on the buffer occupancy. In order to capture the realism in communication systems, the transmission power is assumed to be an increasing and convex function of the number of packets transmitted in each timeslot. This problem is modeled as a constrained Markov decision process (CMDP). We first prove that the optimal policy of the Lagrangian relaxation of the CMDP is deterministic and threshold-based. We then show that the optimal delay-power tradeoff curve is convex and piecewise linear, and the optimal policies of the original problem are also threshold-based. Based on the results, we propose an algorithm to obtain the optimal policy and the optimal tradeoff curve. We also show that the proposed algorithm is much more efficient than using general methods. The theoretical results and the algorithm are validated by linear programming and simulations.

Index Terms— Cross-layer design, joint channel and buffer aware scheduling, markov decision process, queueing, energy efficiency, average delay, delay-power tradeoff, linear programming.

I. INTRODUCTION

SCHEDULING for minimizing delay or power has been studied widely and is getting increasingly important, as many delay sensitive applications are emerging, such as instant messenger (IM), social network service (SNS), streaming media and so on. On the other hand, the requirements of mobility and portability for communication terminals incur stringent energy constraints.

In typical communication systems, for fixed channel conditions, the power efficiency (per bit transmitted) rapidly
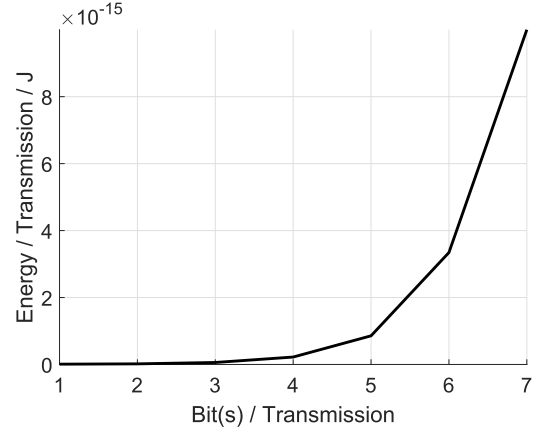
Fig. 1. Energy/Transmission versus Bit/Transmission with adaptive M-PSK (Target ber$=10^{-5}$, Noise Power Spectral Density $N_0 = -150$ dBm/Hz).

decreases as the transmission rate is increased. In other words, the power cost is convex in transmission rate. Below are two canonical examples of communication systems that demonstrate this convex behavior.

1) The information-theoretically optimal transmission rate $R = \frac{1}{2} \log_2(1 + \frac{P}{N})$. Therefore the power to transmit s bit(s) is $\Theta_s = N(4^s - 1)$, which is strictly increasing and convex.

2) Consider an adaptive M-PSK transmission system with a fixed bit error rate (ber). The ber expression for M-PSK is shown in [1, eq(8.31)]. We fix the ber $= 10^{-5}$ and the one-sided noise power spectral density $N_0 = -150$ dBm/Hz. The energy-bit curve is shown in Fig. 1, which is strictly increasing and convex.

The convexity of power cost in transmission rate brings a natural tradeoff between power and delay. As we increase the transmission rate, the delay becomes shorter with the cost of low power efficiency, and vice versa. Our main goal is to characterize the optimal delay-power tradeoff and obtain an optimal scheduling policy for a given average power constraint.

The optimal delay-power tradeoff and the optimal scheduling policy in the point-to-point communication scenario have been studied in [2]–[12], under the convexity assumption for power cost. Among these works, the power cost is modeled based on Shannon's formula in [3]–[7]. Since there is no interference in the point-to-point scenario, the power cost is convex in the transmission rate (bits/transmission), similar to the information-theoretical example we introduced above. Lagrange multiplier method has been applied in these works, in order to transform the constrained optimization to

unconstrained optimization to simplify the problem. Based on this, the properties of the delay-power tradeoff curve have been studied in [3], [6], [7], and [9], and the monotonicity of the optimal scheduling policy is investigated in [2], [4], [6], and [8]–[12]. However, most papers neither trace back to the original constrained problem, nor prove the equivalence between the original and the Lagrangian relaxation problems. Only in [4] and [9]–[11], properties of the optimal policy for the constrained problem are tackled based on the results from the unconstrained problem. However, in [4], the power cost is fixed by Shannon's formula, thus the results cannot be applied to more generalized power models. In [9], necessary properties for proof such as the unichain property of policies, the multimodularity of costs, and stochastically increasing buffer transition probabilities are not proven, but just assumed to be correct. In [10] and [11], binary control is considered, i.e., the scheduler only determines to transmit or not to transmit.

We studied the optimal scheduling in [13]–[15], considering a single-queue single-server system with fixed transmission rate, and obtained analytical solutions. Interestingly, in these cases, the monotonicity of the optimal policy can be directly obtained by steady-state analysis of the Markov Process and linear programming formulation. Similar approaches have been applied in [16] and [17]. We generalized our model and included the adaptive transmission assumption in [18], which is much harder to analyze because of the more complicated state transition of the Markov chain. In this paper, we continue this line of research, analyze the problem within the CMDP framework, and present our thorough analysis and results. We first consider its Lagrangian relaxed version. In the unconstrained MDP problem, we prove that the optimal policy is deterministic and threshold-based. Then, in the CMDP problem, we fully characterize the optimal delay-power tradeoff. We prove that the tradeoff curve is convex and piecewise linear, whose vertices are obtained by the optimal policies in the relaxed problem. Moreover, the neighboring vertices of the tradeoff curve are obtained by policies which take different actions in only one state. These discoveries enable us to show that the solution to the overall CMDP problem is also of a threshold form, and devise an algorithm to efficiently obtain the optimal tradeoff curve.

The remainder of this paper is organized as follows. The system model is described in Section II, where the delay-power tradeoff problem is formulated as a Constrained Markov Decision Process. In Section III, based on the Lagrangian relaxation of the CMDP problem, it is proven that the optimal policy for the average combined cost is deterministic and threshold-based. Then we conduct steady-state analysis in Section IV, based on which we can prove that the optimal delay-power tradeoff curve is piecewise linear, and the optimal policies for the CMDP problem are also threshold-based. Moreover, in this section, we propose an efficient algorithm to obtain the optimal delay-power tradeoff curve, and an equivalent Linear Programming problem is formulated to confirm the validity and efficiency of the theoretical results and the algorithm. Simulation results are given in Section V, and Section VI concludes the paper.
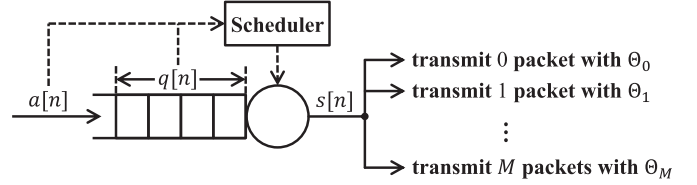

Fig. 2. System Model.

II. SYSTEM MODEL

We consider the system model shown in Fig. 2. Time is divided into timeslots. Assume that at the end of each timeslot, data arrive as a Bernoulli Process with parameter α. Each data arrival contains A packets. Define $\mathcal{A} = \{0, 1\}$. Define $a[n] \in \mathcal{A}$ where $a[n] = 1$ or 0 as whether or not there are data arriving in timeslot n, hence $\Pr\{a[n] = 1\} = \alpha$ and $\Pr\{a[n] = 0\} = 1 - \alpha$.

Let $s[n]$ denote the number of data packets transmitted in timeslot n. We assume that at most S packets can be transmitted in each timeslot because of the constraints of the transmitter. We force $S \geq A$. Define $\mathcal{S} = \{0, 1, \cdots, S\}$, thus $s[n] \in \mathcal{S}$.

Let $p[n]$ denote the power consumption in timeslot n. Transmitting s packet(s) incurs power consumption Θ_s, where $0 \leq s \leq S$. Therefore $p[n] = \Theta_{s[n]}$. Transmitting 0 packet will cost no power, hence $\Theta_0 = 0$. Based on the analysis in the Introduction, being able to capture the convex relationship between power and bits transmitted is important. Therefore we assume that Θ_s is strictly increasing and convex in s.

The arrivals can be stored in a finite buffer up to a maximum of Q packets. For a typical communication system, the buffer size should be larger than the data size of an arrival, thus we set $Q \geq A$. Define $\mathcal{Q} = \{0, 1, \cdots, Q\}$. Let $q[n] \in \mathcal{Q}$ denote the number of packets in the buffer at the beginning of timeslot n. The amount of transmission $s[n]$ will be decided according to our scheduling policy, based on the historical information of the buffer states and the data arrivals. The dynamics of the buffer is given as

$$q[n + 1] = q[n] - s[n] + Aa[n]. \tag{1}$$

To avoid overflow or underflow, the number of transmitted packets in each timeslot n should satisfy $0 \leq q[n] - s[n] \leq Q - A$.

For each timeslot n, respectively consider the queue length $q[n]$ and the transmission $s[n]$ as the state and the action of the system. According to (1), the transition probability

$$\Pr\{q[n + 1] = j | q[n] = q, s[n] = s\}$$
$$= \begin{cases} \alpha & j = q - s + A, \\ 1 - \alpha & j = q - s, \\ 0 & \text{else.} \end{cases} \tag{2}$$

It shows that the probability distribution of the next state is determined by the current state and the chosen action. The queue length $q[n]$ and the transmission power $p[n]$ can be treated as two immediate costs, which are determined by the current state and the current action respectively. Therefore, this system can be considered as a Markov Decision Process (MDP). In the following, we will show that the queue
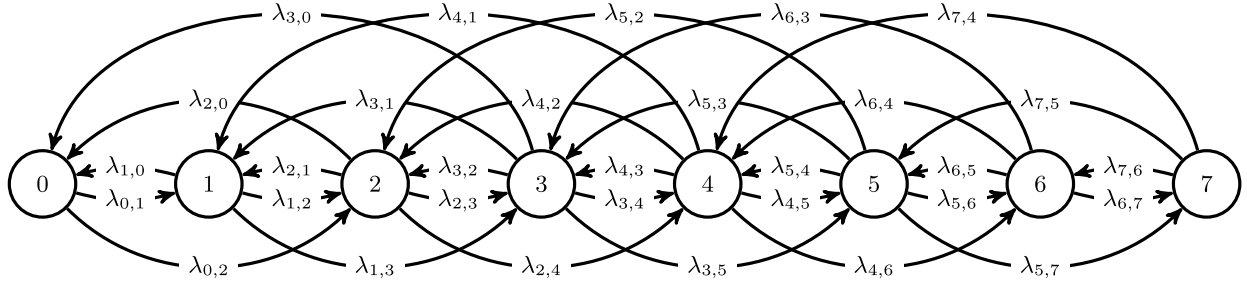
Fig. 3. Markov Chain of $q[n]$ ($Q = 7$, $A = 2$, $S = 3$).

length cost corresponds to the average delay, hence there is a tradeoff between these two costs.

A decision rule $\delta_n : \underbrace{Q \times \mathcal{A} \times Q \times \cdots \times \mathcal{A} \times Q}_{n \text{ state(s) and } (n-1) \text{ action(s)}} \rightarrow \mathcal{P}(S)$
can specify the action $s[n]$ at timeslot n according to a probability distribution $p_{\delta_n(\cdot)}(\cdot)$ on the set of actions S, i.e.,

$$\Pr\{s[n] = s | q[1] = q_1, s[1] = s_1, \cdots, q[n] = q_n\}$$
$$= p_{\delta_n(q_1, s_1, \cdots, q_n)}(s). \tag{3}$$

Define a transmission policy $\gamma = (\delta_1, \delta_2, \cdots)$, which is a sequence of decision rules. Define $\mathrm{E}_{q_0}^{\gamma}\{\cdot\}$ as the notation of the expectation when policy γ is applied and the initial state is q_0. Therefore the average power consumption under policy γ

$$P_{\gamma} = \lim_{N \to \infty} \frac{1}{N} \mathrm{E}_{q_0}^{\gamma} \left\{ \sum_{n=1}^{N} p[n] \right\}. \tag{4}$$

Let D_{γ} denote the average delay under policy γ. According to Little's Law, the average queueing delay is the quotient of the average queue length divided by the average arrival rate, i.e.,

$$D_{\gamma} = \frac{1}{\alpha A} \lim_{N \to \infty} \frac{1}{N} \mathrm{E}_{q_0}^{\gamma} \left\{ \sum_{n=1}^{N} q[n] \right\}. \tag{5}$$

Therefore, policy γ will determine $Z_{\gamma} = (P_{\gamma}, D_{\gamma})$, which is a point in the delay-power plane. Define $\overline{Z_{\gamma} Z_{\gamma'}}$ as the line segment connecting Z_{γ} and $Z_{\gamma'}$. Let Γ denote the set of all feasible policies which can guarantee no overflow or underflow. Define $\mathcal{R} = \{Z_{\gamma} | \gamma \in \Gamma\}$ as the set of all feasible points in the delay-power plane. Intuitively, since the power consumption for each data packet increases if we want to transmit faster, there is a tradeoff between the average queueing delay and the average power consumption. Define the optimal delay-power tradeoff curve $\mathcal{L} = \{(P, D) \in \mathcal{R} | \forall (P', D') \in \mathcal{R},$ either $P' \geq P$ or $D' \geq D\}$.

Since there are two costs in the MDP, by minimizing the average delay given an average power constraint P_{th}, we obtain a CMDP problem.

$$\min_{\gamma \in \Gamma} D_{\gamma} \tag{6a}$$
$$\text{s.t. } P_{\gamma} \leq P_{\text{th}}. \tag{6b}$$

By varying the value of P_{th}, the optimal delay-power tradeoff curve $\mathcal{L}$ can be obtained. In the following, we show that optimizing over a simpler class of policies will minimize the objective in (6).

A. Reduction to Stationary Policies

Here, we show that in order to solve our problem, it is enough to restrict our class of policies to a stationary class of policies. A stationary policy for an MDP means that the probability distribution to determine $s[n]$ is only a function of state $q[n]$, i.e. $\delta_n : Q \rightarrow \mathcal{P}(S)$, and the decision rules for all timeslots are the same. For a CMDP, it is proven in [19, Th. 11.3] that stationary policies are complete, which means stationary policies can achieve as good performance as any other policies. Therefore we only need to consider stationary policies in this problem.

Let $f_{q,s}$ denote the probability to transmit s packet(s) when $q[n] = q$, i.e.

$$f_{q,s} = \Pr\{s[n] = s | q[n] = q\}. \tag{7}$$

Therefore we have $\sum_{s=0}^{S} f_{q,s} = 1$ for all $q = 0, \cdots, Q$. We guarantee the avoidance of overflow or underflow by setting $f_{q,s} = 0$ if $q - s < 0$ or $q - s > Q - A$. Let F denote a $(Q+1) \times (S+1)$ matrix whose element in the $(q+1)$th row and the $(s+1)$th column is $f_{q,s}$. Therefore matrix F can represent a decision rule, and moreover a stationary transmission policy. As explained above, we only need to consider stationary policies. Therefore, we will use F to represent a stationary policy in the rest of the paper. Let P_F and D_F denote the average power consumption and the average queueing delay under policy F. Let $\mathcal{F}$ denote the set of all feasible stationary policies which can guarantee no overflow or underflow. Let $\mathcal{F}_D$ denote the set of all stationary and deterministic policies which can guarantee no overflow or underflow. Thus the optimization problem (6) is equivalent to

$$\min_{F \in \mathcal{F}} D_F \tag{8a}$$
$$\text{s.t. } P_F \leq P_{\text{th}}. \tag{8b}$$

B. Reduction to Unichains

Given a stationary policy for a Markov Decision Process, there is an inherent Markov Reward Process (MRP) with $q[n]$ as the state variable. Let $\lambda_{i,j}$ denote the transition probability from state i to state j. An example of the transition diagram is shown in Fig. 3, where $\lambda_{i,i}$ for $i = 0, \cdots, Q$ are omitted to keep the diagram legible.

The Markov chain could have more than one closed communication class under certain transmission policies. For example, in the example in Fig. 3, if we apply the scheduling policy $f_{0,0} = 1$, $f_{1,0} = 1$, $f_{2,2} = 1$, $f_{3,2} = 1$, $f_{4,2} = 1$, $f_{5,2} = 1$,

$f_{6,2} = 1$, $f_{7,2} = 1$, and $f_{i,j} = 0$ for all others, it can be seen that states 4, 5, 6 and 7 are transient, while states $\{0, 2\}$ and states $\{1, 3\}$ are two closed communication classes. Under this circumstance, the limiting probability distribution and the average cost are dependent on the initial state and the sample paths. However, the following theorem will show that we only need to study the cases with only one closed communication class.

Theorem 1: If the Markov chain generated by policy $\boldsymbol{F}$ has more than one closed communication class, named as $\mathcal{C}_1, \cdots, \mathcal{C}_L$, where $L > 1$, then for all $1 \leq l \leq L$, there exists a policy $\boldsymbol{F}_l$ such that the Markov chain generated by $\boldsymbol{F}_l$ has $\mathcal{C}_l$ as its only closed communication class. Moreover, the limiting distribution and the average cost of the Markov chain generated by $\boldsymbol{F}$ starting from state $c \in \mathcal{C}_l$ are the same as the limiting distribution and the average cost of the Markov chain generated by $\boldsymbol{F}_l$.

Proof: See Appendix A. ∎

According to Theorem 1, the constructed policy $\boldsymbol{F}_l$ and the original policy $\boldsymbol{F}$ will have the same average power and average delay for any closed communication class $\mathcal{C}_l$ which policy $\boldsymbol{F}$ converges to. Therefore, without loss of generality, we can focus on the Markov chains with only one closed communication class, which are called unichains. For a unichain, the initial state or the sample path won't affect the limiting distribution or the average cost, which means the parameter q_0 in $\mathrm{E}_{q_0}^{\gamma}\{\cdot\}$ won't affect the value of the function.

As we will demonstrate in the following two sections, the optimal policies for the Constrained MDP problem and its Lagrangian relaxation problem are threshold-based. Here, we define that, a stationary policy $\boldsymbol{F}$ is threshold-based, if and only if there exist $(S + 1)$ thresholds $0 \leq q_F(0) \leq q_F(1) \leq \cdots \leq q_F(S) \leq Q$, such that $f_{q,s} > 0$ only when $q_F(s - 1) \leq q \leq q_F(s)$ (we set $q_F(-1) = -1$ for the inequality when $s = 0$). It means that, under policy $\boldsymbol{F}$, when the queue state is larger than threshold $q_F(s - 1)$ and smaller than $q_F(s)$, it transmits s packet(s). When the queue state is equal to threshold $q_F(s)$, it transmits s or $(s + 1)$ packet(s). Note that under this definition, probabilistic policies can also be threshold-based.

III. Optimal Deterministic Threshold-Based Policy for the Lagrangian Relaxation Problem

In (8), we formulate the optimization problem as a Constrained MDP, which is difficult to solve in general. Therefore, we first study the Lagrangian relaxation of (8) in this section, and prove that the optimal policy for the relaxation problem is deterministic and threshold-based. We will then use these results to show that the solution to the original non-relaxed CMDP problem is also of a threshold type.

Let $\mu \geq 0$ denote the Lagrange multiplier. Thus the Lagrangian relaxation of (8) is

$$\min_{\boldsymbol{F} \in \mathcal{F}} \ \lim_{N \to \infty} \frac{1}{N} \mathrm{E}_{q_0}^{\boldsymbol{F}} \left\{ \frac{1}{\alpha A} \sum_{n=1}^{N} (q[n] + \alpha A \mu p[n]) \right\} - \mu P_{\text{th}}. \quad (9)$$

In (9), the term $-\mu P_{\text{th}}$ is constant. Therefore, the Lagrangian relaxation problem is minimizing a constructed combined average cost $(q[n] + \alpha A \mu p[n])$. This is an infinite-horizon Markov Decision Process with an average cost criterion, for which it is proven in [20, Th. 9.1.8] that, there exists an optimal stationary deterministic policy. For a stationary deterministic policy $\boldsymbol{F} \in \mathcal{F}_D$, let $s_F(q)$ denote the packet(s) to transmit when $q[n] = q$. In other words, we have $f_{q,s_F(q)} = 1$ for all q. Define $\eta = \alpha A \mu$. Therefore (9) is equivalent to

$$\min_{\boldsymbol{F} \in \mathcal{F}_D} \ \lim_{N \to \infty} \frac{1}{N} \mathrm{E}_{q_0}^{\boldsymbol{F}} \left\{ \sum_{n=1}^{N} \left(q[n] + \eta \Theta_{s_F(q[n])} \right) \right\}. \quad (10)$$

The optimal policy for (10) has the following property.

Theorem 2: An optimal policy $\boldsymbol{F}$ for (10) should satisfy that $s_F(q + 1) - s_F(q) = 0$ or 1 for all $0 \leq q < Q$. Therefore $\boldsymbol{F}$ is a threshold-based policy.

Proof: See Appendix B. ∎

The proof applies a nested induction method to the policy iteration algorithm for the Markov Decision Process. Theorem 2 indicates a very intuitive conclusion that more data should be transmitted if the queue is longer. More specifically speaking, for an optimal deterministic policy $\boldsymbol{F}$, there exists $(S + 1)$ thresholds $q_F(0) \leq q_F(1) \leq \cdots \leq q_F(S)$, such that

$$\begin{cases} f_{q,s} = 1 & q_F(s - 1) < q \leq q_F(s), s = 0, \cdots, S \\ f_{q,s} = 0 & \text{else} \end{cases} \quad (11)$$

where $q_F(-1) = -1$. The form of the optimal policy satisfies our definition of threshold-based policy in Section II.

Moreover, we can have the following two corollaries.

Corollary 1: Under any optimal threshold-based policy $\boldsymbol{F}$, there will be no transmission only when $q[n] = 0$. In other words, threshold $q_F(0) = 0$.

Proof: This is an intuitive result, because every data packet will be transmitted sooner or later, which costs at least Θ_1 power, thus not transmitting when there are backlogs is just a waste of time. The following is its rigorous proof.

If there exists an optimal threshold-based policy $\boldsymbol{F} \in \mathcal{F}_D$ where $s_F(q_1) = 0$, $q_1 > 0$. Since $s_F(q)$ has the threshold-based property, we have $s_F(1) = 0$. Construct a policy $\boldsymbol{F}'$ where $s_{F'}(q) = s_F(q + 1)$ for $0 \leq q < Q$ and $s_{F'}(Q) = A$. It can be seen that $\boldsymbol{F}' \in \mathcal{F}_D$. State 0 is a transient state under policy $\boldsymbol{F}$, and state Q is a transient state under policy $\boldsymbol{F}'$. States $1, \cdots, Q$ under policy $\boldsymbol{F}$ and states $0, \cdots, Q-1$ under policy $\boldsymbol{F}'$ have the exactly same state transition, except that the states for $\boldsymbol{F}'$ are 1 smaller than the states for $\boldsymbol{F}$. Therefore the average power consumption under two policies is the same and the average queue length for $\boldsymbol{F}'$ is 1 smaller, which means the average delay for $\boldsymbol{F}'$ is strictly smaller. Therefore $\boldsymbol{F}$ is not an optimal policy, which conflicts with the assumption. Hence the optimal threshold-based policy should have that there will be no transmissions only when $q[n] = 0$. ∎

Corollary 2: For an optimal threshold-based policy $\boldsymbol{F}$, there is no need to transmit more than A packets. In other words, threshold $q_F(A) = q_F(A + 1) = \cdots = q_F(S) = Q$.

Proof: If there exists an optimal threshold-based policy $\boldsymbol{F} \in \mathcal{F}_D$ where q_1 is the smallest state such that $s_F(q_1) > A$. Since $s_F(q)$ has the threshold-based property, for all $q \geq q_1$,

we have $s(q) > A$. Also, for all $q < q_1$, we have $s(q) \leq A$. Construct a policy $\boldsymbol{F}'$ where $s_{F'}(q) = s_F(q)$ for $q < q_1$ and $s_{F'}(q) = A$ for $q \geq q_1$. It can be seen that $\boldsymbol{F}' \in \mathcal{F}_D$. Since $q_1, \cdots, Q$ are transient states under both policies, and the transmission for recurrent states of the two policies is exactly the same, policy $\boldsymbol{F}'$ has the same performance as policy $\boldsymbol{F}$. Therefore, for an optimal threshold-based policy, there is no need to transmit more than A packets. ∎

IV. OPTIMAL THRESHOLD-BASED POLICY FOR THE CMDP

In Section III, we prove that the optimal policy to minimize the combined cost is deterministic and threshold-based. We will now prove that the solution to the overall CMDP problem also takes on a threshold form. We first conduct steady-state analysis for the Markov Decision Process, discover that the feasible average delay and power region is a convex polygon and the optimal delay-power tradeoff curve is piecewise linear, whose neighboring vertices are obtained by deterministic policies which take different actions in only one state. Based on this, the optimal threshold-based policy obtained in Section III will be shown to correspond to the vertices of the piecewise linear curve. Therefore, the optimal policy for the CMDP problem, which is the convex combination of two deterministic threshold-based policies, will be proven to also take a threshold form. Then, we will provide an efficient algorithm to obtain the optimal delay-power tradeoff curve, and a Linear Programming will be formulated to confirm our results.

Based on Theorem 1, without loss of generality, we can focus on unichains, in which case the steady-state probability distribution exists. Let $\pi_F(q)$ denote the steady-state probability for state q when applying policy $\boldsymbol{F}$. Define $\boldsymbol{\pi}_F = [\pi_F(0), \cdots, \pi_F(Q)]^T$. Let $\boldsymbol{\Lambda}_F$ denote a $(Q+1) \times (Q+1)$ matrix whose element in the $(i+1)$th column and the $(j+1)$th row is $\lambda_{i,j}$, which is determined by policy $\boldsymbol{F}$. Let $\boldsymbol{I}$ denote the identity matrix. Define $\boldsymbol{1} = [1, \cdots, 1]^T$, and $\boldsymbol{0} = [0, \cdots, 0]^T$. We won't specify the size of $\boldsymbol{I}$, $\boldsymbol{1}$ or $\boldsymbol{0}$ if there is no ambiguity.

Define $\boldsymbol{G}_F = \boldsymbol{\Lambda}_F - \boldsymbol{I}$. Define $\boldsymbol{H}_F = \begin{bmatrix} \boldsymbol{1}^T \\ \boldsymbol{G}_F(0:(Q-1),:) \end{bmatrix}$ and $\boldsymbol{c} = \begin{bmatrix} 1 \\ \boldsymbol{0} \end{bmatrix}$.

From the definition of the steady-state distribution, we have $\boldsymbol{G}_F \boldsymbol{\pi}_F = \boldsymbol{0}$ and $\boldsymbol{1}^T \boldsymbol{\pi}_F = 1$. For a unichain, the rank of $\boldsymbol{G}_F$ is Q. Therefore, we can conclude that $\boldsymbol{H}_F$ is invertible and

$$\boldsymbol{\pi}_F = \boldsymbol{H}_F^{-1} \boldsymbol{c}. \tag{12}$$

We can express the average power consumption P_F and the average delay D_F using the steady-state probability distribution. For state q, transmitting s packet(s) will cost Θ_s with probability $f_{q,s}$. Define $\boldsymbol{p}_F = [\sum_{s=0}^{S} \Theta_s f_{0,s}, \cdots, \sum_{s=0}^{S} \Theta_s f_{Q,s}]^T$, which is a function of $\boldsymbol{F}$, thus the average power consumption

$$P_F = \sum_{q=0}^{Q} \pi_F(q) \sum_{s=0}^{S} \Theta_s f_{q,s} = \boldsymbol{p}_F^T \boldsymbol{\pi}_F. \tag{13}$$

Similarly, define $\boldsymbol{d} = [0, 1, \cdots, Q]^T$, thus the average delay

$$D_F = \frac{1}{\alpha A} \sum_{q=0}^{Q} q \pi_F(q) = \frac{1}{\alpha A} \boldsymbol{d}^T \boldsymbol{\pi}_F. \tag{14}$$

A. Partially Linear Property of Scheduling Policies

The mapping from $\boldsymbol{F}$ to $Z_F = (P_F, D_F)$ has a partially linear property shown in the following lemma.

Lemma 1: $\boldsymbol{F}$ and $\boldsymbol{F}'$ are two scheduling policies that are different only when $q[n] = q$, i.e. the two matrices are different only in the $(q+1)$th row. Define $\boldsymbol{F}'' = (1-\epsilon)\boldsymbol{F} + \epsilon \boldsymbol{F}'$ where $0 \leq \epsilon \leq 1$. Then
1) There exists $0 \leq \epsilon' \leq 1$ so that $P_{F''} = (1-\epsilon')P_F + \epsilon' P_{F'}$ and $D_{F''} = (1-\epsilon')D_F + \epsilon' D_{F'}$. Moreover, it should hold that ϵ' is a continuous nondecreasing function of ϵ.
2) When ϵ changes from 0 to 1, point $Z_{F''}$ moves on the line segment $\overline{Z_F Z_{F'}}$ from Z_F to $Z_{F'}$.
    Proof: See Appendix C. ∎

Lemma 1 indicates that the convex combination of scheduling policies which take different actions in only one state will induce the convex combination of points in the delay-power plane. Furthermore, we can have the following two results.

Theorem 3: The set of all feasible points in the delay-power plane, $\mathcal{R}$, is a convex polygon whose vertices are all obtained by deterministic scheduling policies. Moreover, the policies corresponding to neighboring vertices of $\mathcal{R}$ take different actions in only one state.
    Proof: See Appendix D. ∎

Corollary 3: The optimal delay-power tradeoff curve $\mathcal{L}$ is piecewise linear, decreasing, and convex. The vertices of the curve are obtained by deterministic scheduling policies. Moreover, the policies corresponding to neighboring vertices of $\mathcal{L}$ take different actions in only one state.
    Proof: See Appendix E. ∎

B. Optimal Threshold-Based Policy for the CMDP

In the last section, we prove in Theorem 2 that the optimal policy for the combined cost is deterministic and threshold-based. Based on the steady-state analysis, the objective function in the unconstrained MDP problem (10)

$$\lim_{N \to \infty} \frac{1}{N} \mathrm{E}_{q_0}^F \left\{ \sum_{n=1}^{N} \left(q[n] + \eta \Theta_{s_F(q[n])} \right) \right\}$$
$$= \alpha A D_F + \eta P_F = \langle (\eta, \alpha A), Z_F \rangle \tag{15}$$

can be seen as the inner product of vector $(\eta, \alpha A)$ and Z_F. Since $\mathcal{R}$ is a convex polygon, the corresponding Z_F minimizing the inner product will be obtained by vertices of $\mathcal{L}$, as demonstrated in Fig. 4. Since the conclusion in Theorem 2 holds for any η, the vertices of the optimal tradeoff curve are all obtained by optimal policies for the relaxed problem, which are deterministic and threshold-based. Moreover, from Corollary 3, the neighboring vertices of $\mathcal{L}$ are obtained by policies which take different actions in only one state. Therefore, we have the following theorem.

Theorem 4: Given an average power constraint, the scheduling policy $\boldsymbol{F}$ to minimize the average delay takes the following form that, there exists $(S+1)$ thresholds
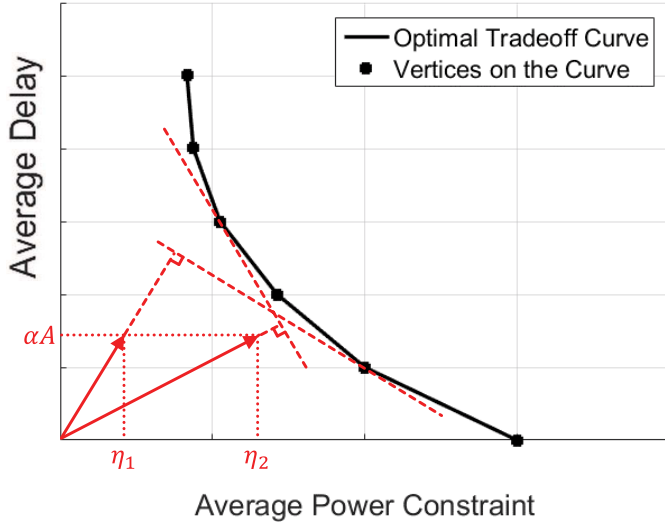
Fig. 4.   The minimum inner product of points on $\mathcal{L}$ and the weighted vector can always be obtained by vertices of $\mathcal{L}$.

$q_F(0) \leq q_F(1) \leq \cdots \leq q_F(S)$, one of which we label as $q_F(s^*)$, such that

$$
\begin{cases}
f_{q,s} = 1 & q_F(s-1) < q \leq q_F(s), \ s \neq s^* \\
f_{q,s^*} = 1 & q_F(s^*-1) < q < q_F(s^*) \\
f_{q_F(s^*),s^*} + f_{q_F(s^*),\ s^*+1} = 1 & \\
f_{q,s} = 0 & \text{else}
\end{cases}
\tag{16}
$$

where $q_F(-1) = -1$. Therefore F is a threshold-based policy.

Proof: According to Corollary 3, the policies corresponding to neighboring vertices of $\mathcal{L}$ are deterministic and take different actions in only one state. In other words, according to (11), the thresholds for F and F' are all the same except one of their thresholds are different by 1. Define the thresholds for F as $q_F(0), q_F(1), \cdots, q_F(s^*), \cdots, q_F(S)$, and the thresholds for F' as $q_F(0), q_F(1), \cdots, q_F(s^*) - 1, \cdots, q_F(S)$, which means the two policies are different only in state $q_F(s^*)$. Since the policy to obtain a point on $\overline{Z_F Z_{F'}}$ is the convex combination of F and F', it should have the form shown in (16). This form satisfies our definition of threshold-based policies in Section II.   ∎

According to Theorem 4, policies corresponding to the points between vertices of the optimal tradeoff curve, as the mixture of two deterministic threshold-based policies different only in one state, also satisfies our definition of threshold-based policy in Section II. When $q_F(s - 1) < q[n] < q_F(s)$, we transmit s packet(s). Any optimal scheduling policy F has at most two decimal elements $f_{q_F(s^*),s^*}$ and $f_{q_F(s^*),s^*+1}$, while the other elements are either 0 or 1.

C. Algorithm to Obtain the Optimal Tradeoff Curve

Here, we propose Algorithm 1 to efficiently obtain the optimal delay-power tradeoff curve. This algorithm is based on the results that the optimal delay-power tradeoff curve is piecewise linear, whose vertices are obtained by deterministic threshold-based policies, and policies corresponding to two

---

Algorithm 1 Constructing the Optimal Delay-Power Tradeoff

1: Construct F whose thresholds $q_F(s) = s$ for $s < A$ and $q_F(s) = Q$ for $s \geq A$
2: Calculate D_F and P_F
3: $\mathcal{F}_c \leftarrow [F]$, $D_c \leftarrow D_F$, $P_c \leftarrow P_F$
4: **while** $\mathcal{F}_c \neq \emptyset$ **do**
5:     $\mathcal{F}_p \leftarrow \mathcal{F}_c$, $D_p \leftarrow D_c$, $P_p \leftarrow P_c$
6:     $\mathcal{F}_c \leftarrow \emptyset$, $slope \leftarrow +\infty$
7:     **while** $\mathcal{F}_p \neq \emptyset$ **do**
8:         $F = \mathcal{F}_p.\text{pop}(0)$
9:         **for all** $0 < s^* \leq A$ **do**
10:            Construct F' where $q_{F'}(s^*) = q_F(s^*) + 1$ and $q_{F'}(s) = q_F(s)$ for $s \neq s^*$
11:            NewPolicyProbing()
12:            // Probing all possible candidates
13:        **end for**
14:    **end while**
15:    Draw the line segment connecting (P_p, D_p) and (P_c, D_c)
16: **end while**

1: **procedure** NEWPOLICYPROBING( )
2:     **if** F' is feasible and threshold-based **then**
3:         Calculate $D_{F'}$ and $P_{F'}$
4:         **if** $D_{F'} = D_p$ and $P_{F'} = P_p$ **then**
5:             $\mathcal{F}_p.\text{append}(F')$
6:             // $Z_{F'}$ coincides with Z_p
7:         **else if** $D_{F'} \geq D_p$ and $P_{F'} < P_p$ **then**
8:             **if** $\frac{D_{F'} - D_p}{P_p - P_{F'}} < slope$ **then**
9:                 $\mathcal{F}_c \leftarrow [F']$, $slope \leftarrow \frac{D_{F'} - D_p}{P_p - P_{F'}}$
10:                $D_c \leftarrow D_{F'}$, $P_c \leftarrow P_{F'}$
11:                // $Z_{F'}$ has the best performance
12:            **else if** $\frac{D_{F'} - D_p}{P_p - P_{F'}} = slope$ **then**
13:                **if** $P_{F'} = P_c$ **then**
14:                    $\mathcal{F}_c.\text{append}(F')$
15:                    // $Z_{F'}$ has the same performance as the current best candidate(s)
16:                **else if** $P_{F'} > P_c$ **then**
17:                    $\mathcal{F}_c \leftarrow [F']$, $slope \leftarrow \frac{D_{F'} - D_p}{P_p - P_{F'}}$
18:                    $D_c \leftarrow D_{F'}$, $P_c \leftarrow P_{F'}$
                     // $Z_{F'}$ has the same slope as the current best candidate(s)
19:                    but is closer to Z_p
20:                **end if**
21:            **end if**
22:        **end if**
23:    **end if**
24: **end procedure**

adjacent vertices take different actions in only one state. With the optimal tradeoff curve obtained, the minimum delay given a specific power constraint can also be obtained.

The basic idea of the algorithm is as follows. We start from the bottom-right vertex of the optimal tradeoff curve, which corresponds to the highest power consumption and the smallest average delay. Therefore the corresponding policy is to transmit as much as possible in any state. According to Corollary 2, there is no need to transmit more than A packets. Thus the policy is $s_F(q) = q$ when $q < A$ and $s_F(q) = A$ when $q \geq A$. The thresholds for this policy are $q_F(s) = s$ when $s < A$ and $q_F(s) = Q$ when $s \geq A$. Then we enumerate all the candidates for the next vertex of the curve, which are different in only one threshold from the current policy, based on the conclusion that policies corresponding to adjacent

vertices take different actions in only one state. The next vertex will be determined by the policy candidate whose connecting line with the current vertex has the minimum absolute slope and the minimum length. Note that a vertex can be obtained by more than one policy, therefore we use lists $\mathcal{F}_p$ and $\mathcal{F}_c$ to restore all policies corresponding to the previous and the current vertices. When conducting the complexity analysis, we assume the situation where a vertex is obtained by multiple policies rarely happens. In total, there are $(A-1)$ alterable thresholds, and the maximum varying range of each threshold is Q. Every time we get to the next vertex, one of the thresholds of the policy will be increased by 1. Therefore, the maximum iteration number for new vertices is $\boldsymbol{O}(AQ)$. Within each iteration, there are A candidates we need to test to determine next vertex, hence the computation times is $\boldsymbol{O}(A^2Q)$. For each candidate, the most time-consuming operation, the matrix inversion, costs $\boldsymbol{O}(Q^3)$. Therefore the complexity of the algorithm is $\boldsymbol{O}(A^2Q^4)$. In the next section, we will see that this algorithm is much more efficient than using the general LP method to solve the problem.

D. Linear Programming Formulation

In the following, we demonstrate that the CMDP problem can also be formulated as a Linear Programming, which can be solved for a certain power constraint. We compare Algorithm 1 and Linear Programming, and demonstrate that our algorithm is superior to the Linear Programming based approach. In the next section, we will use Linear Programming to confirm the properties of the optimal tradeoff curve and the algorithm we have demonstrated.

Based on the steady-state analysis (12) (13) and (14), the optimization problem (8) can be transformed into

$$\min_{\boldsymbol{F},\boldsymbol{\pi}} \frac{1}{\alpha A} \boldsymbol{d}^T \boldsymbol{\pi}_F \tag{17a}$$

$$\text{s.t. } \boldsymbol{p}_F^T \boldsymbol{\pi}_F \leq P_{\text{th}} \tag{17b}$$

$$\boldsymbol{H}_F \boldsymbol{\pi}_F = \boldsymbol{c} \tag{17c}$$

$$\boldsymbol{\pi}_F \succeq \boldsymbol{0} \tag{17d}$$

$$f_{q,s} = 0 \quad \forall q - s < 0 \text{ or } q - s > Q - A \tag{17e}$$

where $\boldsymbol{\pi}_F \succeq \boldsymbol{0}$ means $\boldsymbol{\pi}$ is componentwise nonnegative.

Define $x_{q,s} = \pi(q)f_{q,s}$. By substituting the variables in (17), the optimization problem can be transformed into

$$\min \frac{1}{\alpha A} \sum_{q=0}^{Q} q \sum_{s=0}^{S} x_{q,s} \tag{18a}$$

$$\text{s.t. } \sum_{q=0}^{Q} \sum_{s=0}^{S} \Theta_s x_{q,s} \leq P_{\text{th}} \tag{18b}$$

$$\sum_{l=\max\{0,q-A\}}^{q-1} \sum_{s=0}^{l+A-q} \alpha x_{l,s}$$

$$= \sum_{r=q}^{\min\{q+S-1,Q\}} \sum_{s=r-q+A+1}^{S} x_{r,s}$$

$$+ \sum_{r=q}^{\min\{q+S-1,Q\}} \sum_{s=r-q+1}^{r-q+A} (1-\alpha)x_{r,s} \quad q = 1, \cdots, Q \tag{18c}$$

$$\sum_{q=0}^{Q} \sum_{s=0}^{S} x_{q,s} = 1 \tag{18d}$$

$$x_{q,s} = 0 \quad \forall q - s < 0 \text{ or } q - s > Q - A \tag{18e}$$

$$x_{q,s} \geq 0 \quad \forall 0 \leq q - s \leq Q - A. \tag{18f}$$

It can be observed that this is a Linear Programming problem. Given a feasible solution to (17), $\boldsymbol{F}$ and $\boldsymbol{\pi}$, it can be checked that $x_{q,s} = \pi(q)f_{q,s}$ for all q and s is a feasible solution to (18) with the same objective value. On the other hand, given a feasible solution to (18), $x_{q,s}$ for all q and s, it can be proven that $\pi(q) = \sum_{s=0}^{S} x_{q,s}$ and

$$f_{q,s} = \begin{cases} \dfrac{x_{q,s}}{\pi(q)} & \pi(q) > 0 \\ 1 & \pi(q) = 0, s = \min\{q, S\} \\ 0 & \pi(q) = 0, s \neq \min\{q, S\} \end{cases}$$

is a feasible solution to (17) with the same objective value. This means Linear Programming (18) is equivalent with (17), thus also equivalent with (8).

In (18), there are $(Q+1)(S+1)$ variables in total. From the conclusion in Corollary 2, the number can be reduced to $(A+1)(Q+1)$. If we apply the ellipsoid algorithm to solve the Linear Programming (18), the computational complexity is $\boldsymbol{O}(A^4Q^4)$. It means that the complexity of applying Linear Programming to obtain only one point on the optimal tradeoff curve is larger than that of obtaining the entire curve with Algorithm 1. Moreover, when the average energy constraint is dynamically changing, Linear Programming needs to be solved for each constraint, while the constructed delay-power tradeoff from Algorithm 1 can adapt to the changed constraint instantly. This demonstrates the inherent advantage in using the revealed properties of the optimal tradeoff curve and the optimal policies.

V. NUMERICAL RESULTS

In this section, we validate our theoretical results by conducting numerical computation and simulations. The convex feasible delay-power region and the generated polygons will be demonstrated in a small-scale example. The delay-power tradeoff curves will be obtained in a practical scenario. It will be confirmed that Algorithm 1 can obtain the optimal delay-power tradeoff curve for both cases.

In Fig. 5, we plot all the delay-power points generated by deterministic policies, and connect the points whose corresponding policies take different actions in only one state. By conducting this operation, any generated polygon, a concept introduced in Appendix D, is contained in the figure. For any two deterministic policies, there is a convex polygon generated by them. Any point inside a generated polygon can be obtained by a policy. As we can see, the feasible delay-power region is made up of all the generated polygons. According to our proof in Appendix D, the feasible region is covered by basic polygons. The parameters for this figure are $Q = 6$, $A = 3$, $M = 3$, $\alpha = 0.4$, $\Theta_0 = 0$, $\Theta_1 = 1$, $\Theta_2 = 4$, $\Theta_3 = 9$. As proven in Theorem 3, the feasible delay-power region is the convex hull of all the points obtained by deterministic policies. The optimal delay-power tradeoff curve is obtained by Algorithm 1. Therefore the vertices of the curve
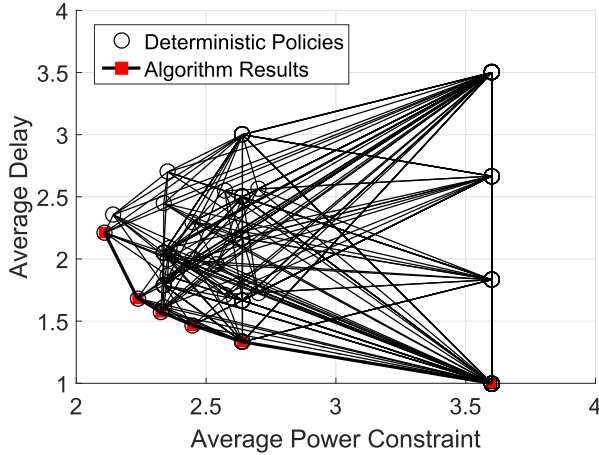
Fig. 5.   Points Corresponding to Deterministic Policies and Generated Polygons in the Delay-Power Plane.
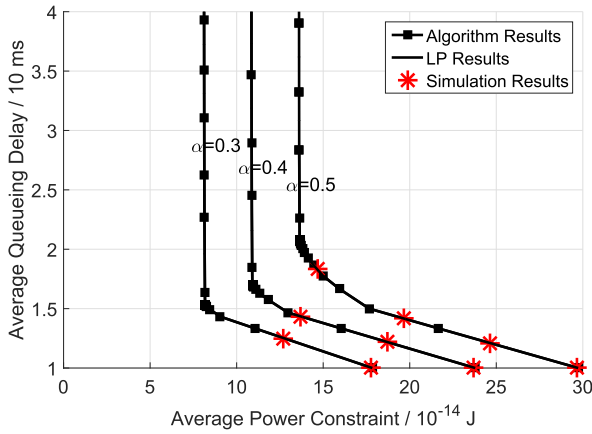

Fig. 6.   Optimal Delay-Power Tradeoff Curves.

are all corresponding to threshold-based deterministic policies, and neighboring vertices are obtained by policies different in only one state. Policies to obtain the points between vertices of the optimal tradeoff curve are the mixture of two deterministic policies.

The optimal delay-power tradeoff curves are demonstrated in Fig. 6, which are validated by Linear Programming and simulations. We consider a more practical scenario with adaptive M-PSK transmissions. The optional modulations are BPSK, QPSK, and 8-PSK. Assume the bandwidth = 1 MHz, the length of a timeslot = 10 ms, and the target bit error rate ber=10^{-5}. Set a data packet contains 10,000 bits. Then by adaptively applying BPSK, QPSK, or 8-PSK, we can respectively transmit 1, 2, or 3 packets in a timeslot, which means S=3. Assume the one-sided noise power spectral density N_0=−150 dBm/Hz. Then the transmission power for different transmission rates can be calculated as Θ_0=0 J, Θ_1=9.0∗10^{-14} J, Θ_2=18.2∗10^{-14} J, and Θ_3=59.5∗10^{-14} J. Assume in a timeslot, data arrive as a Bernoulli process. Each data arrival contains A=3 packets. Set the buffer size Q=100. The optimal delay-power tradeoff curves are shown in Fig. 6, with α = 0.3, 0.4, and 0.5 respectively. It is demonstrated that the optimal delay-power tradeoff curves obtained by Linear Programming completely overlap the optimal tradeoff curves generated by Algorithm 1. The results are further validated by

simulations, which are shown in "∗" markers. As proven in Corollary 3, the optimal tradeoff curves are piecewise linear, decreasing, and convex. The vertices of the curves are marked by squares. The corresponding optimal policies can be checked as threshold-based. With α increasing, the curve gets higher because of the heavier workload. The minimum average delay is 1 for all curves, because when we transmit as much as we can, all data packets will stay in the queue for exactly one timeslot. The curve gets very steep when the power constraint decreases. This is because, when the power constraint gets tighter, we will mainly transmit with BPSK and QPSK. Since $\Theta_1 \approx \frac{\Theta_2}{2}$, different policies will have similar average power consumption.

VI. Conclusion

In this paper, we analyzed the optimal tradeoff between the average delay and the average power consumption in a communication system. The transmission for each timeslot is scheduled according to the buffer state, considering an average power constraint. This problem is formulated as a CMDP. We first study the Lagrangian relaxation of the CMDP problem, and prove that it has deterministic threshold-based optimal policies. Then we show that the feasible delay-power region is a convex polygon, and the optimal delay-power tradeoff curve is piecewise linear, whose vertices are obtained by the optimal solution to the relaxation problem, and the neighboring vertices of which are obtained by policies taking different actions in only one state. Based on these results, the optimal policies for the original problem are proven to be threshold-based. We also propose an algorithm to efficiently obtain the optimal delay-power tradeoff. The theoretical results and the proposed algorithm are validated by Linear Programming and simulations. In the future, we will study a more generalized model combining adaptive transmission and fading channels.

Appendix A
Proof of Theorem 1

Define the set of transient states which have access to C_l as C_l^t. Define the set of transient states which don't have access to C_l as C_{nl}^t. Therefore $\{C_1, \cdots, C_L, C_l^t, C_{nl}^t\}$ is a partition of all the states. It is straightforward that there should be at least one state $c \in \bigcup_{i=1, i \neq l}^{+\infty} C_i \cup C_{nl}^t$ which is adjacent to a state $c' \in C_l \cup C_l^t$, which means $|c - c'| = 1$. If $c - c' = 1$, we transmit 1 packet in state c; if $c' - c = 1$, we transmit $(A-1)$ packet(s) in state c. Therefore we can always modify the transmission policy for state c so that state c can access $C_l \cup C_l^t$. Then c will be a transient state which has access to C_l, and so are the states which communicate with c.

Renew the partition according to the modification of the transmission policy. According to the above operation, the set C_l won't change, but C_l^t will be strictly increasing. Therefore, by repeating the same operation finite times, all the states will be partitioned in C_l and C_l^t. Therefore C_l is its only closed communication class and the corresponding transmission policy is the F_l we request.

Since F and F_l still have the same policy for the states in C_l, the limiting distribution and the average cost of the Markov chain generated by F starting from state $c \in C_l$ are the same

Algorithm 2 Policy Iteration Algorithm for Markov Decision Processes

---
1: $m \leftarrow 0$
2: **for all** q **do**
3:      $h^{(0)}(q) \leftarrow$ arbitrary value // Initialization
4: **end for**
5: **repeat**
6:      **for all** q **do**
7:          // Policy Improvement
8:          $s^{(m+1)}(q) \leftarrow \arg\min_s \{h^{(m+1)}(q, s)\}$
9:      **end for**
10:      **for all** q **do**
11:          // Policy Evaluation
12:          $h^{(m+1)}(q) \leftarrow h^{(m+1)}(q, s^{(m+1)}(q))$
13:      **end for**
14:      $m \leftarrow m + 1$
15: **until** $s^{(m)}(q) = s^{(m-1)}(q)$ holds for all q
16: $s(q) \leftarrow s^{(m)}(q)$ for all q

---

as the limiting distribution and the average cost of the Markov chain generated by $\boldsymbol{F}_l$.

APPENDIX B
PROOF OF THEOREM 2

For the simplicity of notations, in the proof we use $s(q)$ instead of $s_F(q)$. We call that $s(q)$ satisfies the threshold-based property, if $s_F(q + 1) - s_F(q) = 0$ or 1 for all $0 \leq q < Q$. For a deterministic policy $s(q)$ which satisfies the threshold-based property, it can be seen that this policy also satisfies our definition of a threshold-based policy. Define

$$
\begin{aligned}
h^{(m+1)}&(q, s) \\
&= q + \eta \Theta_s + \alpha[h^{(m)}(q - s + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}(q - s) - h^{(m)}(0)].
\end{aligned} \tag{19}
$$

We will prove the theorem by applying a nested induction method to policy iteration algorithm for the Markov Decision Process. In Markov Decision Processes with an average cost, policy iteration algorithm can be applied to obtain the optimal scheduling policy, which is shown in Algorithm 2. In the algorithm, the function $h^{(m)}(q)$ converges to $h(q)$, which is called the potential function or bias function for the Markov Decision Process. The bias function can be interpreted as the expected total difference between the cost starting from a specific state and the stationary cost. The policy iteration algorithm can converge to the optimal solution in finite steps, which is proven in [20, Th. 8.6.6] and [21, Proposition 3.4].

The sketch of the proof is as follows. Because of the mechanism of the policy iteration algorithm, we can assign $h^{(0)}(q)$ as strictly convex in q. In Part I, we will demonstrate by induction that, for any m, if $h^{(m)}(q)$ is strictly convex in q, then $s^{(m+1)}(q)$ has the threshold-based property. In Part II, we will demonstrate that if $s^{(m+1)}(q)$ has the threshold-based property, then $h^{(m+1)}(q)$ is strictly convex in q. Therefore, by mathematical induction, we can prove the theorem.

Part I (Convexity of $h^{(m)}(q)$ in $q \rightarrow$ Threshold-Based Property of $s^{(m+1)}(q)$): Assume $h^{(m)}(q)$ is strictly convex in q.

In this part, we are going to prove $s^{(m+1)}(q)$ has the threshold-based property.

1) Because of the requirements of a feasible policy, we have $s^{(m+1)}(0) = 0$, and $s^{(m+1)}(1) = 0$ or 1. Therefore $s^{(m+1)}(q + 1) - s^{(m+1)}(q) = 0$ or 1 when $q = 0$.

2) We define $s_1 = s^{(m+1)}(q_1)$ for a specific q_1. From the Policy Improvement step in the policy iteration algorithm, we have the following inequalities:

$$
h^{(m+1)}(q_1, s_1) \leq h^{(m+1)}(q_1, s_1 - \delta), \quad \forall 0 \leq \delta \leq s_1, \tag{20}
$$

$$
h^{(m+1)}(q_1, s_1) \leq h^{(m+1)}(q_1, s_1 + \delta), \quad \forall 0 \leq \delta \leq S - s_1. \tag{21}
$$

Since $h^{(m)}(q)$ is strictly convex in q,

$$
\begin{aligned}
h^{(m)}(q_1 &+ 1 - s_1 + A) - h^{(m)}(q_1 - s_1 + A) \\
&< h^{(m)}(q_1 + 1 - (s_1 - \delta) + A) \\
&\quad - h^{(m)}(q_1 - (s_1 - \delta) + A), \tag{22}
\end{aligned}
$$

$$
\begin{aligned}
h^{(m)}(q_1 &+ 1 - s_1) - h^{(m)}(q_1 - s_1) \\
&< h^{(m)}(q_1 + 1 - (s_1 - \delta)) \\
&\quad - h^{(m)}(q_1 - (s_1 - \delta)). \tag{23}
\end{aligned}
$$

Since Θ_s is strictly convex,

$$
\Theta_{s_1+1} - \Theta_{s_1} < \Theta_{s_1+1+\delta} - \Theta_{s_1+\delta}. \tag{24}
$$

By computing $(20) + \alpha \times (22) + (1 - \alpha) \times (23)$, we have

$$
\begin{aligned}
h^{(m+1)}&(q_1 + 1, s_1) \\
&< h^{(m+1)}(q_1 + 1, s_1 - \delta), \quad \forall 0 \leq \delta \leq s_1. \tag{25}
\end{aligned}
$$

By computing (21) and (24), we have

$$
\begin{aligned}
h^{(m+1)}&(q_1 + 1, s_1 + 1) \\
&< h^{(m+1)}(q_1 + 1, s_1 + 1 + \delta), \quad \forall 0 \leq \delta \leq S - s_1 - 1. \tag{26}
\end{aligned}
$$

From (25) and (26), we can see that $s^{(m+1)}(q_1 + 1)$ can only be s_1 or $s_1 + 1$. In other words, we have $s^{(m+1)}(q_1 + 1) - s^{(m+1)}(q_1) = 0$ or 1.

From above, by mathematical induction, we can have that $s^{(m+1)}(q)$ has the threshold-based property.

Part II (Threshold-Based Property of $s^{(m+1)}(q) \rightarrow$ Convexity of $h^{(m+1)}(q)$ in q): Assume $s^{(m+1)}(q)$ has the threshold-based property. We still use the same notation as in the previous part that $s_1 = s^{(m+1)}(q_1)$ for a specific q_1, and $s^{(m+1)}(q_1 + 1) = s_1$ or $s_1 + 1$.

1) If $s^{(m+1)}(q_1 + 1) = s_1$,

$$
\begin{aligned}
h^{(m+1)}&(q_1 + 1) - h^{(m+1)}(q_1) \\
&\leq h^{(m+1)}(q_1 + 1, s_1 + 1) - h^{(m+1)}(q_1, s_1) \tag{27} \\
&= (q_1 + 1) + \eta \Theta_{s_1+1} \\
&\quad + \alpha[h^{(m)}((q_1 + 1) - (s_1 + 1) + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}((q_1 + 1) - (s_1 + 1)) - h^{(m)}(0)] \\
&\quad - [q_1 + \eta \Theta_{s_1} + \alpha[h^{(m)}(q_1 - s_1 + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}(q_1 - s_1) - h^{(m)}(0)]] \tag{28}
\end{aligned}
$$

$$= 1 + \eta(\Theta_{s_1+1} - \Theta_{s_1}). \tag{29}$$

On the other hand,

$$
\begin{aligned}
h^{(m+1)}&(q_1 + 1) - h^{(m+1)}(q_1) \\
&> h^{(m+1)}(q_1 + 1, s_1) - h^{(m+1)}(q_1, s_1 - 1) \tag{30} \\
&= (q_1 + 1) + \eta\Theta_{s_1} \\
&\quad + \alpha[h^{(m)}((q_1 + 1) - s_1 + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}((q_1 + 1) - s_1) - h^{(m)}(0)] \\
&\quad - [q_1 + \eta\Theta_{s_1-1}] \\
&\quad - [\alpha[h^{(m)}(q_1 - (s_1 - 1) + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}(q_1 - (s_1 - 1)) - h^{(m)}(0)]] \tag{31} \\
&= 1 + \eta(\Theta_{s_1} - \Theta_{s_1-1}). \tag{32}
\end{aligned}
$$

2) If $s^{(m+1)}(q_1 + 1) = s_1 + 1$,

$$
\begin{aligned}
h^{(m+1)}&(q_1 + 1) - h^{(m+1)}(q_1) \\
&= (q_1 + 1) + \eta\Theta_{s_1+1} \\
&\quad + \alpha[h^{(m)}((q_1 + 1) - (s_1 + 1) + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}((q_1 + 1) - (s_1 + 1)) - h^{(m)}(0)] \\
&\quad - [q_1 + \eta\Theta_{s_1} + \alpha[h^{(m)}(q_1 - s_1 + A) - h^{(m)}(A)] \\
&\quad + (1 - \alpha)[h^{(m)}(q_1 - s_1) - h^{(m)}(0)]] \tag{33} \\
&= 1 + \eta(\Theta_{s_1+1} - \Theta_{s_1}). \tag{34}
\end{aligned}
$$

To conclude, $1 + \eta(\Theta_{s_1} - \Theta_{s_1-1}) < h^{(m+1)}(q_1 + 1) - h^{(m+1)}(q_1) \le 1 + \eta(\Theta_{s_1+1} - \Theta_{s_1})$ holds for any specific q_1. Therefore $h^{(m+1)}(q + 1) - h^{(m+1)}(q)$ is strictly increasing, which means $h^{(m+1)}(q)$ is strictly convex in q.

Based on the assumption for initial $h^{(0)}(q)$ and the derivations in Part I and II, by mathematical induction, we can prove that $s^{(m)}(q)$ has the threshold-based property for all $m \ge 1$. Since $s^{(m)}(q)$ will converge to the optimal policy $s(q)$ in finite steps, the optimal policy $s(q)$ has the threshold-based property.

APPENDIX C
PROOF OF LEMMA 1

We will prove the two conclusions one by one.

1) From the definition of H_F and p_F, we can see that if $F'' = (1 - \epsilon)F + \epsilon F'$, then $H_{F''} = (1 - \epsilon)H_F + \epsilon H_{F'}$ and $p_{F''} = (1 - \epsilon)p_F + \epsilon p_{F'}$. Define $\Delta H = H_{F'} - H_F$ and $\Delta p = p_{F'} - p_F$. Since F and F' are different only in the $(q+1)$th row, it can be derived that ΔH has nonzero element only in the $(q+1)$th column, and the $(q+1)$th element of Δp is its only nonzero element. Therefore ΔH can be expressed as $[\mathbf{0}, \cdots, \boldsymbol{\delta}_q, \cdots, \mathbf{0}]$, where $\boldsymbol{\delta}_q$ is its $(q+1)$th column. Δp can be defined as $[0, \cdots, \zeta_q, \cdots, 0]^T$, where ζ_q is its $(q+1)$th element. Also, we set

$$
H_F^{-1} = \begin{bmatrix} \boldsymbol{h}_0^T \\ \boldsymbol{h}_1^T \\ \vdots \\ \boldsymbol{h}_Q^T \end{bmatrix}.
$$

Hence

$$
(H_F^{-1}\Delta H)H_F^{-1} = \begin{bmatrix} (\boldsymbol{h}_0^T\boldsymbol{\delta}_q)\boldsymbol{h}_q^T \\ (\boldsymbol{h}_1^T\boldsymbol{\delta}_q)\boldsymbol{h}_q^T \\ \vdots \\ (\boldsymbol{h}_Q^T\boldsymbol{\delta}_q)\boldsymbol{h}_q^T \end{bmatrix}.
$$

By mathematical induction, we can have that for $i \ge 1$,

$$
\begin{aligned}
(H_F^{-1}&\Delta H)^i H_F^{-1} \\
&= \begin{bmatrix} (\boldsymbol{h}_0^T\boldsymbol{\delta}_q)(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T \\ (\boldsymbol{h}_1^T\boldsymbol{\delta}_q)(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T \\ \vdots \\ (\boldsymbol{h}_Q^T\boldsymbol{\delta}_q)(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T \end{bmatrix} \tag{35} \\
&= (\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1}, \tag{36}
\end{aligned}
$$

and $\Delta \boldsymbol{p}^T H_F^{-1}(H_F^{-1}\Delta H)^{i-1} = \zeta_q(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T$.

Therefore the expansion $(H_F + \epsilon\Delta H)^{-1} = H_F^{-1} + \sum_{i=1}^{+\infty}(-\epsilon)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1}$.

From (12), (13) and (14), we have $P_F = \boldsymbol{p}_F^T H_F^{-1}\boldsymbol{c}$ and $D_F = \frac{1}{\alpha A}\boldsymbol{d}^T H_F^{-1}\boldsymbol{c}$. Therefore

$$
\frac{P_{F''} - P_F}{P_{F'} - P_F}
$$

$$
= \frac{(\boldsymbol{p}_F + \Delta \boldsymbol{p})^T(H_F + \epsilon\Delta H)^{-1}\boldsymbol{c} - \boldsymbol{p}_F^T H_F^{-1}\boldsymbol{c}}{(\boldsymbol{p}_F + \Delta \boldsymbol{p})^T(H_F + \Delta H)^{-1}\boldsymbol{c} - \boldsymbol{p}_F^T H_F^{-1}\boldsymbol{c}} \tag{37}
$$

$$
= \frac{\boldsymbol{p}_F^T\left[(H_F + \epsilon\Delta H)^{-1} - H_F^{-1}\right]\boldsymbol{c} + \epsilon\Delta \boldsymbol{p}^T(H_F + \epsilon\Delta H)^{-1}\boldsymbol{c}}{\boldsymbol{p}_F^T\left[(H_F + \Delta H)^{-1} - H_F^{-1}\right]\boldsymbol{c} + \Delta \boldsymbol{p}^T(H_F + \Delta H)^{-1}\boldsymbol{c}} \tag{38}
$$

$$
= \frac{\boldsymbol{p}_F^T\left[\sum_{i=1}^{+\infty}(-\epsilon)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1}\right]\boldsymbol{c} - \Delta \boldsymbol{p}^T\left[\sum_{i=1}^{+\infty}(-\epsilon)^i(H_F^{-1}\Delta H)^{i-1}H_F^{-1}\right]\boldsymbol{c}}{\boldsymbol{p}_F^T\left[\sum_{i=1}^{+\infty}(-1)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1}\right]\boldsymbol{c} - \Delta \boldsymbol{p}^T\left[\sum_{i=1}^{+\infty}(-1)^i(H_F^{-1}\Delta H)^{i-1}H_F^{-1}\right]\boldsymbol{c}} \tag{39}
$$

$$
= \frac{\sum_{i=1}^{+\infty}(-\epsilon)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{p}_F^T(H_F^{-1}\Delta H)H_F^{-1}\boldsymbol{c} - \sum_{i=1}^{+\infty}(-\epsilon)^i\zeta_q(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T\boldsymbol{c}}{\sum_{i=1}^{+\infty}(-1)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{p}_F^T(H_F^{-1}\Delta H)H_F^{-1}\boldsymbol{c} - \sum_{i=1}^{+\infty}(-1)^i\zeta_q(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}\boldsymbol{h}_q^T\boldsymbol{c}} \tag{40}
$$

$$
= \frac{\sum_{i=1}^{+\infty}(-\epsilon)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}}{\sum_{i=1}^{+\infty}(-1)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}} \tag{41}
$$

$$
= \frac{\epsilon + \epsilon\boldsymbol{h}_q^T\boldsymbol{\delta}_q}{1 + \epsilon\boldsymbol{h}_q^T\boldsymbol{\delta}_q} \tag{42}
$$

and

$$
\frac{D_{F''} - D_F}{D_{F'} - D_F}
$$

$$
= \frac{\boldsymbol{d}^T(H_F + \epsilon\Delta H)^{-1}\boldsymbol{c} - \boldsymbol{d}^T H_F^{-1}\boldsymbol{c}}{\boldsymbol{d}^T(H_F + \Delta H)^{-1}\boldsymbol{c} - \boldsymbol{d}^T H_F^{-1}\boldsymbol{c}} \tag{43}
$$

$$
= \frac{\boldsymbol{d}^T(\sum_{i=1}^{+\infty}(-\epsilon)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1})\boldsymbol{c}}{\boldsymbol{d}^T(\sum_{i=1}^{+\infty}(-1)^i(\boldsymbol{h}_q^T\boldsymbol{\delta}_q)^{i-1}(H_F^{-1}\Delta H)H_F^{-1})\boldsymbol{c}} \tag{44}
$$

$$= \frac{\sum_{i=1}^{+\infty}(-\epsilon)^i (\boldsymbol{h}_q^T \boldsymbol{\delta}_q)^{i-1}}{\sum_{i=1}^{+\infty}(-1)^i (\boldsymbol{h}_q^T \boldsymbol{\delta}_q)^{i-1}} \tag{45}$$

$$= \frac{\epsilon + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q}{1 + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q}. \tag{46}$$

Hence $\frac{P_{\boldsymbol{F}''} - P_{\boldsymbol{F}}}{P_{\boldsymbol{F}'} - P_{\boldsymbol{F}}} = \frac{D_{\boldsymbol{F}''} - D_{\boldsymbol{F}}}{D_{\boldsymbol{F}'} - D_{\boldsymbol{F}}} = \frac{\epsilon + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q}{1 + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q} = \epsilon'$, so that $P_{\boldsymbol{F}''} = (1 - \epsilon')P_{\boldsymbol{F}} + \epsilon' P_{\boldsymbol{F}'}$ and $D_{\boldsymbol{F}''} = (1 - \epsilon')D_{\boldsymbol{F}} + \epsilon' D_{\boldsymbol{F}'}$. Moreover, it can be observed that $\epsilon' = \frac{\epsilon + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q}{1 + \epsilon \boldsymbol{h}_q^T \boldsymbol{\delta}_q}$ is a continuous nondecreasing function.

2) From the first part, we know $\frac{P_{\boldsymbol{F}''} - P_{\boldsymbol{F}}}{P_{\boldsymbol{F}'} - P_{\boldsymbol{F}}} = \frac{D_{\boldsymbol{F}''} - D_{\boldsymbol{F}}}{D_{\boldsymbol{F}'} - D_{\boldsymbol{F}}} = \epsilon'$ and ϵ' is a continuous nondecreasing function of ϵ. When $\epsilon = 0$, we have $\epsilon' = 0$. When $\epsilon = 1$, we have $\epsilon' = 1$. Therefore when ϵ changes from 0 to 1, the point $(P_{\boldsymbol{F}''}, D_{\boldsymbol{F}''})$ moves on the line segment from $(P_{\boldsymbol{F}}, D_{\boldsymbol{F}})$ to $(P_{\boldsymbol{F}'}, D_{\boldsymbol{F}'})$. The slope of the line is

$$\frac{D_{\boldsymbol{F}'} - D_{\boldsymbol{F}}}{P_{\boldsymbol{F}'} - P_{\boldsymbol{F}}}$$

$$= \frac{\frac{1}{\alpha A}\boldsymbol{d}^T (\boldsymbol{H}_{\boldsymbol{F}} + \Delta \boldsymbol{H})^{-1} \boldsymbol{c} - \frac{1}{\alpha A}\boldsymbol{d}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{c}}{(\boldsymbol{p}_{\boldsymbol{F}} + \Delta \boldsymbol{p})^T (\boldsymbol{H}_{\boldsymbol{F}} + \Delta \boldsymbol{H})^{-1} \boldsymbol{c} - \boldsymbol{p}_{\boldsymbol{F}}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{c}} \tag{47}$$

$$= \frac{\frac{1}{\alpha A}\boldsymbol{d}^T (\sum_{i=1}^{+\infty}(-1)^i (\boldsymbol{h}_q^T \boldsymbol{\delta}_q)^{i-1} (\boldsymbol{H}_{\boldsymbol{F}}^{-1} \Delta \boldsymbol{H}) \boldsymbol{H}_{\boldsymbol{F}}^{-1}) \boldsymbol{c}}{\sum_{i=1}^{+\infty}(-1)^i (\boldsymbol{h}_q^T \boldsymbol{\delta}_q)^{i-1} \boldsymbol{p}_{\boldsymbol{F}}^T (\boldsymbol{H}_{\boldsymbol{F}}^{-1} \Delta \boldsymbol{H}) \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{c} \\ - \sum_{i=1}^{+\infty}(-1)^i \zeta_q (\boldsymbol{h}_q^T \boldsymbol{\delta}_q)^{i-1} \boldsymbol{h}_q^T \boldsymbol{c}} \tag{48}$$

$$= \frac{\frac{1}{\alpha A}\boldsymbol{d}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \Delta \boldsymbol{H} \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{c}}{\boldsymbol{p}_{\boldsymbol{F}}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \Delta \boldsymbol{H} \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{c} - \zeta_q \boldsymbol{h}_q^T \boldsymbol{c}} \tag{49}$$

$$= \frac{\boldsymbol{d}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{\delta}_q}{\alpha A (\boldsymbol{p}_{\boldsymbol{F}}^T \boldsymbol{H}_{\boldsymbol{F}}^{-1} \boldsymbol{\delta}_q - \zeta_q)}. \tag{50}$$

APPENDIX D
PROOF OF THEOREM 3

Define $\mathcal{C} = \mathbf{conv}\,\{Z_{\boldsymbol{F}} | \boldsymbol{F} \in \mathcal{F}_D\}$ as the convex hull of points in the delay-power plane corresponding to deterministic scheduling policies. By proving $\mathcal{R} = \mathcal{C}$, we can have that $\mathcal{R}$ is a convex polygon whose vertices are all obtained by deterministic scheduling policies.

The proof contains three parts. In the first part, we will prove $\mathcal{R} \subseteq \mathcal{C}$ by the construction method. In the second part, we define the concepts of basic polygons and compound polygons, and prove that they are convex, based on which $\mathcal{R} \supseteq \mathcal{C}$ can be proven. By combining the results in these two parts, we will have $\mathcal{R} = \mathcal{C}$. Finally, in the third part, we will prove the policies corresponding to neighboring vertices of $\mathcal{R}$ are different in only one state.

Part I (Prove $\mathcal{R} \subseteq \mathcal{C}$): For any specific probabilistic policy $\boldsymbol{F}$ where $0 < f_{q^*,s^*} < 1$, we construct

$$\boldsymbol{F}' = \begin{cases} f'_{q,s} = 1 & q = q^*, s = s^* \\ f'_{q,s} = 0 & q = q^*, s \neq s^* \\ f'_{q,s} = f_{q,s} & \text{else} \end{cases} \tag{51}$$

and

$$\boldsymbol{F}'' = \begin{cases} f''_{q,s} = 0 & q = q^*, s = s^* \\ f''_{q,s} = \dfrac{f_{q,s}}{1 - f_{q^*,s^*}} & q = q^*, s \neq s^* \\ f''_{q,s} = f_{q,s} & \text{else.} \end{cases} \tag{52}$$

Since $0 \leq \frac{f_{q,s}}{1 - f_{q^*,s^*}} \leq 1$, and whenever $f_{q,s} = 0$, we have $f'_{q,s} = f''_{q,s} = 0$, the constructed policies $\boldsymbol{F}'$ and $\boldsymbol{F}''$ are feasible. It can be seen that $\boldsymbol{F} = f_{q^*,s^*}\boldsymbol{F}' + (1 - f_{q^*,s^*})\boldsymbol{F}''$. Since $\boldsymbol{F}$ is a convex combination of $\boldsymbol{F}'$ and $\boldsymbol{F}''$, also $\boldsymbol{F}'$ and $\boldsymbol{F}''$ are only different in the (q^*+1)th row, from Lemma 1, we know that $Z_{\boldsymbol{F}}$ is a convex combination of $Z_{\boldsymbol{F}'}$ and $Z_{\boldsymbol{F}''}$. Note that f'_{q^*,s^*} and f''_{q^*,s^*} are integers. Also, in $\boldsymbol{F}'$ and $\boldsymbol{F}''$, no new decimal elements will be introduced. Hence we can conclude that, in finite steps, the point $Z_{\boldsymbol{F}}$ can be expressed as a convex combination of points in the delay-power plane corresponding to deterministic scheduling policies, which means $Z_{\boldsymbol{F}} \in \mathcal{C}$. From the arbitrariness of $\boldsymbol{F}$, we can see $\mathcal{R} \subseteq \mathcal{C}$ is proven.

Part II (Prove $\mathcal{R} \supseteq \mathcal{C}$): In the second part, we will first define the concepts of basic polygons and compound polygons in Part II.0. Then basic polygons and compound polygons will be proven convex in Part II.1 and Part II.2 respectively. Based on these results, we will prove $\mathcal{R} \supseteq \mathcal{C}$ in Part II.3.

Part II.0 (Introduce the Concepts of Basic Polygons and Compound Polygons): For two deterministic policies $\boldsymbol{F}$ and $\boldsymbol{F}'$ which are different in K states, namely $q_1, \cdots, q_K$, define

$$F_{b_1,b_2,\cdots,b_K}(q,:) \\ = \begin{cases} (1 - b_k)\boldsymbol{F}(q,:) + b_k \boldsymbol{F}'(q,:) & q = q_k, \\ \boldsymbol{F}(q,:) & q \neq q_1, \cdots, q_K, \end{cases} \tag{53}$$

where $0 \leq b_k \leq 1$ for all k. Thus $\boldsymbol{F}_{0,0,\cdots,0} = \boldsymbol{F}$, and $\boldsymbol{F}_{1,1,\cdots,1} = \boldsymbol{F}'$. With more b_k equal to 0, the policy is more like $\boldsymbol{F}$. With more b_k equal to 1, the policy is more like $\boldsymbol{F}'$. For policies $\boldsymbol{F}_{b_1,\cdots,b_k,\cdots,b_K}$ and $\boldsymbol{F}_{b_1,\cdots,b'_k,\cdots,b_K}$ where $b_k \neq b'_k$, since they are different in only one state, according to Lemma 1, the delay-power point corresponding to their convex combination $Z_{\epsilon \boldsymbol{F}_{b_1,\cdots,b_k,\cdots,b_K} + (1-\epsilon)\boldsymbol{F}_{b_1,\cdots,b'_k,\cdots,b_K}}$ is the convex combination of $Z_{\boldsymbol{F}_{b_1,\cdots,b_k,\cdots,b_K}}$ and $Z_{\boldsymbol{F}_{b_1,\cdots,b'_k,\cdots,b_K}}$. However, for policies different in more than one state, the delay-power point corresponding to their convex combination is not necessarily the convex combination of their own delay-power points. Therefore, we introduce the concept of generated polygon to demonstrate the delay-power region of convex combinations of two policies. We plot $Z_{\boldsymbol{F}_{b_1,\cdots,b_K}}$, where $b_k = 0$ or 1 for all $1 \leq k \leq K$, and connect the points whose corresponding policies are different in only one state. Therefore any point on any line segment can be obtained by a certain policy. We define the figure as a polygon generated by $\boldsymbol{F}$ and $\boldsymbol{F}'$. The red polygon in Fig. 7a and the polygon in Fig. 8a are demonstrations where $\boldsymbol{F}$ and $\boldsymbol{F}'$ are different in 2 and 3 states respectfully. If $K = 2$, we call the polygon a basic polygon. If $K > 2$, we call it a compound polygon. As demonstrated in Fig. 8a, a compound polygon contains multiple basic polygons.

Part II.1 (Prove a Basic Polygon Is Convex and Any Point Inside a Basic Polygon can be Obtained by a Policy): For
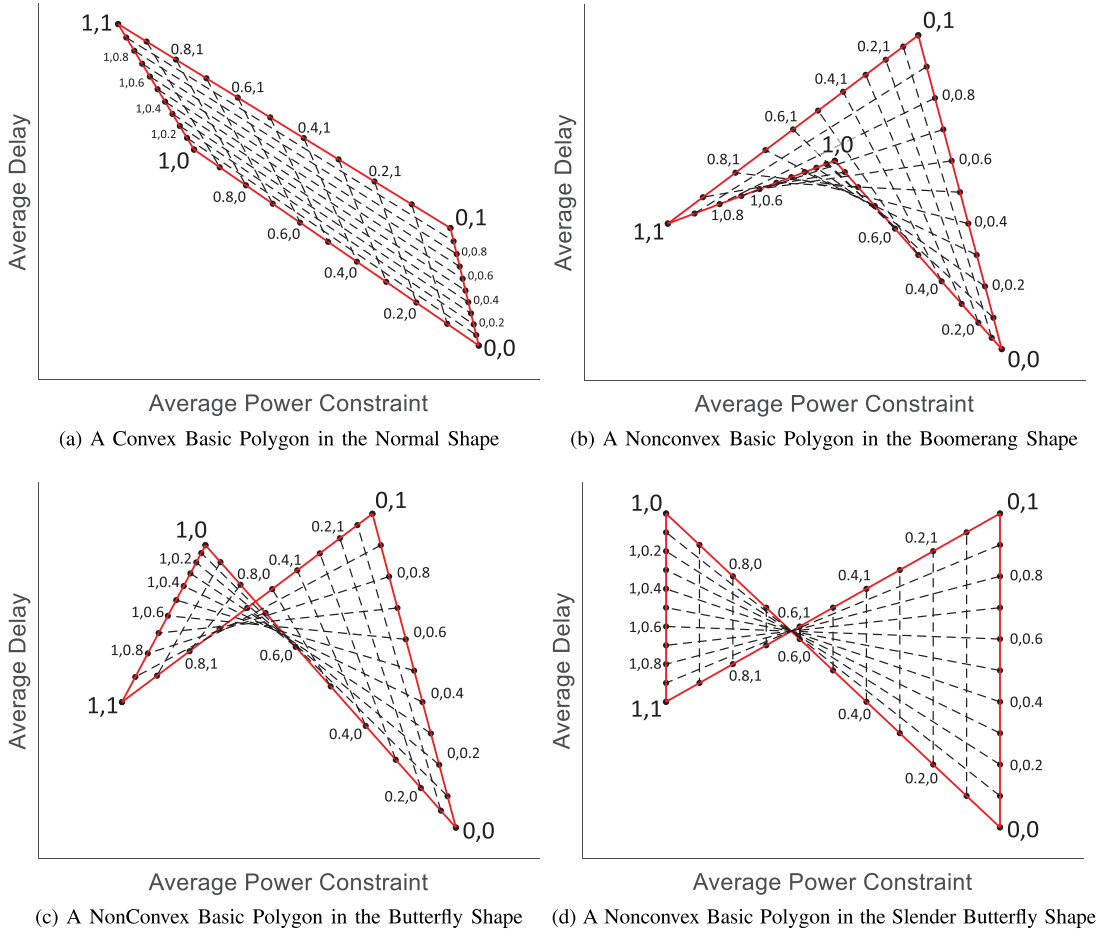
(a) A Convex Basic Polygon in the Normal Shape

(b) A Nonconvex Basic Polygon in the Boomerang Shape

(c) A NonConvex Basic Polygon in the Butterfly Shape

(d) A Nonconvex Basic Polygon in the Slender Butterfly Shape

Fig. 7. Demonstration for Basic Polygons.

(a) A Convex Compound Polygon
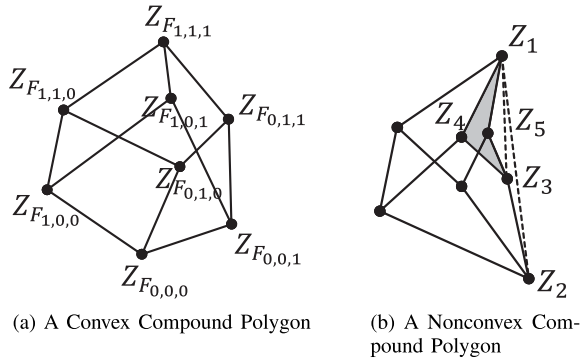
(b) A Nonconvex Compound Polygon

Fig. 8. Demonstration for Compound Polygons.

better visuality, in Fig. 7, we simplify the notation $Z_{F_{b_1,b_2}}$ as b_1, b_2. According to different relative positions of $Z_{F_{0,0}}$, $Z_{F_{0,1}}$, $Z_{F_{1,0}}$, and $Z_{F_{1,1}}$, there are in total 3 possible shapes of basic polygons, as shown in Fig. 7a-7c respectfully. We name them as the normal shape, the boomerang shape, and the butterfly shape. The degenerate polygons such as triangles, line segments and points are considered included in the above three cases. Besides F_{b_1,b_2} with integral b_1, b_2 and the line segments connecting them, we also plot the points corresponding to policy F_{b_1,b_2} where one of b_1, b_2 is integer and the other one is decimal. We connect the points corresponding

to policies with the same b_1 or b_2 in dashed lines. As demonstrated in Fig. 7, we draw line segments $\overline{Z_{F_{b_1,0}}Z_{F_{b_1,1}}}$ where $b_1 = 0.1, 0.2, \cdots, 0.9$ and $\overline{Z_{F_{0,b_2}}Z_{F_{1,b_2}}}$ where $b_2 = 0.1, 0.2, \cdots, 0.9$. For any specific b_1 and b_2, the point $Z_{F_{b_1,b_2}}$ should be on both $\overline{Z_{F_{b_1,0}}Z_{F_{b_1,1}}}$ and $\overline{Z_{F_{0,b_2}}Z_{F_{1,b_2}}}$. Because of the existence of $Z_{F_{b_1,b_2}}$, line segments $\overline{Z_{F_{b_1,0}}Z_{F_{b_1,1}}}$ and $\overline{Z_{F_{0,b_2}}Z_{F_{1,b_2}}}$ should always have an intersection point for any specific b_1 and b_2. However, if there exist line segments outside the polygon, there exist b_1 and b_2 whose line segments don't intersect. Therefore, in the boomerang shape, there will always exist b_1 and b_2 whose line segments don't intersect. In the butterfly shape, there will exist b_1 and b_2 whose line segments don't intersect except the case that all the line segments are inside the basic polygon, as shown in Fig. 7d, which is named as the slender butterfly shape. In the slender butterfly shape, there exists a specific b_1^* such that $\overline{Z_{F_{b_1^*,0}}Z_{F_{b_1^*,1}}}$ degenerates into a point, or there exists a specific b_2^* such that $\overline{Z_{F_{0,b_2^*}}Z_{F_{1,b_2^*}}}$ degenerates into a point. Without loss of generality, we assume it is the b_1^* case. It means that under policy $F_{b_1^*,b_2}$, state q_2, the state corresponding to b_2, is a transient state. For $b_1 \in (b_1^* - \epsilon, b_1^* + \epsilon)$ when ϵ is small enough, the Markov chain applying policy F_{b_1,b_2} also has q_2 as a transient state, therefore $\overline{Z_{F_{b_1,0}}Z_{F_{b_1,1}}}$ also degenerates into a point. Thus $\overline{Z_{F_{0,0}}Z_{F_{1,0}}}$ and $\overline{Z_{F_{0,1}}Z_{F_{1,1}}}$ overlap, which

means the slender butterfly shape always degenerates to a line segment, which can also be considered as a normal shape. Since the normal shape is the only possible shape of a basic polygon, the basic polygon is convex. Since the transition from the point $Z_{F_{0,0}}$ to $Z_{F_{1,1}}$ is termwise monotone and continuous, every point inside the basic polygon can be obtained by a policy.

Part II.2 (Prove a Compound Polygon Is Convex): For any two deterministic policies F and F', if their generated compound polygon is not convex, then there exist two vertices whose connecting line is outside the compound polygon, as demonstrated by $\overline{Z_1 Z_2}$ in Fig. 8b. Thus, there must exist two vertices who are connecting to the same point such that their connecting line is outside the compound polygon, as demonstrated by $\overline{Z_1 Z_3}$. The policy corresponding to these two vertices must be different in only two states, therefore there must be a basic polygon generated by them, as demonstrated by the filled polygon. Since $\overline{Z_1 Z_3}$ is outside the compound polygon, it is outside the basic polygon too, which is not possible because basic polygons are always convex. Therefore all generated compound polygons are convex.

Part II.3 (Prove $\mathcal{R} \supseteq \mathcal{C}$): For any point $C \in \mathcal{C}$, it will surely fall into one of the compound polygons. Because otherwise, there will be at least one point corresponding to a deterministic policy which is outside any compound polygons, which is impossible. Any compound polygon is covered by basic polygons, therefore C is inside at least one basic polygon. Since any point inside a basic polygon can be obtained by a policy, the point $C \in \mathcal{R}$. From the arbitrariness of C, we have $\mathcal{R} \supseteq \mathcal{C}$.

From Part II.1 and Part II.2, it can be seen that $\mathcal{R} = \mathcal{C}$. Since there are only finite deterministic policies in total, the set $\mathcal{R}$ is a convex polygon whose vertices are all obtained by deterministic scheduling policies.

Part III. (Neighboring Vertices of $\mathcal{R}$): For any two neighboring vertices Z_F and $Z_{F'}$ of $\mathcal{R}$, if F and F' are different in more than one state, their generated polygon is convex. If the line segment $\overline{Z_F Z_{F'}}$ is inside the generated polygon, Z_F and $Z_{F'}$ are not neighboring vertices. If the line segment $\overline{Z_F Z_{F'}}$ is on the boundary of the generated polygon, there will be other vertices between them, such that Z_F and $Z_{F'}$ are not neighboring, neither. Therefore, policies F and F' are deterministic and different in only one state.

i.e., $\mathcal{R} = \bigcap_{i=1}^{I}\{(P, D)|a_i P + b_i D \geq c_i\}$. We divide (a_i, b_i, c_i) into 2 categories according to the value of a_i and b_i as (a_i^+, b_i^+, c_i^+) for $i = 1, \cdots, I^+$ if $a_i > 0$ and $b_i > 0$, and (a_i^-, b_i^-, c_i^-) for $i = 1, \cdots, I^-$ if $a_i \leq 0$ or $b_i \leq 0$. We have $I = I^+ + I^-$ and $I^+, I^- > 0$. Then $\mathcal{R} = \bigcap_{i=1}^{I^+}\{(P, D)|a_i^+ P + b_i^+ D \geq c_i^+\} \cap \bigcap_{i=1}^{I^-}\{(P, D)|a_i^- P + b_i^- D \geq c_i^-\}$. For $1 \leq l \leq I^+$, define $\mathcal{L}_l = \{(P, D)|a_l^+ P + b_l^+ D = c_l^+\} \cap \bigcap_{i=1, i \neq l}^{I^+}\{(P, D)|a_i^+ P + b_i^+ D \geq c_i^+\} \cap \bigcap_{i=1}^{I^-}\{(P, D)|a_i^- P + b_i^- D \geq c_i^-\}$.

For all $(P, D) \in \mathcal{L}_l$, immediately we have $(P, D) \in \mathcal{R}$. For all $(P', D') \in \mathcal{R}$, since $a_l^+ P' + b_l^+ D' \geq c_l^+ = a_l^+ P + b_l^+ D$, it should hold that $P' \geq P$ or $D' \geq D$. According to the definition of $\mathcal{L}$, we have $(P, D) \in \mathcal{L}$. Therefore $\mathcal{L}_l \subseteq \mathcal{L}$.

For all $(P, D) \in \mathcal{L}$, we investigate three cases: 1) If $a_i^+ P + b_i^+ D > c_i^+$ for all $1 \leq i \leq I^+$ and $a_i^- P + b_i^- D > c_i^-$ for all $b_i^- > 0$, set $\epsilon = \min_{b_i > 0} \frac{a_i P + b_i D - c_i}{b_i}$ so that $a_i P + b_i(D - \epsilon) \geq c_i$ for all $b_i > 0$. Since $(P, D) \in \mathcal{R}$, for all $b_i \leq 0$ $a_i P + b_i D \geq c_i$, therefore $a_i P + b_i(D - \epsilon) \geq c_i$ for all $b_i \leq 0$. Hence $(P, D - \epsilon) \in \mathcal{R}$, which is against the definition of $\mathcal{L}$. 2) If $a_i^+ P + b_i^+ D > c_i^+$ for all $1 \leq i \leq I^+$ and $a_i^- P + b_i^- D > c_i^-$ for all $a_i^- > 0$, set $\epsilon = \min_{a_i > 0} \frac{a_i P + b_i D - c_i}{a_i}$ so that $a_i(P - \epsilon) + b_i D \geq c_i$ for all $a_i > 0$. Since $(P, D) \in \mathcal{R}$, for all $a_i \leq 0$ $a_i P + b_i D \geq c_i$, therefore $a_i(P - \epsilon) + b_i D \geq c_i$ for all $a_i \leq 0$. Hence $(P - \epsilon, D) \in \mathcal{R}$, which is against the definition of $\mathcal{L}$. 3) If $a_i^+ P + b_i^+ D > c_i^+$ for all $1 \leq i \leq I^+$, and there exists i^* and j^* such that $a_{i^*}^- \leq 0, b_{i^*}^- > 0, a_{j^*}^- > 0, b_{j^*}^- \leq 0, a_{i^*}^- P + b_{i^*}^- D = c_{i^*}^-, a_{j^*}^- P + b_{j^*}^- D = c_{j^*}^-$. For all $(P', D') \in \mathcal{R}$, either $P' \geq P, D' \geq D$ or $P' \leq P, D' \leq D$. If there exists $P' < P$ and $D' < D$, then (P, D) is against the definition of $\mathcal{L}$. If $P' \geq P$ and $D' \geq D$ for all (P', D'), since for all $1 \leq i \leq I^+$, we have $a_i^+ P + b_i^+ D > c_i^+$, therefore $a_i^+ P' + b_i^+ D' > c_i^+$. Hence $\mathcal{L}_i \cap \mathcal{R} = \emptyset$, which is against the condition. From the above three cases, for all $(P, D) \in \mathcal{L}$, there exists at least one certain l^* such that $a_{l^*}^+ P + b_{l^*}^+ D = c_{l^*}^+$, which means $(P, D) \in \mathcal{L}_{l^*}$.

From above we can see that $\mathcal{L} = \bigcup_{l=1}^{I^+} \mathcal{L}_l$. Therefore $\mathcal{L}$ is piecewise linear.

Properties of Vertices of $\mathcal{L}$: The vertices of $\mathcal{L}$ are also the vertices of $\mathcal{R}$, and neighboring vertices of $\mathcal{L}$ are also neighboring vertices of $\mathcal{R}$. From the results in Theorem 3, vertices of $\mathcal{L}$ are obtained by deterministic scheduling policies, and the policies corresponding to neighboring vertices of $\mathcal{L}$ are different in only one state.

APPENDIX E
PROOF OF COROLLARY 3

Monotonicity: Since $\mathcal{L} = \{(P, D) \in \mathcal{R}|\forall(P', D') \in \mathcal{R}$, either $P' \geq P$ or $D' \geq D\}$, for any $(P_1, D_1), (P_2, D_2) \in \mathcal{L}$ where $P_1 < P_2$, we should have $D_1 \geq D_2$. Therefore $\mathcal{L}$ is decreasing.

Convexity: Since $\mathcal{R}$ is a convex polygon, for any $(P_1, D_1), (P_2, D_2) \in \mathcal{L}$, their convex combination is $(\theta P_1 + (1 - \theta)P_2, \theta D_1 + (1 - \theta)D_2) \in \mathcal{R}$. Therefore, there should be a point (P_θ, D_θ) on $\mathcal{L}$ where $P_\theta = \theta P_1 + (1 - \theta)P_2$, and $D_\theta \leq \theta D_1 + (1 - \theta)D_2$. Therefore $\mathcal{L}$ is convex.

Piecewise Linearity: Since $\mathcal{R}$ is a convex polygon, it can be expressed as the intersection of a finite number of halfspaces,

REFERENCES

[1] M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels*. New York, NY, USA: Wiley, 2005.

[2] B. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. 37th Allerton Conf. Commun. Control, Comput.*, Monticello, IL, USA, 1999, pp. 709–717.

[3] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[4] M. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," in *Proc. IEEE INFOCOM*, Mar. 2003, pp. 311–320.

[5] I. Bettesh and S. S. Shamai (Shitz), "Optimal power and rate control for minimal average delay: The single-user case," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4115–4141, Sep. 2006.

[6] R. Berry, "Optimal power-delay tradeoffs in fading channels—Small-delay asymptotics," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3939–3952, Jun. 2013.

[7] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.

[8] M. Agarwal, V. S. Borkar, and A. Karandikar, "Structural properties of optimal transmission policies over a randomly varying channel," *IEEE Trans. Autom. Control*, vol. 53, no. 6, pp. 1476–1491, Jul. 2008.

[9] D. V. Djonin and V. Krishnamurthy, "MIMO transmission control in fading channels—A constrained Markov decision process formulation with monotone randomized policies," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5069–5083, Oct. 2007.

[10] M. H. Ngo and V. Krishnamurthy, "Optimality of threshold policies for transmission scheduling in correlated fading channels," *IEEE Trans. Commun.*, vol. 57, no. 8, pp. 2474–2483, Aug. 2009.

[11] M. H. Ngo and V. Krishnamurthy, "Monotonicity of constrained optimal transmission policies in correlated fading channels with ARQ," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 438–451, Jan. 2010.

[12] M. H. Cheung and J. Huang, "DAWN: Delay-aware Wi-Fi offloading and network selection," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1214–1223, Jun. 2015.

[13] W. Chen, Z. Cao, and K. B. Letaief, "Optimal delay-power tradeoff in wireless transmission with fixed modulation," in *Proc. IEEE IWCLD*, Sep. 2007, pp. 60–64.

[14] X. Chen and W. Chen, "Joint channel-buffer aware energy-efficient scheduling over fading channels with short coherent time," in *Proc. IEEE ICC*, Jun. 2015, pp. 2810–2815.

[15] X. Chen and W. Chen, "Delay-optimal probabilistic scheduling for low-complexity wireless links with fixed modulation and coding: A cross-layer design," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8036–8051, Oct. 2015.

[16] J. Yang and S. Ulukus, "Delay-minimal transmission for average power constrained multi-access communications," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2754–2767, Sep. 2010.

[17] G. Saleh, A. El-Keyi, and M. Nafie, "Cross-layer minimum-delay scheduling and maximum-throughput resource allocation for multiuser cognitive networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 4, pp. 761–773, Apr. 2013.

[18] X. Chen and W. Chen, "Delay-optimal buffer-aware probabilistic scheduling with adaptive transmission," in *Proc. IEEE/CIC ICCC*, Nov. 2015, pp. 1–6.

[19] E. Altman, *Constrained Markov Decision Processes*. London, U.K.: Chapman & Hall, 1999.

[20] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 2014.

[21] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 2. Belmont, MA, USA: Athena Scientific, 1995.

Xiang Chen (S'14) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. Since then, he has been pursuing the Ph.D. degree with the Department of Electronic Engineering, Tsinghua University. From 2015 to 2016, he visited the Department of Electrical and Computer Engineering, The Ohio State University. His research interests are in the areas of communications and networking.

Wei Chen (S'05–M'07–SM'13) received the B.S. and Ph.D. degrees (Hons.) from Tsinghua University in 2002 and 2007, respectively. From 2005 to 2007, he was a Visiting Ph.D. Student with The Hong Kong University of Science and Technology. Since 2007, he has been a Faculty Member with Tsinghua University, where he is currently a tenured Full Professor, the Director of the Academic Degree Office, Tsinghua University, and a member of the University Council. During 2014 to 2016, he served as the Deputy Head of the Department of Electronic Engineering. He has also held visiting appointments at several universities, including most recently at Princeton. His research interests are in the areas of wireless networks and information theory.

Dr. Chen is a member of the National 10000-Talent Program, a Cheung Kong Young Scholar, and a Chief Scientist of the National 973 Youth Project. He is also supported by the NSFC Excellent Young Investigator Project, the New Century Talent Program of Ministry of Education, and the Beijing Nova Program. He received the first prize of 14th Henry Fok Ying-Tung Young Faculty Award, the 17th Yi-Sheng Mao Science & Technology Award for Beijing Youth, the 2017 Young Scientist Award of the China Institute of Communications, and the 2015 Information Theory New Star Award of the China Institute of Electronics. He received the IEEE Marconi Prize Paper Award and the IEEE Comsoc Asia Pacific Board Best Young Researcher Award in 2009 and 2011, respectively. He is a winner of the National May 1st Medal. He serves as an Editor of the IEEE TRANSACTIONS COMMUNICATIONS and the IEEE TRANSACTIONS ON EDUCATION. He also serves as the Executive Chairman of the Youth Forum of the China Institute of Communications.
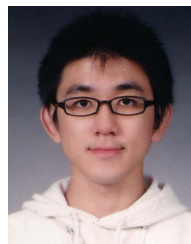
Joohyun Lee (S'11–M'14) received the B.S. and integrated M.S. and Ph.D. degrees from the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2008 and 2014, respectively. He is currently a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, The Ohio State University. His research interests are in the areas of context-aware networking and computing, mobility-driven cellular traffic offloading, energy-efficient mobile networking, protocol design and analysis for delay-tolerant networks, and network economics and pricing. He received the IEEE William R. Bennett Prize Paper Award in 2016, given to the best original paper published in the IEEE/ACM TRANSACTIONS ON NETWORKING in the previous three calendar years.

Ness B. Shroff (S'91–M'93–SM'01–F'07) received the Ph.D. degree in electrical engineering from Columbia University in 1994. He joined the School of ECE, Purdue University, as an Assistant Professor. He became a Full Professor of ECE in 2003 and the Director of CWSA in 2004, Purdue University, a university-wide center on wireless systems and applications. In 2007, he joined The Ohio State University, where he holds the Ohio Eminent Scholar Endowed Chair in networking and communications with the Department of ECE and the Department of CSE. He holds or has held visiting (chaired) professor positions at Tsinghua University, Beijing, China, Shanghai Jiaotong University, Shanghai, China, and IIT Bombay, India. He is currently an Editor-at-Large of the IEEE/ACM TRANSACTIONS ON NETWORKING, a Senior Editor of the IEEE TRANSACTIONS ON CONTROL OF NETWORKED SYSTEMS, and a Technical Editor of the *IEEE Network Magazine*. He has received numerous best paper awards for his research and is listed in Thomson Reuters' *The World's Most Influential Scientific Minds*, and is noted as a Highly Cited Researcher by Thomson Reuters. He also received the IEEE INFOCOM Achievement Award for seminal contributions to scheduling and resource allocation in wireless networks.