Joint Scheduling- Traffic Admission Control: Structural Results and Online Learning Algorithm

Khoa T. Phan[†], Tho Le-Ngoc[†], Mihaela van der Schaar[‡], and Fangwen Fu[§]
[†] Department of Electrical and Computer Engineering, McGill University, Montreal, Canada
[‡] Electrical Engineering Department, University of California, Los Angeles (UCLA), Los Angeles, USA
[§] Intel Corporation, California, USA
Email: khoa.phan@mail.mcgill.ca, tho.le-ngoc@mcgill.ca, mihaela@ee.ucla.edu, fu.fangwen@gmail.com

***Abstract*—This work studies the joint scheduling- admission control (SAC) problem over a fading channel. In particular, the optimal trade-off between maximizing the throughput and minimizing the queue size (or average congestion) is investigated. The SAC problem is formulated as a constrained Markov decision process (MDP) to maximize a utility defined as a function of the throughput and the queue size. The structural properties of the optimal policies are subsequently derived. When the statistical knowledge of the traffic arrival and channel processes is not available, we propose an online learning algorithm for the optimal policies. The analysis and algorithm development are relied on the reformulation of the Bellman's optimality dynamic programming equation using suitably defined value functions which can be learned using online time-averaging.**

–Key words–: Scheduling, traffic admission control, Markov decision process (MDP), learning, structural results.

I. Introduction

On the communications over time-varying channels, when the probability distribution functions (PDFs) of the channel and traffic arrival processes are known a-priori, optimal scheduling policies can be analyzed and computed off-line [1]–[4]. However, such knowledge is often unavailable a-priori in real-life communications, hence, developing online scheduling algorithms without requiring known PDFs is important [5]–[8]. While the above works have addressed these issues for the scheduling problem *without* traffic admission control, our current work studies the joint scheduling- traffic admission control (SAC) problem.

In the scheduling without admission control, the central concept is the power- delay trade-off [1]. That says, a delay (or an average congestion) requirement can be attained by increasing the transmission power, i.e., increasing the service rate. However, when there is a constraint on the maximum power, a delay bound might be impossible to achieve. One solution is to implement admission control to limit the traffic entering the buffer by admitting only a portion of the arrival traffic. Also, admission control is required to ensure queue stability (finite queue) when the power budget is smaller than the minimum power required to stabilize the queue without admission control. It is clear that in the systems with SAC, there is a trade-off between maximizing the throughput and minimizing the average queue size. The work in [9] proposes the energy constrained control algorithm (ECCA) to stabilize the queue and maximize the throughput using Lyapunov optimization theory. Although simple, ECCA cannot achieve the optimal throughput- queue size trade-off because it does not learn the system dynamics. Alternatively, this work focuses on the control policies achieving the *optimal* trade-off in *all* traffic loading regions.

This work formulates the SAC problem as a constrained Markov decision process (MDP) to maximize a utility (or reward) defined as the difference between the *throughput benefit* and the *buffer cost* (or congestion cost). The benefit and cost functions are increasing functions of the throughput, and the buffer size, respectively. Such utility functions capture the inherent trade-off between maximizing the throughput and minimizing the queue size. Then, using the stochastic control tools, we can derive the structural properties of the optimal policies. Moreover, this work develops an online learning algorithm for the optimal policies without requiring the explicit knowledge on the system dynamics. Our approach is to introduce new value functions which are used to rewrite the Bellman's equation. The resulting equation is amenable to online learning via online time-averaging.

II. Optimal Joint Scheduling- Admission Control

A. System description

We consider a SAC model where a single user (a transmitter- receiver pair) transmits data stored in a buffer over a fading channel. Time is divided into slots of equal duration. The dynamics of the buffer (or queue) is controlled using admission control and scheduling actions. Specifically, in each slot, the scheduling action computes the amount of traffic removed from the buffer for transmission to the receiver. Also, the admission control action determines the amount of traffic (from the newly-arriving traffic) to be stored into the buffer.

The wireless channel is assumed to be block-fading over the time slots. Denote h^t as the channel state representing the power gain in slot t, $t = 0, 1, \ldots$. We assume:

(A1) The channel process $\{h^t\} \in \mathcal{H}$ is independent and identically distributed (i.i.d.) over slots with general PDF $p_{\mathcal{H}}(h^t)$ over a finite channel state space $\mathcal{H}$.

978-1-4673-3122-7/13/$31.00 ©2013 IEEE

Denote $\mathcal{B} \in [0, \infty)$[1] and $b^t \in \mathcal{B}$ as the queue state space and the queue state representing the queue size (in number of bits) in slot t, respectively. Let a^t, $a^t \in [0, b^t]$ (in number of bits) denote the scheduling action in slot t. Moreover, let y^t and r^t (in number of bits) represent the amount of new arrivals to the system and the amount of arrivals admitted into the queue in slot t, $r^t \in [0, y^t]$. We assume:

(A2) The traffic arrival process $\{y^t\} \in \mathcal{Y} = [0, y_{\max}]$ is i.i.d. over slots with general PDF $p_{\mathcal{Y}}(y^t)$ [1].

Given b^0 as the initial backlog, the queue dynamics across the time slots satisfy the Lindley's recursion:

$$b^{t+1} = [b^t - a^t]^+ + r^t \tag{1}$$

where $[x]^+$ denotes $\max\{x, 0\}$. Note that without admission control, $r^t = y^t$ for all slots.

The reliable transmission of a^t (bits) under channel state h^t in slot t incurs a power $c(h^t, a^t)$.[2] We assume:

(A3) The power functions $c(h, a)$ are strictly convex increasing differential with a; strictly decreasing with h; $c(h, 0) = 0$, and $\lim_{a\to\infty} c(h, a) = \infty$.

We define the throughput R as $R \triangleq \liminf_{t\to\infty} \frac{1}{t} E\left\{\sum_{\tau=0}^{t-1} r^\tau\right\}$ where the expectation operator $E\{.\}$ is taken over the probability measure induced by the random processes and some SAC policy (to be defined later). The (average) queue size (or average congestion) and power consumption are, respectively, $B \triangleq \limsup_{t\to\infty} \frac{1}{t} E\left\{\sum_{\tau=0}^{t-1} b^\tau\right\}$ and $C \triangleq \limsup_{t\to\infty} \frac{1}{t} E\left\{\sum_{\tau=0}^{t-1} c(h^\tau, a^\tau)\right\}$. It is assumed that the power C does not exceed a maximum value $C_{\max}$.

The utility u^t obtained in slot t is defined as the difference between the *throughput benefit* obtained $f_b(r^t)$ and the *buffer cost* $f_c(b^t)$ incurred in the same slot, i.e., $u^t \triangleq f_b(r^t) - f_c(b^t)$. The (average) utility is defined as:

$$U \triangleq \liminf_{t\to\infty} \frac{1}{t} E\left\{\sum_{\tau=0}^{t-1} f_b(r^\tau) - f_c(b^\tau)\right\}. \tag{2}$$

We make the following assumption:

(A4) The benefit function $f_b(r)$ is increasing concave differential with r; The cost function $f_c(b)$ is increasing convex differential with b [5].

B. Optimal SAC problem formulation

The optimal SAC problem can be posed as:

$$\max_{\pi\in\Pi} \quad U \quad \text{such that:} \quad C \le C_{\max} \tag{3}$$

where Π is the set of all feasible (or admissible) SAC control policies π.

The formulation (3) with $f_b(r) = r$ and $f_c(b) = \kappa b$ for some positive κ can be used to study the SAC problem to maximize the throughput under the constraint on the maximum queue size [9] because they have similar Lagrangian functions.

1) Optimal throughput-queue size trade-off: To study the trade-off, we let the functions be $f_b(r) = r$ and $f_c(b) = \kappa b$ for some coefficients $\kappa \in [0, 1)$.[3] The corresponding maximum objective value of (3) is $U^* = R^* - \kappa B^*$ where R^* and B^* are the throughput and the queue size. Since U^* is maximized, R^* is the maximum throughput such that the queue size is less than or equal to B^*. More generally, now for any B, define $R(B)$ to be the maximum throughput such that the queue size is less than or equal to B. With this definition, we have $R(B^*) = R^*$. Proposition 1[4] characterizes the optimal trade-off $R(B)$.

Proposition 1: Under maximum power constraint, $R(B)$ is concave increasing of B.

The points on the curve $R(B)$ are obtained by varying the coefficients $\kappa \in (0, 1)$.

Since the cost function $f_c(b)$ is unbounded increasing with the queue size (assumption (A4)), the objective function in (3) is unbounded decreasing with the queue size. Hence, the optimal solutions of (3) must result in finite queue size, and hence, the underlying Markov chain is irreducible. Consequently, according to Theorem 12.7 in [10], the constrained MDP problem (3) admits an optimal solution that can be found using the Lagrangian approach:

$$\min_{\beta>0} \left\{\max_{\pi\in\Pi} \left\{U - \beta C\right\} + \beta C_{\max}\right\}. \tag{4}$$

Therefore, to study (4) (and thus (3)), we can first study the inner maximization for a given positive multiplier β:

$$\max_{\pi\in\Pi} \left\{U - \beta C\right\}. \tag{5}$$

III. Optimal SAC Policies: Structural Results and Online Learning Algorithm

It is assumed that the scheduling controller cannot observe the arrival state y^t when making the scheduling decision a^t. Moreover, the action a^t is determined first based on the state (b^t, h^t) and the action r^t is determined after based on the state $([b^t - a^t]^+, y^t)$. Hence, a stationary control policy π for (5) consists of a scheduling policy represented by a function $a : \mathcal{B} \times \mathcal{H} \to \mathbb{R}^+$ and an admission control policy represented by a function $r : \mathcal{B} \times \mathcal{Y} \to \mathbb{R}^+$. The scheduling policy specifies a^t as a function of the state (b^t, h^t), i.e., $a^t = a(b^t, h^t) \in [0, b^t]$; The admission control policy specifies r^t as a function of the state $(\hat{b}^t = b^t - a^t, y^t)$, i.e., $r^t = r(\hat{b}^t, y^t) \in [0, y^t]$.

A. Post-transmission and post-admission states and corresponding state value functions

In (5), define $J(b, h)$ as the state value function for the state $(b, h) \in \mathcal{B} \times \mathcal{H}$, i.e., the optimal value of (5) with starting

[1] We allow the buffer to be an arbitrary real value for mathematical convenience.

[2] One possible power function is derived from the Shannon theoretic function $c(h, a) = (2^a - 1)/h$ which will be used in the simulation section.

[3] Linear buffer cost model has been used in several works [4], [8], and is related to the queuing delay by Little's theorem. Moreover, since it holds true that $R < B$, $\kappa \in [0, 1)$ to avoid triviality, otherwise, no traffic is admitted.

[4] The proofs of the presented results are omitted due to space limitation.

state $(b^0, h^0) = (b, h)$. The Bellman's optimality dynamic programming equation for (5) is:

$$J(b,h) = \max_{a:a\le b} \Bigg\{ -f_c(b) - \beta c(h,a) + \sum_{y'\in\mathcal{Y}} p_{\mathcal{Y}}(y') \Bigg(\max_{r:r\le y'} \Big\{ f_b(r) + \sum_{h'\in\mathcal{H}} p_{\mathcal{H}}(h') J(b-a+r, h') \Big\} \Bigg) - J(b_0, h_0) \Bigg\} \quad (6)$$

for some arbitrary but fixed state $(b_0, h_0) \in \mathcal{B}\times\mathcal{H}$. The optimal policy π^* consists of the optimal solutions of the two maximization operators in (6). When the PDFs are known, $J(b,h)$ can be found using relative value iteration algorithm (RVIA). When the PDFs are unknown, we propose an approach which allows online learning of the optimal policies.

We introduce two new states and their corresponding state value functions. The post-admission state value function $J_{\mathrm{p-ad}}(\bar{b})$ is defined as:

$$J_{\mathrm{p-ad}}(\bar{b}) = \sum_{h'\in\mathcal{H}} p_{\mathcal{H}}(h') J(\bar{b}, h') \quad (7)$$

for post-admission backlog state $\bar{b}$, $\bar{b} \in \mathcal{B}$. Hence, the post-admission state $\bar{b}^t$ in slot t equals to the backlog state b^{t+1} in slot $t+1$. The post-transmission state value function $J_{\mathrm{p-tr}}(\hat{b})$ is defined as:

$$J_{\mathrm{p-tr}}(\hat{b}) = \sum_{y'\in\mathcal{Y}} p_{\mathcal{Y}}(y') \Big(\max_{r:r\le y'} \Big\{ f_b(r) + J_{\mathrm{p-ad}}(\hat{b}+r) \Big\} \Big). \quad (8)$$

for post-transmission states $\hat{b}$, $\hat{b} \in \mathcal{B}$. By definition, in slot t, we have $\hat{b}^t = [b^t - a^t]^+$ and $\bar{b}^t = \hat{b}^t + r^t$. From (6), the optimal policy π^* consists of the solutions of the following optimization problems:

$$a^*(b,h) = \arg\max_{a:a\le b} \Big\{ -f_c(b) - \beta c(h,a) + J_{\mathrm{p-tr}}(b-a) \Big\} \quad (9)$$

$$r^*(\hat{b}, y) = \arg\max_{r:r\le y} \Big\{ f_b(r) + J_{\mathrm{p-ad}}(\hat{b}+r) \Big\}. \quad (10)$$

Hence, if the value functions are known, the optimal policy can be derived. Later, we show that online learning of the value functions is possible using online time-averaging.

From (6), we also have the following relationship:

$$J_{\mathrm{p-ad}}(\bar{b}) = \sum_{h'\in\mathcal{H}} p_{\mathcal{H}}(h') \max_{a:a\le\bar{b}} \Big\{ -f_c(\bar{b}) - \beta c(h',a) + J_{\mathrm{p-tr}}(\bar{b}-a) \Big\}. \quad (11)$$

The structural properties of the optimal policy π^* are stated.

Theorem 1: The optimal policy π^* of (5) has the following properties:

1. The value functions $J_{\mathrm{p-ad}}(\bar{b})$, and $J_{\mathrm{p-tr}}(\hat{b})$ are concave decreasing.
2. The admission control action $r^*(\hat{b}, y)$ is non-increasing with $\hat{b}$, non-decreasing with y, and has the following form:
$$r^*(\hat{b}, y) = \min\{\bar{B}, \hat{b} + y\} \quad (12)$$
where $\bar{B}$ is some threshold.
3. The scheduling action $a^*(b,h)$ is non-decreasing with b and non-decreasing with h.

Theorem 1 says that the admission control policy can be emulated using a finite buffer with size $\bar{B}$ and the queue dynamics in (1) can be represented as follows for $t = 0, 1, \ldots$:

$$b^{t+1} = \min\Big\{ \bar{B}, [b^t - a^t]^+ + y^t \Big\}. \quad (13)$$

The EECA in [9] prescribes that, in every slot, all new arrivals are admitted whenever the current backlog is below a predetermined threshold. Else, all new arrivals are dropped. Such admission control policy is sub-optimal.

B. Stochastic approximation based online learning algorithm

Using (8) and (11), the sequential RVIA equations for the value functions can be written as follows for $t = 0, 1, \ldots$:

$$J^{t+1}_{\mathrm{p-tr}}(\hat{b}) = \sum_{y'\in\mathcal{Y}} p_{\mathcal{Y}}(y') \Big(\max_{r:r\le y'} \Big\{ f_b(r) + J^t_{\mathrm{p-ad}}(\hat{b}+r) \Big\} \Big) - J^t_{\mathrm{p-tr}}(\hat{b}_0) \quad (14)$$

$$J^{t+1}_{\mathrm{p-ad}}(\bar{b}) = \sum_{h'\in\mathcal{H}} p_{\mathcal{H}}(h') \Big(\max_{a:a\le\bar{b}} \Big\{ f_c(\bar{b}) - \beta c(h',a) + J^{t+1}_{\mathrm{p-tr}}(\bar{b}-a) \Big\} \Big) - J^t_{\mathrm{p-ad}}(\bar{b}_0) \quad (15)$$

with initial conditions $J^0_{\mathrm{p-ad}}(\bar{b}) = 0$, $J^0_{\mathrm{p-tr}}(\hat{b}) = 0$, $\bar{b}, \hat{b} \in \mathcal{B}$ and $\hat{b}_0$, $\bar{b}_0$ are arbitrary but fixed states. The iterations converge to the state value functions satisfying (7), (8), and (11) [10].

The iterations (14)–(15) require known PDFs to evaluate the expectations. Fortunately, since the expectations are outside of the maximization operators in (14)–(15), a learning algorithm can be developed by removing the expectation operators and then using online time averaging to learn the value functions under unknown PDFs, i.e., it solves the MDP (5) for a fixed β. Moreover, to find the solution of (4), the multiplier β can be updated using stochastic sub-gradient method. The updating equations are as follows for $t = 0, 1, \ldots$:

$$J^{t+1}_{\mathrm{p-tr}}(\hat{b}) = (1-\phi_t) J^t_{\mathrm{p-tr}}(\hat{b}) + \phi_t \Big(\max_{r:r\le y^t} \Big\{ f_b(r) + J^t_{\mathrm{p-ad}}(\hat{b}+r) \Big\} - J^t_{\mathrm{p-tr}}(\hat{b}_0) \Big) \quad (16)$$

$$\beta^{t+1} = \Big[\beta^t + \varepsilon_t \Big(c\big(h^t, a^*(b^t, h^t)\big) - C_{\max} \Big) \Big]^+ \quad (17)$$

$$J^{t+1}_{\mathrm{p-ad}}(\bar{b}) = \max_{a:a\le\bar{b}} \Big\{ -f_c(\bar{b}) - \beta^{t+1} c(h^{t+1}, a) + J^{t+1}_{\mathrm{p-tr}}(\bar{b}-a) \Big\} - J^t_{\mathrm{p-ad}}(\bar{b}_0) \quad (18)$$

for $\bar{b}$, $\hat{b} \in \mathcal{B}$. The initial conditions are $J^0_{\mathrm{p-ad}}(\bar{b}) = J^0_{\mathrm{p-tr}}(\hat{b}) = 0$, $\beta^0 > 0$, and the learning sequences satisfy the requirement [6]:

$$\sum_{\tau=0}^{\infty} \phi_\tau = \sum_{\tau=0}^{\infty} \varepsilon_\tau = \infty; \sum_{\tau=0}^{\infty} \phi_\tau^2 + \varepsilon_\tau^2 < \infty; \lim_{\tau\to\infty} \frac{\varepsilon_\tau}{\phi_\tau} = 0.$$

Note that in (16) and (18), we batch-update the state value functions for all possible backlog states $\hat{b}$, $\bar{b} \in \mathcal{B}$, not only the previously visited state. This is possible because the traffic arrival and the channel processes are independent of the queue states. Also, $J^{t+1}_{\mathrm{p-ad}}(\bar{b})$ in (18) needs not to be time-averaged

since time-averaging has been carried out for $J^{t+1}_{\mathrm{p-tr}}$ in the same slot. The iterations (16) and (18) can be viewed as the stochastic estimates of their counterparts (14)–(15) and are updated based on the instantaneous arrival y^t and channel h^t states without requiring known PDFs. The convergence of the proposed learning algorithm is established next.

Theorem 2: The functions $J^t_{\mathrm{p-ad}}(\bar{b})$ and $J^t_{\mathrm{p-tr}}(\hat{b})$ in (16), (18) for $t = 0, 1, \ldots$ are concave decreasing. Moreover, $\lim_{t\to\infty} J^t_{\mathrm{p-ad}}(\bar{b}) = J^*_{\mathrm{p-ad}}(\bar{b})$, $\lim_{t\to\infty} J^t_{\mathrm{p-tr}}(\hat{b}) = J^*_{\mathrm{p-tr}}(\hat{b})$; $\lim_{t\to\infty} \beta^t = \beta^*$ where β^* is the optimal multiplier of (4) and $J^*_{\mathrm{p-tr}}(\hat{b})$, $J^*_{\mathrm{p-ad}}(\bar{b})$ are the value functions of (5) with β being replaced by β^*.

The online learning algorithm does not assume any specific PDFs. Hence, it is very robust to the channel and traffic arrival model variations.

IV. Illustrative Results

A. Simulation setup

We implement the proposed learning algorithms using MATLAB. We assume that the slot duration is equal to $1/W$ where W (Hz) is the bandwidth.

We use the exponential power function derived from the Shannon theoretic rate $c(h, a) = (2^a - 1)/h$ where $h \in \mathcal{H}$ is the power gain.

The channel state space consists of 8 states $\mathcal{H} = \{0.0131, 0.0418, 0.0753, 0.1157, 0.1661, 0.2343, 0.3407, 0.6200\}$ with probabilities $[1, 1, 2, 3, 3, 2, 1, 1]/14$ [7].

We assume (truncated) Poisson arrival process with an average rate 15 (bits) per slot with $y_{\min} = 0$ and $y_{\max} = 30$. The learning rate sequences are chosen as $\phi_t = (1/t)^{.7}$ and $\varepsilon_t = (1/t)^{.85}$. The learning duration is 50000 slots.
To obtain the trade-off curves, we let the functions be $f_b(r) = r$ and $f_c(b) = \kappa b$ for different values of $\kappa \in (0, 1)$.

B. Numerical results

We plot in Fig. 1 the optimal power- queue size trade-off. We can see that without admission control, the queue size $B_{\max}$ is approximately 29 (bits) for $C_{\max} = 6.5$. Also, in order to have queue stability (finite queue size), the power must be greater than $C^\infty \approx 5.8$.

We are now looking at the SAC policies. Fig. 2 plots the optimal trade-off curve achieved by the the proposed online learning algorithm. We also plot the trade-off obtained by the ECCA in [9]. We can observe that for the same queue size, the proposed learning algorithm is able to achieve higher throughput than the ECCA. Alternatively, for the same throughput, the learning algorithm achieves smaller queue size. When the queue size approaches $B_{\max}$ (by setting κ sufficiently small in the learning algorithm), the throughput approaches the average arrival rate, i.e., almost all the arrival is buffered.

Figures 3 demonstrate the convergence of the proposed learning algorithm for some values of κ. It shows the convergence of the Lagrange multiplier updated using stochastic sub-gradient method and of the power consumption. In all cases, the learning algorithm consumes maximum available power $C_{\max}$.

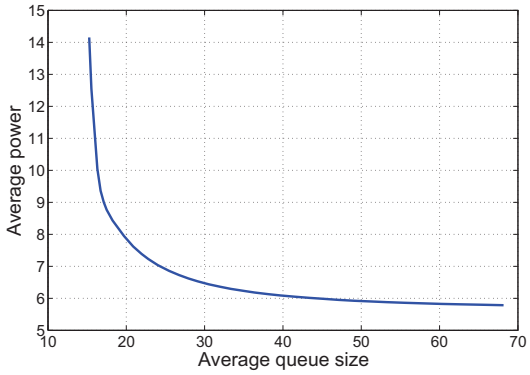

Fig. 1. Optimal power-queue size trade-off for the simulation example.

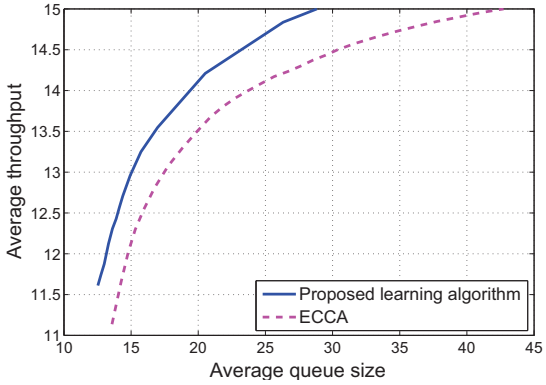

Fig. 2. Throughput-queue size trade-off for $C_{\max} > C^\infty$.

We next demonstrate the use of the proposed learning algorithm to stabilize the queue when $C_{\max} = 4.5 < C^\infty$. Fig. 4 shows the trade-off curves obtained by the proposed algorithm and the ECCA. Again, the proposed algorithm is more efficient in terms of higher throughput for a given average queue size or smaller average queue size for a given throughput. By setting κ small, the queue size increases but finite, ensuring queue stability and at the same time, the throughput is maximized but is strictly less than $\bar{y}$ since traffic has to be dropped.

C. Connection with Lyapunov optimization based approach

We now draw a (simple) connection between the proposed optimal learning and ECCA. For convenience, we rewrite the scheduling action in slot t under optimal learning:

$$a^t = \arg\max_{a: a \le b^t} \left\{ -\beta^t c(h^t, a) + V^t_{\mathrm{p-tr}}(b^t - a; \beta^t) \right\}. \quad (19)$$

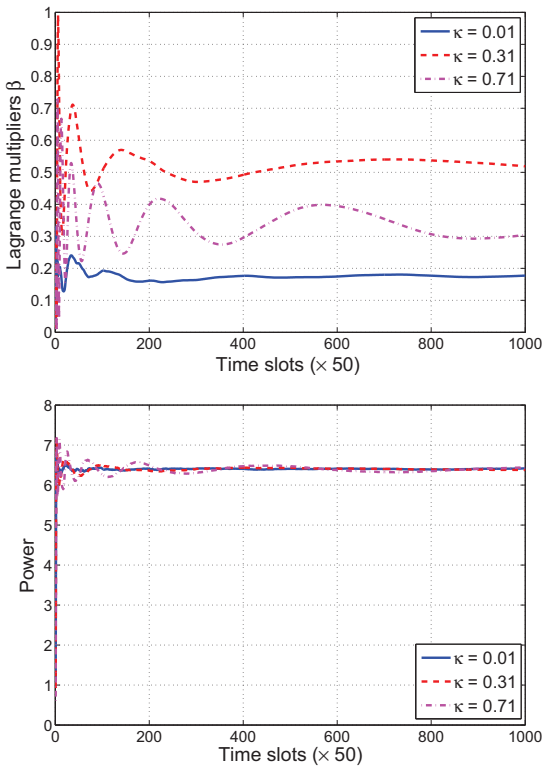

Fig. 3. Convergence of the learning algorithm: Lagrange multiplier and power consumption.

On the other hands, ECCA minimizes the following metric to compute the scheduling action in slot t [9]:

$$a^t_{\text{ECCA}} = \arg\max_{a:a\le b^t} \left\{-q^t c(h^t,a) + b^t a\right\}$$
$$= \arg\max_{a:a\le b^t} \left\{-q^t c(h^t,a) - b^t(b^t-a)\right\} \quad (20)$$

where a term $(b^t)^2$ is added without changing the optimal solution in (20). $\{q^t\}$ is virtual power queue state in slot t and is updated as $q^{t+1} = [q^t - c(h^t, a^t_{\text{ECCA}})]^+ + C_{\max}$. Comparing (19) and (20), we shall have:

$$V^t_{\text{p-tr}}(\hat{b};\beta^t) \approx -\alpha^t b^t \hat{b}$$

where $\alpha^t = \beta^t/q^t$ is some scaling coefficient. Hence, ECCA can be considered as an approximate learning algorithm where the value function $V^t_{\text{p-tr}}(\hat{b};\beta^t)$ is approximated by a linear decreasing function with the slope $-\alpha^t b^t$. Remind that in the optimal learning, $V^t_{\text{p-tr}}$ is concave decreasing. Such approximation has different effects in different traffic loading regions. For example, in the large queue size region, i.e., b^t is large, linear decreasing function is a 'good' approximation of the optimal concave decreasing function. Hence, ECCA performs well in high traffic loading region. However, in the small/medium queue size region, such approximation is coarse, which leads to a worse performance of the ECCA as seen in Fig. 2.

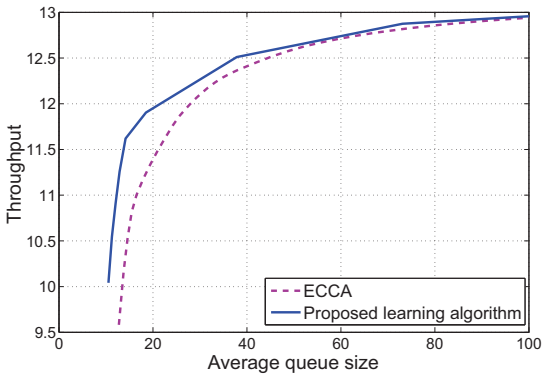

Fig. 4. Throughput-queue size trade-off for $C_{\max} < C^{\infty}$.

V. Conclusions

While existing works have mainly concentrated on the scheduling problem over a fading channel without traffic admission control, this work has studied the joint scheduling-admission control problem. We analyzed the structural properties of the optimal policies. Online learning algorithm for the optimal policies is proposed without requiring a-priori known probability distribution functions of the system dynamics.

References

[1] R. Berry, and R. Gallager, "Communication Over Fading Channels with Delay Constraints," *IEEE Trans. Infor. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[2] M. Goyal, A. Kumar, and V. Sharma, "Optimal Cross-layer Scheduling of Transmissions over a Fading Multiaccess Channel," *IEEE Trans. Infor. Theory*, vol. 54, no. 8, pp. 3518–3536, Aug. 2008.

[3] M. Agarwal, V. 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, July 2008.

[4] D. Djonin, and V. Krishnamurthy, "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, October 2007.

[5] D. Djonin, and V. Krishnamurthy, "Q-Learning Algorithms for Constrained Markov Decision Processes with Randomized Monotone Policies: Applications to MIMO Transmission Control," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2170–2181, May 2007.

[6] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "On-Line Learning Algorithm for Energy Efficient Delay Constrained Scheduling over Fading Channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 732–742, May 2008.

[7] F. Fu, and M. van der Schaar, "Structure-Aware Stochastic Control for Transmission Scheduling," *IEEE Trans. Veh. Tech.*, vol. 61, no. 9, pp. 3931–3945, Nov. 2012.

[8] N. Mastronarde, and M. van der Schaar, "Fast Reinforcement Learning for Energy Efficient Wireless Communications," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6262–6266, Dec. 2011.

[9] L. Georgiadis, M. J. Neely, and L. Tassiulas. *Resource Allocation and Cross-Layer Control in Wireless Networks*. Foundations and Trends in Networking, vol. 1, no. 1, pp. 1–144, 2006.

[10] E. Altman, *Constrained Markov Decision Processes: Stochastic Modeling.* London, UK.: Chapman & Hall CRC, 1999.

[11] V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.