

# Throughput–Delay Tradeoff in Interference-Free Wireless Networks With Guaranteed Energy Efficiency

Yuzhou Li, Min Sheng, *Member, IEEE*, Cheng-Xiang Wang, *Senior Member, IEEE*, Xijun Wang, *Member, IEEE*, Yan Shi, *Member, IEEE*, and Jiandong Li, *Senior Member, IEEE*

**Abstract**—Existing works have addressed the tradeoffs between any two of the three performance metrics: throughput, energy efficiency (EE), and delay. In this paper, we unveil the intertwined relations among these three metrics under a unifying framework and particularly investigate the problem of EE-guaranteed throughput–delay tradeoff in interference-free wireless networks. We first propose two admission control schemes, referred to as the first-out and first-in schemes. We then formulate it as two stochastic optimization problems, aiming at throughput maximization (in the first-out scheme) or dropping rate minimization (in the first-in scheme) subject to requirement of EE (RoE), stability, admission control, and transmit power. To solve the problems, the EE-Guaranteed algorithm for throughput–delay tradeoff (eGuard), respectively called eGuard-I and eGuard-II in the first-out and first-in schemes, is devised. Moreover, with guaranteed RoE, we theoretically show that the eGuard (I and II) can not only push the throughput arbitrarily close to the optimal with tradeoffs in delay but also quantitatively control the throughput–delay performance on demand. Simulation results consolidate the theoretical analysis and particularly show the pros and cons of the two schemes.

**Index Terms**—Throughput, energy efficiency, delay, admission control, power allocation.

## I. INTRODUCTION

### A. Motivation and Related Works

**T**HE dramatic increase in the number of smart phones and tablets with ubiquitous broadband connectivity has triggered an explosive growth in throughput demand to be served by wireless networks [1], [2], which is also called 1000×

Manuscript received January 23, 2014; revised August 6, 2014; accepted November 2, 2014. Date of publication November 13, 2014; date of current version March 6, 2015. This work was supported in part by the National Natural Science Foundation of China under Grants 61231008, 61172079, 61201141, 61301176, and 91338114, the 863 project under Grant 2014AA01A701, the 111 Project under Grant B08038, the Basic Research Fund in Xidian University under Grant 7214497201, and the EU FP7 QUICK project under Grant PIRSES-GA-2013-612652. The associate editor coordinating the review of this paper and approving it for publication was M. C. Vuran. (*Corresponding author: Min Sheng.*)

Y. Li, M. Sheng, X. Wang, Y. Shi, and J. Li are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: msheng@mail.xidian.edu.cn; jdli@mail.xidian.edu.cn; yuzhou\_li@stu.xidian.edu.cn; yshi@xidian.edu.cn; xijunwang@xidian.edu.cn).

C.-X. Wang is with the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, U.K. He is also with School of Information Science and Engineering, Shandong University, Jinan, Shandong, 250100, China (e-mail: cheng-xiang.wang@hw.ac.uk).

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

Digital Object Identifier 10.1109/TWC.2014.2370647

data challenge [3]. From the forecast by Cisco, global mobile data traffic will increase 13-fold between 2012 and 2017 at a compound annual growth rate (CAGR) of 66%, reaching 11.2 exabytes per month by 2017, up from 885 petabytes per month at the end of 2012 [4]. This growth in turn skyrockets the energy consumption and greenhouse gas emission [5]. As a result, it is another thirsty need to improve or guarantee energy efficiency (EE) of systems [6].

Apart from throughput and EE, delay is also a key metric to measure the quality of service (QoS). Thus, it is worthwhile to incorporate throughput, EE, and delay into a theoretical framework, and effectively balance them from a perspective of systematic design. To unveil their intertwined relations, we here provide a skeleton on the relevant researches. Note that throughput in this paper, as in [7], [8], is defined as the maximum admissible traffic rate that the system can stably carry and thus, it to some extent reflects spectral efficiency (SE) of the system. This is because throughput characterizes the data admittance capability of the system, it is thereby related to the transmission ability, i.e., SE. Existing works have extensively studied the EE-SE tradeoff, the power (or energy)-delay tradeoff, the throughput-delay tradeoff, and the EE-delay tradeoff.

1) *EE-SE Tradeoff*: This is one of the main research branches being carried on [9], [10]. Power allocation approaches were adopted in [11], [12] to address the EE-SE tradeoff in interference-free and interference-limited wireless networks, respectively. The authors in [13], [14] jointly considered multi-dimensional resource allocation, such as power allocation, sub-carrier assignment, and antenna selection, to investigate the EE-SE tradeoff in orthogonal frequency division multiple access (OFDMA) networks. The EE-SE tradeoff was also in depth explored in relay-aided and virtual MIMO cellular systems [15]–[17].

As a common feature, [11]–[17] investigated the relevant problems based on the full buffer assumptions and snapshot-based models (i.e., only concerned the performance at the observation time). This indicates that, the stochastic and time-varying features of traffic arrivals were not incorporated into the formulations and thus, all of them did not take admission control into account. In addition, delay was also neglected in these models.

2) *Power (or Energy)-Delay Tradeoff*: Contrary to the static models used in the EE-SE tradeoff, this thread usually goes from the long-term average perspective. By devising network selection algorithms, [18], [19] studied the energy-delay

tradeoff in heterogeneous wireless networks. In [20], a theoretical framework was presented to analyze power-delay tradeoff in systems with adjustable length of transmission and idle time. Dynamic power allocation was adopted in [21] to compromise power and delay. The authors in [22] aimed to find the sleeping control and power matching configurations to achieve the Pareto optimal tradeoff between sum power (or energy) consumption (PC) and average delay. Adaptive user association was exploited in [23] to strike a balance between energy and delay in cellular networks.

However, the results for power (or energy)-delay tradeoff can hardly give insights into EE problems. This is because PC is adopted to measure how small the energy is needed to guarantee users' QoS, while EE to depict how efficiently the energy is consumed [12], [17]. Moreover, the optimality of them cannot be achieved simultaneously [9]. Besides, the application scenarios adopting PC or EE as performance metric are different. In general, PC is suitable for traffics with strict rate requirements such as voice, while EE for delay-tolerant applications such as file transfers. We focus on the latter in this paper. Further, in [18]–[23], admission control is unnecessary and throughput equals to the traffic arrival rate, as they assumed it is inside the network capacity region. Meanwhile, it is also well worth studying how to devise admission control strategies to balance throughput and delay when EE is taken into account for any traffic arrival rate.

3) *Throughput-Delay and EE-Delay Tradeoffs*: With admission control and power allocation taken into account, [7], [8] explored the throughput maximization problems. However, neither of them considered the requirement of EE (RoE) and thus, the RoE of the system or links might not be guaranteed. In fact, improvements in throughput or delay are achieved at the cost of EE, which will be further clarified in Section VI.

Studies on the EE-delay tradeoff have so far been scarce. In [24], the authors showed the loss in EE due to the delay constraints at equilibrium by a game-theoretic approach. Nevertheless, there is no admission control in [24], as it focused on the infinite-backlog case as in [11]–[17]. In addition, throughput was not formulated into the problem as well.

## B. Summary of Results

Motivated by [7], [8], [11]–[24], this paper devotes to achieving flexible tradeoffs among throughput, EE, and delay for any traffic arrival rate whether inside or outside the network capacity region. Specifically, we study the problem of EE-guaranteed throughput-delay tradeoff in interference-free wireless networks.<sup>1</sup> To this end, we first extend the definition of EE presented in the existing works [11]–[17] to depict time-varying and random channel conditions and traffic arrivals. We then propose two admission control schemes, referred to as the first-out scheme and first-in scheme. The first-out scheme first decides whether the newly arrived data is admitted to the system or not. On the contrary, the first-in scheme directly admits all

<sup>1</sup>Note that our idea, formulation, and analysis can be readily extended to interference-limited wireless networks. Specifically, it is only required to replace  $\text{SNR}_m(P(t), G(t))$  in (2) by  $\text{SNR}_m(P(t), G(t), I(t))$ , where  $\text{SNR}_m(P(t), G(t), I(t))$  denotes the signal-to-interference-plus-noise-ratio (SINR) at receiver  $m$ .

the arrived data to the system and drops the queued data only when the system cannot bear so much data.

In the following, under the constraints of RoE, stability, admission control, and power allocation, two stochastic optimization formulations are presented. The formulation for the first-out scheme maximizes throughput and the one for the first-in scheme minimizes dropping rate. We then propose two algorithms, named as the EE-Guaranteed algorithm for throughput-delay tradeoff (eGuard-I) for the first-out scheme and the eGuard-II for the first-in scheme. Each of the proposed algorithms can be further decomposed into two sub-algorithms, which explicitly consist of admission control and power allocation. Theoretical analysis shows that the eGuard (I and II) can achieve the optimal throughput at the cost of delay performance. Anyhow, the RoE is guaranteed.

The main contributions of this work are twofold.

- We put forward two formulations to quantitatively strike a balance between throughput and average delay, meanwhile guarantee the RoE of the system. Only by adjusting a control parameter, the proposed algorithms provide a simple method to control the throughput-delay performance on demand, which in turn adaptively affects admission control and power allocation.
- We propose two admission control schemes, namely the first-out and first-in schemes. In comparison, the first-out scheme has relatively poor performance in throughput and average delay given the RoE, but it benefits from the compatibility with transmission control protocol (TCP) for different intensities of traffic loads. On the contrary, the first-in scheme is preferred to work in the light or middle traffic states, as it is incompatible with TCP.

## C. Paper Organization

In Section II, we present the concerned scenario, introduce the proposed admission control schemes, referred to as the first-out scheme and first-in scheme, and give the formal definitions of stability and EE. We then formulate the EE-guaranteed throughput-delay tradeoff problem in the first-out scheme and devise the eGuard-I to solve it in Section III. Section IV focuses on the EE-guaranteed throughput-delay tradeoff problem in the first-in scheme and the eGuard-II is proposed. We analyze the performance of the proposed algorithms in Section V. Numerical results and performance discussions are presented in Section VI. Finally, we conclude our paper in Section VII.

## II. ADMISSION CONTROL SCHEMES AND PRELIMINARY

In this section, we begin with describing the concerned system scenario, followed by introducing the proposed admission control schemes. We then formally give the definitions of stability and EE.

### A. Description of the Concerned Scenario

We consider one-hop interference-free wireless networks that operate in discrete time with normalized time slots  $t \in \{0, 1, 2, \dots\}$ . There are  $N$  transmitters and  $M$  receivers, both

of which are equipped with a single antenna. The channels assigned to receivers are orthogonal and thus, there is no interference among them. Typical scenarios include single-cell uplink or downlink cellular communications and distributed antenna systems (DAS) [2], [11], [25]–[27].

Data arrives randomly every slot and are queued separately for transmission to each receiver. Let  $A(t) = (A_m(t))$  be the process of random data arrivals, where  $A_m(t)$  is the amount of new data that arrives to receiver  $m$  on slot  $t$ . For simplicity, we assume that  $A(t)$  is independent and identically distributed (i.i.d.) over slots with arrival rate  $\lambda$ , so that  $\mathbb{E}\{A(t)\} = \lambda$  for all  $t$ . In the paper, we do not specify traffic models and their variances, as neither of them affects the formulations and analysis. Besides,  $Q(t) = (Q_m(t))$  denotes the number of data currently stored in each of the  $M$  queues.

We denote Link conditions by  $G(t) = (g_{n,m}(t))$ , where  $g_{n,m}(t)$  represents the link condition from transmitter  $n$  to user  $m$  at slot  $t$ , which accounts for path loss, shadowing, and fading. We assume that  $G(t)$  can be accurately known at the transmitters and is i.i.d. over time slots with probabilities  $\pi_G$  in a finite (but arbitrarily large) state space  $\mathbb{G}$ . In addition, it keeps constant for the duration of a time slot, but potentially changes on slot boundaries. Besides, power matrix is denoted by  $P(t) = (P_{n,m}(t))$ , with  $P_{n,m}(t)$  being the transmit power from transmitter  $n$  to receiver  $m$ .

The transmit rate to receiver  $m$  is given by the following general rate-power function

$$R_m(t) = R_m(P(t), G(t)). \quad (1)$$

A typical example of continuous rate-power functions is the Shannon formula, i.e.,

$$R_m(P(t), G(t)) = \log_2(1 + \text{SNR}_m(P(t), G(t))) \quad (2)$$

where  $\text{SNR}_m(P(t), G(t))$  is the signal-to-noise-ratio (SNR) at receiver  $m$ . Assuming that the maximum ratio combining (MRC) is adopted at the receiver [11], [28], then the SNR is expressed as

$$\text{SNR}_m(P(t), G(t)) = \frac{\sum_{n=1}^N P_{n,m}(t) g_{n,m}(t)}{\sigma^2} \quad (3)$$

where  $\sigma^2$  denotes the noise power. It is worthwhile to note that the rate-power function can also take values in a predefined discrete modulation and coding set (MCS), which is designed to achieve a target bit error rate (BER) [28].

As in [11]–[13], [29], [30], we model the PC of transmitter  $n$  as

$$\text{PC}_n(t) = \xi_n \sum_{m=1}^M P_{n,m}(t) + P_n^C \quad (4)$$

where  $\xi_n$  and  $P_n^C$  are constants accounting for the inefficiency of the power amplifier and the circuit power of transmitter  $n$ , respectively.

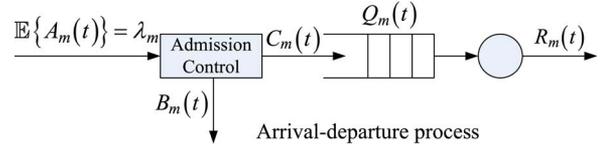


Fig. 1. An intuitional illustration of the first-out scheme.

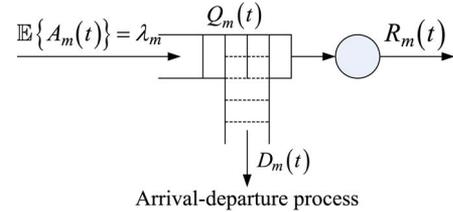


Fig. 2. An intuitional illustration of the first-in scheme.

Accordingly, the sum transmit rate and the total PC of the system are respectively given by

$$R_{\text{tot}}(t) = R_{\text{tot}}(P(t), G(t)) = \sum_{m=1}^M R_m(t) \quad (5)$$

$$P_{\text{tot}}(t) = P_{\text{tot}}(P(t), G(t)) = \sum_{n=1}^N \text{PC}_n(t) \quad (6)$$

## B. Admission Control

To explore the maximum amount of admissible traffics the system can stably carry, which is called throughput in this paper, it is required to devise effective admission control schemes to adjust the traffic rate injected into the system. To this end, we propose two schemes, referred to as the first-out scheme and first-in scheme.

1) *The First-out Scheme:* As illustrated in Fig. 1, the controller in this scheme first makes decisions on whether the arrived data is admitted to the system or not over slots. If not, it directly drops the data.

Let  $C_m(t)$  and  $B_m(t)$  respectively represent the amount of data admitted to queue  $m$  and dropped at queue  $m$  at slot  $t$ , we thus have  $B_m(t) = A_m(t) - C_m(t)$ . Then, the actual data queue  $Q_m(t)$  is updated as

$$Q_m(t+1) = \max[Q_m(t) - R_m(t), 0] + C_m(t). \quad (7)$$

2) *The First-in Scheme:* The controller in this scheme directly admits all the arrived data to the system. It takes actions to drop some of the queued data only when the system cannot bear so much data. Fig. 2 depicts this process.

We denote the amount of dropped data at slot  $t$  by  $D_m(t)$ , then the actual data queue  $Q_m(t)$  evolves according to

$$Q_m(t+1) = \max[Q_m(t) - R_m(t) - D_m(t), 0] + A_m(t). \quad (8)$$

## C. Definitions of Stability and EE

To maximize throughput, it is necessary to increase  $C_m(t)$  as large as possible in the first-out scheme or decrease  $D_m(t)$  as small as possible in the first-in scheme. However, excessive traffic injection certainly leads to a large queue length

in the buffer under the constraints of the limited resources (e.g., limited transmit rate  $R_m(t)$ ) and thus, resulting in large unacceptable delay and poor user experiences. This shows that throughput and delay are two conflict metrics and thus, it is necessary to effectively balance them. To quantitatively reveal such a tradeoff, we require the queue to be stable.

*Definition 1:* A discrete time queue process  $U(t)$  is mean rate stable [31] if

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}\{|U(t)|\}}{t} = 0. \quad (9)$$

*Definition 2:* A network of queues is stable if all individual queues are stable.

However, a good performance in throughput or delay does not mean a good EE, as there are tradeoffs between EE and SE [11]–[17] and between EE and delay [24], [32]. Thus, it is necessary to guarantee the RoE of the system or links when dynamically balancing throughput and delay. To this end, we give the definition of EE.

*Definition 3:* EE  $\eta_{EE}$  of networks is defined as the ratio of the long-term aggregate data delivered to the corresponding long-term total power consumption in unit of bit/Hz/Joule. Specifically,  $\eta_{EE}$  is given by

$$\begin{aligned} \eta_{EE} &= \lim_{t \rightarrow \infty} \frac{\sum_{\tau=0}^{t-1} \mathbb{E}\{R_{Tot}(P(\tau), G(\tau))\}}{\sum_{\tau=0}^{t-1} \mathbb{E}\{P_{Tot}(P(\tau), G(\tau))\}} \\ &= \frac{\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{R_{Tot}(P(\tau), G(\tau))\}}{\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{P_{Tot}(P(\tau), G(\tau))\}} = \frac{\bar{R}_{Tot}}{\bar{P}_{Tot}}. \end{aligned} \quad (10)$$

Here, we assume temporarily that the limits are well defined.

*Remark 1:* It is quite different between (10) and the classical EE definitions in [11]–[17], which are defined as  $\eta_{EE}(t) = R_{Tot}(t)/P_{Tot}(t)$ . As can be seen, they are snapshot-based, i.e., only concern EE at the observation time. However, defined from the long-term perspective, (10) thereby implies the concept of time diversity (this will be further clear in Section VI-E). It is thus suitable to explore throughput, as both channel conditions and traffic arrivals are time-variant. Further, (10) can be also seen as an extension of those given in [11]–[17], because it degenerates to them if there are no time averages and expectations in (10).

Similarly, the time averages of throughput and dropping rate of receiver  $m$  are respectively defined as

$$\bar{C}_m = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{C_m(\tau)\}, \quad \bar{D}_m = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{D_m(\tau)\}. \quad (11)$$

Therefore, the system throughputs for the first-out and first-in schemes are individually given by

$$\text{Thr}^{\text{out}} = \sum_{m=1}^M \bar{C}_m \quad (12)$$

$$\text{Thr}^{\text{in}} = \sum_{m=1}^M \lambda_m - \sum_{m=1}^M \bar{D}_m. \quad (13)$$

### III. THE FIRST-OUT SCHEME

In this section, we consider the first-out scheme. We first formulate the problem, and then present the eGuard-I to solve the proposed formulation.

#### A. Problem Formulation

In the first-out scheme, we formulate the EE-guaranteed throughput-delay tradeoff problem as the following stochastic optimization problem

$$\begin{aligned} \max \quad & \text{Thr} = \sum_{m=1}^M w_m \bar{C}_m \\ \text{s.t.} \quad & \text{C1: } \eta_{EE} \geq \eta_{EE}^{\text{req}} \\ & \text{C2: Queues } Q_m(t) \text{ are mean rate stable, } \forall m, t \\ & \text{C3: } 0 \leq C_m(t) \leq A_m(t), \forall m, t \\ & \text{C4: } R_m(t) \leq Q_m(t), \forall m, t \\ & \text{C5: } \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\text{max}}, \forall n, t \\ & \text{C6: } P_{n,m}(t) \geq 0, \forall n, m, t. \end{aligned} \quad (14)$$

In (14),  $w_m$  denotes the weight of each receiver,  $\eta_{EE}^{\text{req}}$  is the RoE of the system, and  $P_n^{\text{max}}$  is the maximum transmit power of transmitter  $n$ . The objective is to maximize the sum weighted throughput Thr. C1 ensures the RoE. C2 is the network stability constraint to guarantee a finite queue length for each queue. C3 is the admission control constraint. C4 is the energy-saving constraint. C5 and C6 are the peak and nonnegative transmit power constraints.

*Remark 2:* 1) The queue stability constraints are used to depict and control the average delay. Definition 1 shows that network stability is guaranteed if the average queue length is finite [31], i.e.,  $\bar{U} = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{U(\tau)\} < \infty$ . Further, it is required to control the size of all queues to stabilize the network from Definition 2. Note that average delay is proportional to average queue length for a given traffic arrival rate from Little's Theorem [33] (i.e., average delay = average queue length/traffic arrival rate). As a result, the average queue length bridges stability and delay. Thus, we can address the average delay by queue size and further by queue stability. 2) Formulation (4) is suitable for non-real-time (i.e., delay-tolerant) systems, which also holds for (21). This is because, instead of guaranteeing deterministic delay requirements, both of them can quantitatively adjust the average delay from the later analysis in Section V.

#### B. The Virtual EE Queue

To handle the RoE constraint C1 in (14), we introduce the concept of the virtual EE queue. Specifically, we define it as  $Z(t)$  with  $Z(0) = 0$  and the updated equation as

$$Z(t+1) = \max[Z(t) + y(t), 0] \quad (15)$$

where  $y(t) = \eta_{EE}^{\text{req}} P_{Tot}(t) - R_{Tot}(t)$ .

*Lemma 1:* If an admission control and power allocation policy stabilizes the virtual EE queue  $Z(t)$ , then C1 is satisfied.

*Proof:* See Appendix A.  $\blacksquare$

*Remark 3:* Lemma 1 shows that we can transform the original problem (14) into a problem of minimizing the time average objective subject to queue stability (both the actual data queue  $Q_m(t)$  and the virtual EE queue  $Z(t)$ ) and C3–C6.

### C. EE-Guaranteed Algorithm for Throughput-Delay Tradeoff (eGuard-I)

At this point, we can exploit the drift-plus-penalty algorithm [31], which is developed to solve stochastic optimization models by Lyapunov optimization technique, to tackle (14). We refer it to as the eGuard-I in the paper, shown in Algorithm 1. Note that  $V$  used in the eGuard-I represents an arbitrary positive control parameter, and an explanation for its implication will be given in Section V.

---

#### Algorithm 1 EE-Guaranteed Algorithm for throughput-delay Tradeoff (eGuard-I).

---

1: At each slot  $t$ , observe the queue states  $Q_m(t)$ ,  $Z(t)$ , and the channel condition  $G(t)$ .

2: Make the admission control and power allocation policy according to

$$\begin{aligned} \min \quad & \sum_{m=1}^M [Q_m(t) - w_m V] C_m(t) + Z(t) \eta_{EE}^{\text{req}} \sum_{n=1}^N \text{PC}_n(t) \\ & - \sum_{m=1}^M [Q_m(t) + Z(t)] R_m(t) \\ \text{s.t.} \quad & \text{C3: } 0 \leq C_m(t) \leq A_m(t), \forall m, t \\ & \text{C4: } R_m(t) \leq Q_m(t), \forall m, t \\ & \text{C5: } \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\text{max}}, \forall n, t \\ & \text{C6: } P_{n,m}(t) \geq 0, \forall n, m, t. \end{aligned} \quad (16)$$

3: Update  $Q_m(t)$  and  $Z(t)$  respectively according to (7) and (15) based on the above obtained solutions.

---

As can be seen from (16), there is no terms coupling  $C_m(t)$  and  $P_m(t)$  in the objective and constraints. Thus, we can equivalently decompose (16) into the following two subproblems and solve them separately.

1) Dynamic admission control:

$$\begin{aligned} \min \quad & \sum_{m=1}^M [Q_m(t) - w_m V] C_m(t) \\ \text{s.t.} \quad & \text{C3: } 0 \leq C_m(t) \leq A_m(t), \forall m, t. \end{aligned} \quad (17)$$

Further, the solution of (17) has the following simple on-off structure:

$$C_m(t) = \begin{cases} A_m(t), & \text{if } Q_m(t) \leq w_m V \\ 0, & \text{Otherwise.} \end{cases} \quad (18)$$

2) Dynamic power allocation:

$$\begin{aligned} \min \quad & Z(t) \eta_{EE}^{\text{req}} \sum_{n=1}^N \text{PC}_n(t) - \sum_{m=1}^M [Q_m(t) + Z(t)] R_m(t) \\ \text{s.t.} \quad & \text{C4: } R_m(t) \leq Q_m(t), \forall m, t \\ & \text{C5: } \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\text{max}}, \forall n, t \\ & \text{C6: } P_{n,m}(t) \geq 0, \forall n, m, t. \end{aligned} \quad (19)$$

From (2), C4 in (19) can be equivalently transformed to

$$\text{C4}^*: 1 + \frac{\sum_{n=1}^N P_{n,m}(t) g_{n,m}(t)}{\sigma^2} \leq 2Q_m(t) \quad (20)$$

which is a linear convex constraint.

After replacing C4 by C4\*, (19) becomes a standard convex optimization problem [34] and thus, it can be effectively and optimally solved by off-the-shelf solvers, such as CVX [35]. Attributed to the optimality in each slot, the eGuard-I can achieve an asymptotically optimal throughput. Specifically, the eGuard-I can attain a throughput that arbitrarily approaches the theoretical optimum of (14) by choosing a large enough  $V$ . However, throughput improves at the cost of the average delay, and their tradeoff can further be explicitly adjusted by the eGuard-I. We will discuss the details on how to compute the drift-plus-penalty of (14) and how the performance of the eGuard-I is respectively in Lemma 2 and Theorem 2 in Section V, and verify the claims by Figs. 4 and 5 in Section VI-B.

*Remark 4:* 1) The eGuard-I is an online algorithm, as it requires none of prior-knowledge of traffic arrival rates  $\lambda$  and channel statistics  $\pi_G$ . 2) The eGuard-I can be easily applied to practical applications. As can be known from (16), it is only required for the controller to know  $G(t)$ ,  $Z(t)$ , and  $Q_m(t)$  to determine  $C_m(t)$  and  $P_m(t)$  in each slot. This can be done by simply updating and saving  $Z(t)$  and  $Q_m(t)$  at the controller.

## IV. THE FIRST-IN SCHEME

In this section, we apply the obtained results in the first-out scheme to address the related problem in the first-in scheme.

### A. Problem Formulation

In the first-in scheme, to explore the maximum throughput of the system, i.e., (13), we need to minimize the amount of dropped data. To this end, the EE-guaranteed throughput-delay tradeoff problem is modeled as

$$\begin{aligned} \min \quad & \text{Dro} = \sum_{m=1}^M w_m \bar{D}_m \\ \text{s.t.} \quad & \text{C1}' : \eta_{EE} \geq \eta_{EE}^{\text{req}} \\ & \text{C2}' : \text{Queues } Q_m(t) \text{ are mean rate stable}, \forall m, t \\ & \text{C3}' : D_m(t) + R_m(t) \leq Q_m(t), \forall m, t \\ & \text{C4}' : \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\text{max}}, \forall n, t \\ & \text{C5}' : P_{n,m}(t) \geq 0, \forall n, m, t \\ & \text{C6}' : D_m(t) \geq 0, \forall m, t. \end{aligned} \quad (21)$$

In (21),  $Dro$  is the sum weighted dropping rate and  $C3'$  is used to confine  $D_m(t)$  and  $R_m(t)$  (see Fig. 2).

### B. Algorithm Design

Similar to the eGuard-I,<sup>2</sup> the eGuard-II for (21) is given in Algorithm 2.

---

#### Algorithm 2 EE-Guaranteed Algorithm for Throughput-Delay tRaDeoff (eGuard-II)

---

1: At each slot  $t$ , observe the queue states  $Q_m(t)$ ,  $Z(t)$ , and the channel condition  $G(t)$ .

2: Make the admission control and power allocation policy according to

$$\begin{aligned}
 \min \quad & \sum_{m=1}^M [w_m V - Q_m(t)] D_m(t) + Z(t) \eta_{EE}^{\text{req}} \sum_{n=1}^N PC_n(t) \\
 & - \sum_{m=1}^M [Q_m(t) + Z(t)] R_m(t) \\
 \text{s.t.} \quad & C3' : D_m(t) + R_m(t) \leq Q_m(t), \forall m, t \\
 & C4' : \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\max}, \forall n, t \\
 & C5' : P_{n,m}(t) \geq 0, \forall n, m, t \\
 & C6' : D_m(t) \geq 0, \forall m, t. \tag{22}
 \end{aligned}$$

3: Update  $Q_m(t)$  and  $Z(t)$  respectively according to (8) and (15) based on the above obtained solutions.

---

To effectively solve (22), we introduce a new optimization variable  $X_m(t)$  to replace  $R_m(t)$  and then recast (22) as

$$\begin{aligned}
 \min \quad & \sum_{m=1}^M [w_m V - Q_m(t)] D_m(t) + Z(t) \eta_{EE}^{\text{req}} \sum_{n=1}^N PC_n(t) \\
 & - \sum_{m=1}^M [Q_m(t) + Z(t)] X_m(t) \\
 \text{s.t.} \quad & C3' : D_m(t) + X_m(t) \leq Q_m(t), \forall m, t \\
 & C4' : \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\max}, \forall n, t \\
 & C5' : P_{n,m}(t) \geq 0, \forall n, m, t \\
 & C6' : D_m(t) \geq 0, \forall m, t \\
 & C7' : R_m(t) \geq X_m(t), \forall m, t \\
 & C8' : X_m(t) \geq 0, \forall m, t. \tag{23}
 \end{aligned}$$

Note that the inequality constraint  $C7'$  will be satisfied with equality at optimality. At this point, (23) is a standard convex

<sup>2</sup>If we transform the objective function of (21) to  $\max -\sum_{m=1}^M w_m \bar{D}_m$ , then (21) takes the similar form to (14)

optimization problem [34] and thus, it can be optimally solved as (19).

Furthermore, by comparing (22) with (16), we can find that  $C3'$  in (22) couples  $D_m(t)$  and  $R_m(t)$  and thus, it is difficult to devise explicitly independent (or separate) admission control and power allocation algorithms as in the eGuard-I. However, (maybe) sub-optimal decomposition algorithms can be constructed as follows.

- 1) Dynamic power allocation: Power allocation for the eGuard-II takes the same form as (19).
- 2) Dynamic admission control At any slot  $t$ , transmitters first deliver data to users, say  $m$ , at the speed of  $R_m(t)$  determined by the transmit power  $P(t)$  obtained from (19), and then the controller executes admission control as

$$D_m(t) = \begin{cases} Q_m(t) - R_m(t), & \text{if } w_m V \leq Q_m(t) \\ 0, & \text{Otherwise.} \end{cases} \tag{24}$$

## V. PERFORMANCE ANALYSIS

To obtain the performance of the eGuard,<sup>3</sup> we first provide some necessary but practical boundedness assumptions in this section. We then formally derive the performance of the eGuard leveraging on the Lyapunov optimization technique.

### A. Boundedness Assumptions

Let  $A_{G(t)}$  represent the set of all admission control and power allocation options available under a given  $G(t)$ . For any given channel condition  $G(t)$ , any actual queue  $Q_m(t)$ , any virtual queue  $Z(t)$ , and  $\mathcal{R}(t) \in A_{G(t)}$ , we assume that

$$\mathbb{E} \{y(t)^2\} \leq \theta \tag{25}$$

$$\mathbb{E} \{R_m(t)^2\} \leq \theta, \forall m \tag{26}$$

$$\mathbb{E} \{C_m(t)^2\} \leq \theta, \forall m \tag{27}$$

$$\text{Thr}^{\min} \leq \mathbb{E} \{\text{Thr}(t)\} \leq \text{Thr}^{\max} \tag{28}$$

where  $\theta$ ,  $\text{Thr}^{\min}$ , and  $\text{Thr}^{\max}$  are some finite constants.

*Note 1:* The assumptions (25)–(28) are very reasonable since admissible rates, departure rates, and values of power allocation are bounded in realistic systems. Moreover, we do not use the specific values of  $\theta$ ,  $\text{Thr}^{\min}$ , and  $\text{Thr}^{\max}$  in the following analysis.

<sup>3</sup>In this paper, we take the eGuard-I as an example to analyze the performance for brevity. However, the method for performance analysis can be easily applied to the case of the eGuard-II. More specifically, only two changes need to be modified. First, change (27) and (28) to  $\mathbb{E} \{D_m(t)^2\} \leq \theta$  and  $\text{Dro}^{\min} \leq \mathbb{E} \{\text{Dro}(t)\} \leq \text{Dro}^{\max}$  ( $\text{Dro}^{\min}$  and  $\text{Dro}^{\max}$  are some finite constants.), respectively. Second, change  $\min \text{Dro} = \sum_{m=1}^M w_m \bar{D}_m$  in (21) to  $\max -\text{Dro} = -\sum_{m=1}^M w_m \bar{D}_m$ . We then can totally follow the method in the eGuard-I to analyze the performance of the eGuard-II by setting  $\text{Dro}'(t) = -\text{Dro}(t)$ , as  $\text{Dro}'(t)$  plays the same role as  $\text{Thr}(t)$ .

### B. Lyapunov Drift and Optimization

Let  $\Theta(t) = [Q(t), Z(t)]$ . The Lyapunov function is constructed as

$$L(\Theta(t)) = \frac{1}{2} \sum_{m=1}^M Q_m(t)^2 + \frac{1}{2} Z(t)^2. \quad (29)$$

Then the one-slot conditional Lyapunov drift  $\Delta(\Theta(t))$  is given by [31]

$$\Delta(\Theta(t)) = \mathbb{E}\{L(\Theta(t+1)) - L(\Theta(t)) | \Theta(t)\}. \quad (30)$$

*Lemma 2:* Suppose  $G(t)$  is i.i.d. over slots. Under any admission control and power allocation algorithm, all parameters  $V \geq 0$ , and all possible values of  $\Theta(t)$ , the upper bound of the drift-plus-penalty of (14), denoted by  $\Delta(\Theta(t)) - V\mathbb{E}\{Thr(t) | \Theta(t)\}$ , is given by

$$\begin{aligned} & \Delta(\Theta(t)) - V\mathbb{E}\{Thr(t) | \Theta(t)\} \\ & \leq B - V\mathbb{E}\{Thr(t) | \Theta(t)\} \\ & \quad + Z(t)\mathbb{E}\{\eta_{EE}^{req} P_{tot}(t) - R_{tot}(t) | \Theta(t)\} \\ & \quad + \sum_{m=1}^M Q_m(t)\mathbb{E}\{C_m(t) - R_m(t) | \Theta(t)\} \end{aligned} \quad (31)$$

where  $Thr(t) = \sum_{m=1}^M w_m C_m(t)$ , and  $B$  is a positive constant (according to (25)–(28)) and for all  $t$  satisfies

$$B \geq \frac{1}{2} \sum_{m=1}^M \mathbb{E}\{C_m(t)^2 + R_m(t)^2 | \Theta(t)\} + \frac{1}{2} \mathbb{E}\{y(t)^2 | \Theta(t)\}. \quad (32)$$

*Proof:* See Appendix B.  $\blacksquare$

As questions left in Section III-C, the design principle behind the eGuard-I and the implication of the control parameter  $V$  become clear hereto.

- By plugging (5) and (6) into (31), we can obtain that the eGuard-I (16) is devised to minimize the right-hand-side (RHS) of (31) subject to C3–C6,<sup>4</sup> i.e., minimize the upper bound of the drift-plus-penalty of (14).
- We can obtain from (31) that  $V$  plays the role to balance throughput and queue length (i.e., delay), i.e.,  $V$  represents a design knob of throughput and delay tradeoff. More specifically, to minimize the RHS, the controller tries to increase throughput  $Thr(t)$  as large as possible when a large  $V$  is adopted by the system. That is to say, a large  $V$  implies a preference on throughput. On the contrary, it is desired for the controller to decrease queue size  $Q_m(t)$  as small as possible to minimize the RHS for a small  $V$ . Thus, a small  $V$  reflects the priority on queue size (i.e., delay).

So far, we have shown why the eGuard-I runs as (16). To evaluate its performance, we need the following lemma, which will be used in Theorem 1.

<sup>4</sup>By the principle of opportunistically minimizing an expectation [31], minimizing  $f(t)$  ensures that  $E\{f(t) | \Theta(t)\}$  is minimized.

*Lemma 3:* Suppose (14) is feasible, i.e., there exists at least an admission control and power allocation solution to satisfy C1–C6, and the boundedness assumptions (25)–(28) hold. Then, there exists a stationary randomized algorithm  $\mathcal{R}^*(t)$  for any  $\varphi > 0$  that satisfies

$$\mathbb{E}\{C_m^*(t) - R_m^*(t) | \Theta(t)\} = \mathbb{E}\{C_m^*(t) - R_m^*(t)\} \leq -\varepsilon, \forall m \quad (33)$$

$$\mathbb{E}\{y^*(t) | \Theta(t)\} = \mathbb{E}\{y^*(t)\} \leq \varphi, \forall n \quad (34)$$

$$\mathbb{E}\{Thr^*(t) | \Theta(t)\} = \mathbb{E}\{Thr^*(t)\} \geq Thr^{opt} - \varphi \quad (35)$$

where  $\varepsilon$  is an arbitrarily small positive number,  $Thr^*(t)$ ,  $y^*(t)$ ,  $C_m^*(t)$ , and  $R_m^*(t)$  are the resulting values under  $\mathcal{R}^*(t)$ , and  $Thr^{opt}$  is the theoretical optimum of (14).

*Proof:* Similar proof can be found in [31].  $\blacksquare$

By exploiting Lemmas 1, 2, and 3, we are now ready to address the only remaining question left in Section III-C, that is, how the performance of the eGuard-I is.

*Theorem 1:* Suppose that  $G(t)$  is i.i.d. over slots with probabilities  $\pi_G$ , the problem (14) is feasible, and  $\mathbb{E}\{L(\Theta(0))\} < \infty$ . Then, the following properties hold for the eGuard-I with any  $V > 0$ .

- All queues  $Q_m(t)$  and  $Z(t)$  are mean rate stable. Hence, the RoE constraint C1 is guaranteed from Lemma 1.
- The performance bound of the weighted throughput satisfies

$$Thr \geq Thr^{opt} - \frac{B}{V}. \quad (36)$$

- The average queue length  $\bar{Q}$  has the following performance bound

$$\begin{aligned} \bar{Q} &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=0}^{K-1} \left[ \sum_{m=1}^M \mathbb{E}\{Q_m(t)\} \right] \\ &\leq \frac{B + V(Thr^{\max} - Thr^{opt})}{\varepsilon}. \end{aligned} \quad (37)$$

*Proof:* Because the eGuard-I minimizes the RHS of (31) over C3–C6, we have

$$\begin{aligned} & \Delta(\Theta(t)) - V\mathbb{E}\{Thr(t) | \Theta(t)\} \leq B \\ & \quad + \sum_{m=1}^M Q_m(t)\mathbb{E}\{C_m^*(t) - R_m^*(t) | \Theta(t)\} \\ & \quad + Z(t)\mathbb{E}\{y^*(t) | \Theta(t)\} - V\mathbb{E}\{Thr^*(t) | \Theta(t)\} \end{aligned} \quad (38)$$

where  $C_m^*(t)$ ,  $R_m^*(t)$ ,  $y^*(t)$ , and  $Thr^*(t)$  are the resulting values under any alternative (possibly stationary randomized) strategy  $\mathcal{R}^*(t)$ .

Plugging (33), (34), and (35) into (38) and taking a limit as  $\varphi \rightarrow 0$  yield

$$\Delta(\Theta(t)) - V\mathbb{E}\{Thr(t) | \Theta(t)\} \leq B - VThr^{opt} - \varepsilon \sum_{m=1}^M Q_m(t). \quad (39)$$

By taking iterated expectation and using telescoping sums [31] over  $t \in \{0, 1, \dots, K-1\}$  in the above inequality, we get

$$\begin{aligned} & \mathbb{E}\{L(\Theta(K))\} - \mathbb{E}\{L(\Theta(0))\} - V \sum_{t=0}^{K-1} \mathbb{E}\{\text{Thr}(t)\} \\ & \leq K(B - V\text{Thr}^{\text{opt}}) - \varepsilon \sum_{t=0}^{K-1} \left[ \sum_{m=1}^M \mathbb{E}\{Q_m(t)\} \right]. \end{aligned} \quad (40)$$

a) Rearranging (40) and using the condition (28) and the fact that  $Q_m(t) \geq 0$ , we obtain

$$\begin{aligned} \mathbb{E}\{Z(K)^2\} & \leq 2K(B - V\text{Thr}^{\text{opt}}) \\ & \quad - 2VK\text{Thr}^{\text{min}} + 2\mathbb{E}\{L(\Theta(0))\}. \end{aligned} \quad (41)$$

Due to the fact that the variance of  $|Z(K)|$  can not be negative, we have  $\mathbb{E}\{|Z(K)|\}^2 \leq \mathbb{E}\{Z(K)^2\}$ . Thus, we have

$$\begin{aligned} & \mathbb{E}\{|Z(K)|\} \\ & \leq \sqrt{2K(B - V\text{Thr}^{\text{opt}}) - 2VK\text{Thr}^{\text{min}} + 2\mathbb{E}\{L(\Theta(0))\}}. \end{aligned} \quad (42)$$

Dividing (42) by  $K$  and taking a limit as  $K \rightarrow \infty$  results in

$$\lim_{K \rightarrow \infty} \frac{\mathbb{E}\{|Z(K)|\}}{K} = 0. \quad (43)$$

Hence,  $Z(t)$  is mean rate stable from Definition 1 and thus, C1 is satisfied based on Lemma 1. Similar proof can be applied to  $Q_m(t)$ .

b) Dividing (40) by  $VK$ , rearranging terms, and using the fact that  $\mathbb{E}\{L(\Theta(K))\} \geq 0$  and  $Q_m(t) \geq 0$  yields

$$\frac{1}{K} \sum_{t=0}^{K-1} \mathbb{E}\{\text{Thr}(t)\} \geq \text{Thr}^{\text{opt}} - \frac{B}{V} - \frac{\mathbb{E}\{L(\Theta(0))\}}{VK}. \quad (44)$$

Taking a limit as  $K \rightarrow \infty$  proves part (b).

c) Similarly, we can rewrite (40) as

$$\begin{aligned} \varepsilon \sum_{t=0}^{K-1} \left[ \sum_{m=1}^M \mathbb{E}\{Q_m(t)\} \right] & \leq K(B - V\text{Thr}^{\text{opt}}) \\ & \quad + V \sum_{t=0}^{K-1} \mathbb{E}\{\text{Thr}(t)\} + \mathbb{E}\{L(\Theta(0))\}. \end{aligned} \quad (45)$$

Dividing (45) by  $\varepsilon K$  and taking a limit as  $K \rightarrow \infty$ , we have

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=0}^{K-1} \left[ \sum_{m=1}^M \mathbb{E}\{Q_m(t)\} \right] \leq \frac{B + V(\text{Thr}^{\text{max}} - \text{Thr}^{\text{opt}})}{\varepsilon} \quad (46)$$

where condition (28) is exploited.  $\blacksquare$

To readily understand the obtained results presented in Theorem 1, we further provide some intuitional observations.

- Eq. (36) shows  $\text{Thr} \geq \text{Thr}^{\text{opt}} - \frac{B}{V}$ . Besides,  $\text{Thr} \leq \text{Thr}^{\text{opt}}$ . Hence,  $\text{Thr}^{\text{opt}} - \frac{B}{V} \leq \text{Thr} \leq \text{Thr}^{\text{opt}}$ . This indicates that  $\text{Thr}$  can be arbitrarily close to  $\text{Thr}^{\text{opt}}$  by setting large enough  $V$ s to make  $\frac{B}{V}$  arbitrarily small.

- An increase in throughput is accompanied by a growth in queue length (i.e., delay). This is because the variation in  $V$  gives rise to opposite performance changes in throughput and delay. For example, a large  $V$  leads to a good throughput (see (36)) but a bad delay (see (37)).
- Eqs. (36) and (37) together show the throughput-delay tradeoff of  $[O(1/V), O(V)]$ , which consolidates and quantifies the analysis made in Section II-C. Moreover, it provides an important and simple guideline on how to explicitly balance the throughput-delay performance. That is, it is only required to select appropriate  $V$ s. This will be further clarified in Section VI.

*Remark 5:* Although we consider i.i.d. traffic arrivals with constant arrival rate  $\lambda$  in the paper, it is worthwhile to demonstrate that our proposed algorithms and the obtained theoretical results hold for other arrivals that are independent from slot to slot, but the arrival rates  $\lambda$  are time-varying and ergodic (possibly non-i.i.d.). That is, the dynamic admission control and power allocation policies, (16) in the first-out scheme and (22) in the first-in scheme, are robust to the changes in the arrival rates. An intuitional explanations are as follows. Suppose that the arrival rate is  $\lambda^{(1)}$  for a certain duration of time, then changes to  $\lambda^{(2)}$  (perhaps due to changes in users' demands). This change will be reflected in the queue size  $Q(t)$  that builds up in the queues of the system. Further, as can be seen from (16) and (22), the controller makes the admission control and power allocation policies only based on the size of the queues without requiring the knowledge of traffic arrival rates  $\lambda$  (see Remark 4, Section III-C). Thus, the proposed algorithms (the eGuard-I and II) react smoothly to such changes in the statistics of traffic arrivals.

## VI. SIMULATION RESULTS AND ANALYSIS

In this section, we take distributed antenna systems (DAS), which have emerged as a promising cellular communication architecture for 5G networks [2], as an example to evaluate the performance of our proposed algorithms.

### A. Parameters Setting

As shown in Fig. 3, we consider single-cell downlink DAS (as in [11], [26], [27], [36]) with 8 distributed antenna units (DAUs) and 10 users (i.e.,  $N = 8$  and  $M = 10$ ), both of which are equipped with a single antenna. All DAUs are randomly deployed throughout the cell area and connected to the base station (BS) central processing unit (CPU) via optical fiber or other high-bandwidth wired connections that can be assumed to be an ideal back-haul. In the CPU, all the signal processing and resource allocation are done. The channels assigned to different users are orthogonal and thus, it is interference-free among users [11], [26], [27], [36].

We assume the system adopts continuous rate scheme,<sup>5</sup> i.e., rate-power function is given by (2). For simplicity of

<sup>5</sup>Note that we here exploit the continuous rate scheme only for the simulations, and as stated in Section I, the proposed formulations and algorithms also hold for discrete cases.

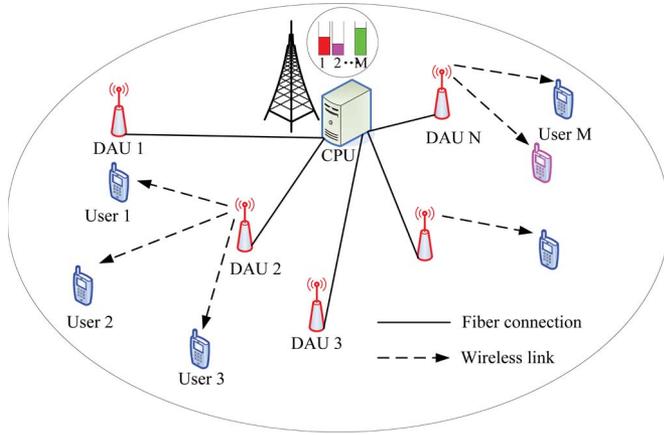


Fig. 3. Simulation scenario.

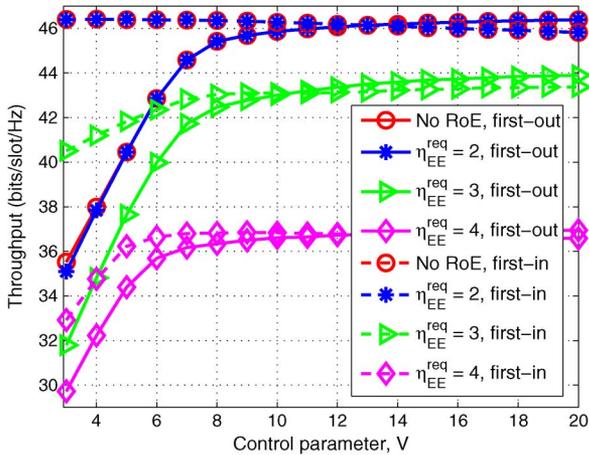


Fig. 4. Throughput against V under different  $\eta_{EE}^{req}$  in the first-out and first-in schemes.

simulations, we suppose that the wireless channel condition  $H_{n,m}(t) = g_{n,m}(t)/\sigma^2$  at any slot  $t$  takes values in the integer interval [3], [17], each of which having a probability of 1/15 [8], [26]. We set  $\xi_n = 1$  and  $w_m = 1$  for all  $n$  and  $m$ <sup>6</sup>,  $P_n^{max} = 2$  W, and  $P_n^C = 0.2$  W [11], [36], and assume that the traffic arrival is a Poisson distribution with  $\lambda_m = 10$  bits/slot/Hz for all receivers over slots in the simulations. Note that each point of the following curves is plotted based on a 6000-run and averaged over these values.

**B. Impacts of the Control Parameter on System Performance**

1) *The First-Out Scheme:* Figs. 4–6 quantitatively reveal how the control parameter  $V$  affects the performances of throughput, EE, and delay. Fig. 4 displays the evolution of throughput against  $V$ . It shows the convergence of Thr and thus, we can be sure that it inevitably converges to  $Thr^{opt}$  and can arbitrarily approach  $Thr^{opt}$  by setting  $V$  large enough from the analysis made in the last section. Further, it is clear that Thr increases to the optimal at the speed of  $O(1/V)$ , while average delay (i.e., queue length) grows linearly in  $O(V)$ , as shown in

<sup>6</sup>We set  $\xi_n = 1$  to denote ideal power amplifier (otherwise,  $\xi_n > 1$ ) [13] and  $w_m = 1$  to reflect the same priority on all links.

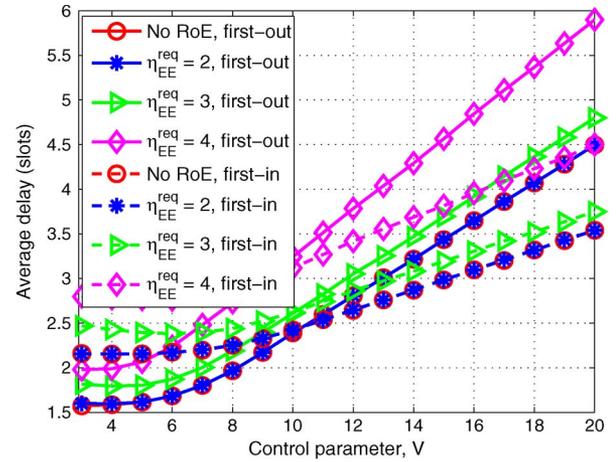


Fig. 5. Average delay against V under different  $\eta_{EE}^{req}$  in the first-out and first-in schemes.

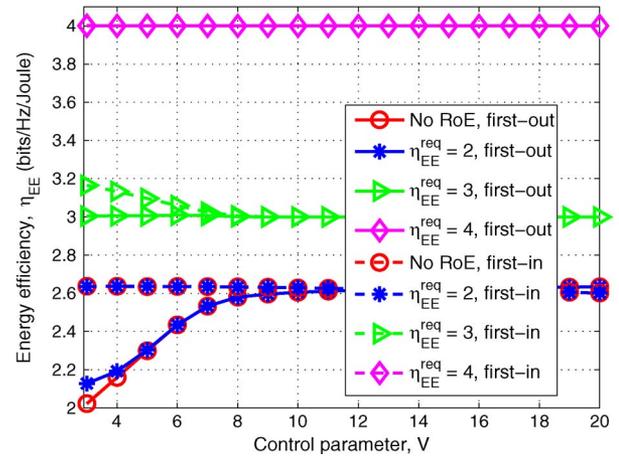


Fig. 6. Actually obtained EE against V under different  $\eta_{EE}^{req}$  in the first-out and first-in schemes.

Fig. 5, which consolidate (b) and (c) of the theoretical analysis in Theorem 1. Besides, Fig. 6 plots the actual obtained EE against  $V$ . It is observed that the actual obtained EE is always larger than or equal to  $\eta_{EE}^{req}$ . That is to say, constraint C1 in (14) is always satisfied. Thus, (a) of Theorem 1 is verified.

Under the constraint of the RoE, Figs. 4 and 5 together show the  $[O(1/V), O(V)]$  tradeoff of throughput-delay. This provides a significant guideline to dynamically balance throughput and delay by only setting an appropriate control parameter  $V$ . Specifically, if the system pursues a larger throughput, then a larger  $V$  is required (see Fig. 4). Otherwise, a smaller  $V$  is desired to ensure a smaller delay (see Fig. 5). This verifies the analysis presented in Section V-B. Moreover, the RoE is guaranteed (see Fig. 6) in any case.

From Figs. 4–6, we further find that there is a threshold  $\eta_{EE}^{th}$  for EE in the system when balancing throughput and delay, say  $\eta_{EE}^{th} = 2.63$  bits/Joule/Hz in our simulations (see Fig. 6). The value of  $\eta_{EE}^{th}$  is given by the actually obtained EE in the case without RoE. This indicates that the system is able to guarantee some degree of EE when maximizing throughput (see (10) and (14)). Naturally, for the cases with RoE below  $\eta_{EE}^{th}$ , the curves for the actually obtained EE and the tradeoff between

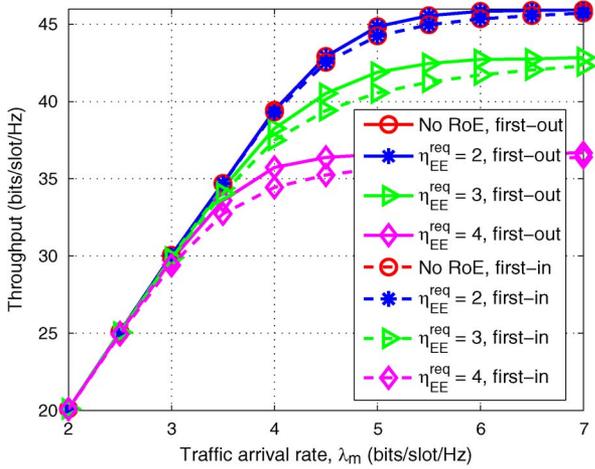


Fig. 7. Throughput against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{rmreq}$  with  $V = 10$  in the first-out and first-in schemes.

throughput and delay almost overlap with the case without RoE (e.g., for the case of  $\eta_{EE}^{req} = 2$  bits/Joule/Hz). Once  $\eta_{EE}^{rmreq}$  exceeds  $\eta_{EE}^{th}$ , say  $\eta_{EE}^{rmreq} = 4$ , the system throughput sharply decreases (see Fig. 4), and delay also has an increase (see Fig. 5). This is because, to satisfy the RoE, the system decreases transmit power (moreover, the obtained EE in this case is equal to RoE due to the fact that larger obtained EE results in a lower throughput) and thus, throughput decreases and delay increases (this follows from the decrease in the transmit rate resulting from the power reduction).

Furthermore, it is necessary to give deep insights into the obtained results. These indeed provide us with a big picture (from a perspective of systematic design) on how to achieve flexible control or coordination among throughput, EE, and delay. Specifically speaking, Figs. 4–6 show the intertwined but controllable (by  $V$ ) relations among throughput, delay, and EE. To guarantee a required RoE, there is an optimal traffic arrival rate to match it, i.e., there is a best flow control for TCP. For example, when  $V = 14$  and  $\eta_{EE}^{rmreq} = 4$  bits/Joule/Hz, the optimal throughput is about 36 bits/slot/Hz (see Fig. 4). Hence, it is not necessary for the window control of TCP to admit traffic rate larger or less than 36 bits/slot/Hz. Conversely, the traffic arrival rate affects EE and delay as well. Such as, to pursue a high EE and a low delay, a small traffic rate is required (This will be further clear in Figs. 7–9).

2) *The First-in Scheme:* The explanations to the first-in scheme in Figs. 4–6 are similar to those for the first-out scheme.

### C. Impacts of the Traffic Arrival Rate on System Performance

Figs. 7–9 show the variation tendencies of throughput, EE, and delay with the changes in the traffic arrival rate  $\lambda$  for a given  $V$ . From Fig. 7, we see that throughputs for both schemes are equal to the traffic arrival rates at the first, then grow to the maximums and keep unchanged as  $\lambda$  increases. This is because the system can stably work for small  $\lambda$ , but has to start admission control for large  $\lambda$ . Fig. 8 shows that the queue length always increases with  $\lambda$ , especially for large  $\lambda$ . This follows from the facts that more arrived data means longer queue length, and the transmit rate cannot increase to be large

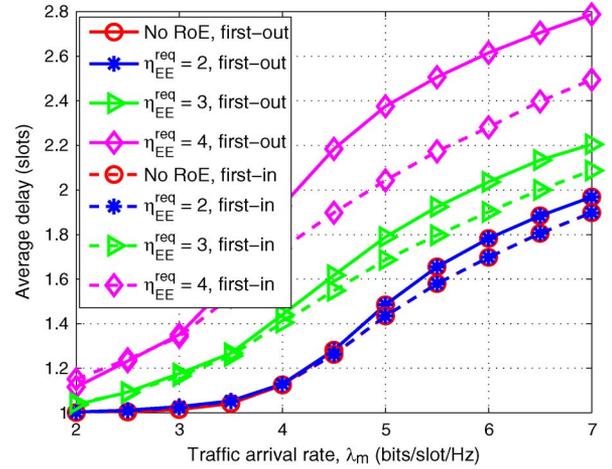


Fig. 8. Average delay against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{rmreq}$  with  $V = 10$  in the first-out and first-in schemes.

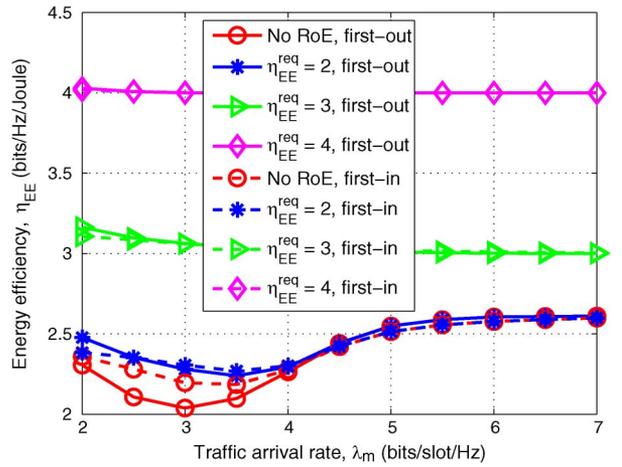


Fig. 9. The actually obtained EE against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{rmreq}$  with  $V = 10$  in the first-out and first-in schemes.

enough due to the RoE constraint. Besides, the RoE is always guaranteed from Fig. 9. Furthermore, it is observed that the variation in the actually obtained EE is almost contrary to that of the throughput when  $\text{RoE} < \eta_{EE}^{th}$ , while the EE is equal to the RoE for the case of  $\text{RoE} \geq \eta_{EE}^{th}$ , say  $\eta_{EE}^{th} = 4$  bits/Joule/Hz. This is because the biggest burden for the latter is to guarantee the RoE. Furthermore, Figs. 7–9 again consolidate that setting the RoE below or above the  $\eta_{EE}^{th}$  affects the system performance dramatically, as well as show that it is necessary to strike a balance among throughput, EE, and delay from a systematic perspective.

### D. Eguard Versus the Rate Adaptive Scheme

Further, we take the first-out scheme (i.e., the eGuard-I) as an example to compare its performance with the classical rate adaptive scheme (RAS) [37], which is specified as

$$\begin{aligned} & \max \sum_{m=1}^M R_m(t) \\ & \text{s.t. } C1: \eta_{EE}^{\text{RAS}}(t) \geq \eta_{EE}^{\text{req}} \end{aligned}$$

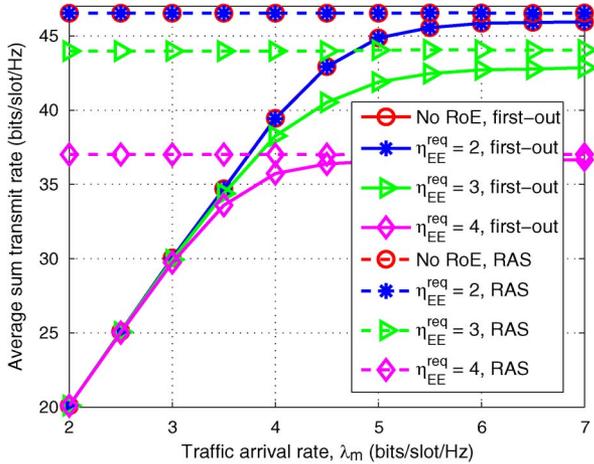


Fig. 10. Average sum transmit rate against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{req}$  in the RAS and first-out scheme with  $V = 10$ .

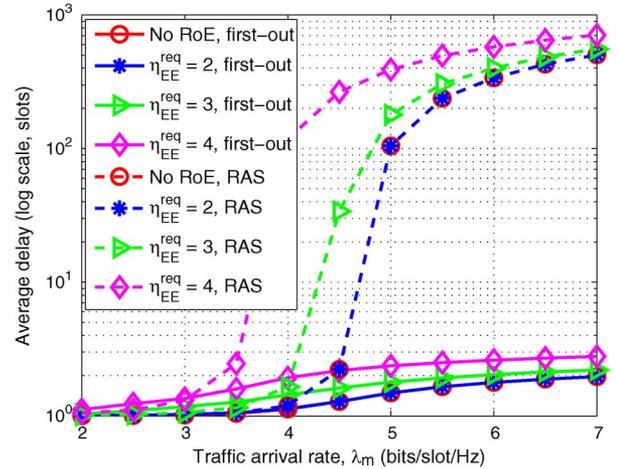


Fig. 12. Average delay against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{req}$  in the RAS and first-out scheme with  $V = 10$ .

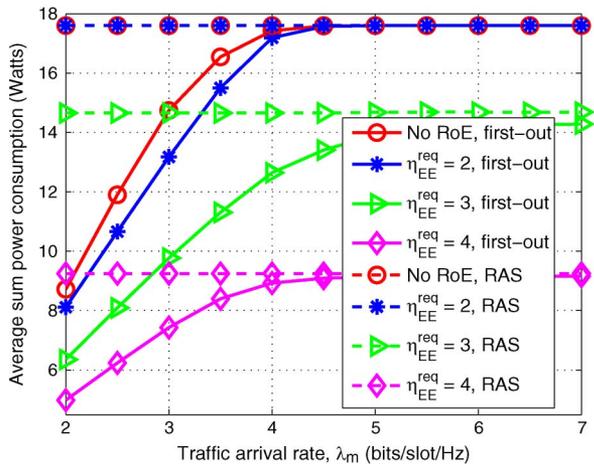


Fig. 11. Average power consumption against traffic arrival rate  $\lambda$  under different  $\eta_{EE}^{req}$  in the RAS and first-out scheme with  $V = 10$ .

$$\begin{aligned}
 C2: & \sum_{m=1}^M P_{n,m}(t) \leq P_n^{\max}, \forall n, t \\
 C3: & P_{n,m}(t) \geq 0, \forall n, m, t
 \end{aligned} \tag{47}$$

where  $\eta_{EE}^{RAS}(t) = \frac{R_{tot}(t)}{P_{tot}(t)}$ .

As can be seen from Fig. 10, the variation tendencies of the sum transmit rates (i.e., SE) against  $\lambda_m$  in the first-out scheme are similar to those of throughput in Fig. 7. This verifies the claim presented in Section I-A that SE can be used to characterize throughput. On the contrary, the sum transmit rates in the RAS keep unchanged as  $\lambda_m$  varies. This is because the RAS does not consider stochastic traffic arrivals and delivers data under the full buffer assumption (i.e., there are always data bits for transmission). As a result, the RAS fails to adapt to traffic arrivals and thus, leading to the RAS consuming much more power than that in the first-out scheme when the network is in the light traffic states, which is shown in Fig. 11. For example, for the case of  $\lambda_m = 3.5$  bits/slot/Hz (thus the sum traffic arrival rates is 35 bits/slot/Hz) and  $\eta_{EE}^{req} = 3$  bits/Joule/Hz, it is totally not necessary to deliver data at the speed of 44 bits/slot/Hz (see Fig. 10), and it is also a waste of energy (see Fig. 11).

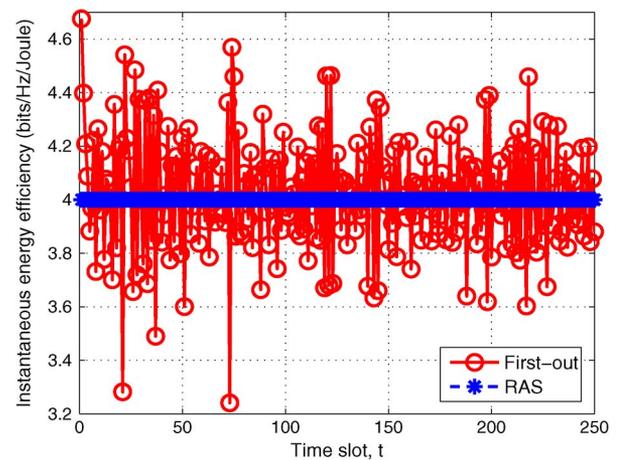


Fig. 13. The actually obtained EE over slots in the RAS and first-out scheme with  $V = 10$ ,  $\eta_{EE}^{req} = 4$  bits/Joule/Hz, and  $\lambda_m = 7$  bits/slot/Hz.

In Fig. 12, we compare the average delay in the first-out and RAS schemes. It is observed that the average delay in the two schemes is almost the same and very small for a given RoE when  $\lambda_m$  is small, e.g.,  $\lambda_m \leq 3.5$  bits/slot/Hz when  $\eta_{EE}^{req} = 3$  bits/Joule/Hz. This is because both schemes are able to timely transmit all the arrived data. In the following, the average delay in the RAS sharply increases as  $\lambda_m$  is large than a certain threshold, and it will tend to infinity as time elapses. That is to say, the system becomes unstable. However, it will never happen in the first-out scheme. The average delay in the first-out scheme is controlled to slowly increase with the traffic arrival rates and is far smaller than that in the RAS in all cases. This follows from the facts that 1) the queue stability constraint in (14) ensures the average delay to be controllable, and 2) the first-out scheme adapts to traffic arrivals attributed to the queue stability constraints, and it will automatically start admission control when the network is in the heavy traffic states.

### E. Illustration of the Definition of EE

Fig. 13 shows the instantaneously obtained EE over slots. As can be seen, the actually obtained EE  $\eta_{EE}^{RAS}(t)$  in the RAS

always equals  $\eta_{EE}^{rmreq}$  and keeps unchanged over slots. This again shows that the RAS cannot adapt to channel conditions and traffic arrivals. On the contrary, although  $\eta_{EE}^{rmreq} = 4$  bits/Joule/Hz, the actually obtained EE in the first-out scheme is not necessary larger than or equal to 4 bits/Joule/Hz at each slot. It may be larger or less than 4 bits/Joule/Hz at a given slot. However, the EE defined by (10) during the communication process, as shown in Fig. 6, is ensured to satisfy the RoE. This demonstrates that the proposed algorithms, the eGuard (I and II), can adapt to channel conditions and match traffic loads. As a result, the EE definition given by (10) implies the concept of time diversity and thus, it has the advantages of modeling time-variant channel conditions and stochastic traffic arrivals.

### F. Comparison of the First-out and First-in Schemes

In comparison, for a given traffic arrival rate, it is observed from Figs. 4–6 that the first-out scheme has smaller throughput and EE but better delay performance when  $V$  is set to be small. As  $V$  increases, both schemes have almost the same throughput and EE, but the delay performance of the first-in scheme is much better than that of the first-out scheme. Therefore,

- If the system concerns delay very much, it should work in the first-out scheme with small  $V$ .
- If the system prefers throughput and delay simultaneously, it should work in the first-in scheme with a relatively large  $V$ .
- If the system pursues EE, it must sacrifice throughput and delay.

Apart from the aforementioned scheme selection issues, we further discuss the differences about the two schemes in practical implementations.

- The first-out scheme has relatively poor performance throughout the whole variation region of  $V$ . However, it benefits from the convenience and flexibility of the flow (window) control of TCP for a given  $V$  to optimally match throughput, EE, and delay with little (even no) data dropping (see Fig. 1). More specifically, the controller adjusts  $A_m(t)$  to match  $C_m(t)$  according to RoE and delay.
- The first-in scheme is more brutal. As shown in Fig. 2, it may drop a lot of (even all) data in the buffer. Brutal data dropping is not suitable for reliability-sensitive applications such as data downloading. What's worse, it will in turn lead to the retransmissions of data and thus, further aggravate the burden of the system. Besides, it is not convenient to do flow control. Hence, this scheme is suitable to work in the light or middle traffic states.

## VII. CONCLUSIONS AND FUTURE WORK

This paper has studied the problem of EE-guaranteed throughput-delay tradeoff. Two admission control schemes, referred to as the first-out and first-in schemes, have been proposed. We have formulated two stochastic optimization problems to maximize the throughput (in the first-out scheme) or minimize the dropping rate (in the first-in scheme) subject to RoE, stability, admission control, and transmit power constraints. The eGuard (called the eGuard-I and II in the first-out

and first-in schemes, respectively) has been proposed to solve the problems. Theoretical analysis has shown that the eGuard (I and II) can achieve the optimal throughput with a tradeoff in the average delay, but the RoE can always be guaranteed. Simulation results have verified the theoretical results and shown the differences between the two schemes.

Our work is suitable for non-real-time (i.e., delay-tolerant) traffics such as web browsing and file transfers, as we focus on the average delay in the paper. However, there are some other real-time applications in realistic systems, e.g., voice and mobile video, which impose hard-deadline (or maximum delay) constraints. Thus, it is an interesting research direction to extend our proposed models and analysis to real-time systems to provide deterministic delay guarantee. Moreover, for more complicated systems with coexistence of non-real-time and real-time traffics, how to flexibly balance throughput, EE, and delay from a perspective of systematic design and further analyze its performance are also well worth studying.

### APPENDIX A PROOF OF LEMMA 1

From (15), we naturally have

$$Z(t+1) \geq Z(t) + y(t) \quad (48)$$

where  $y(t) = \eta_{EE}^{\text{req}} P_{\text{Tot}}(t) - R_{\text{Tot}}(t)$  and  $Z(0) = 0$ .

Summing (48) over  $t \in \{0, 1, \dots, K-1\}$  and taking expectation, we obtain

$$\mathbb{E}\{Z(K)\} \geq \sum_{t=0}^{K-1} \mathbb{E}\{\eta_{EE}^{\text{req}} P_{\text{Tot}}(t)\} - \sum_{t=0}^{K-1} \mathbb{E}\{R_{\text{Tot}}(t)\}. \quad (49)$$

Dividing (49) by  $K$  and taking  $K \rightarrow \infty$ , we have

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{\mathbb{E}\{Z(K)\}}{K} \\ & \geq \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=0}^{K-1} \mathbb{E}\{\eta_{EE}^{\text{req}} P_{\text{Tot}}(t)\} - \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=0}^{K-1} \mathbb{E}\{R_{\text{Tot}}(t)\}. \end{aligned} \quad (50)$$

According to Jensen's inequality,  $0 \leq \mathbb{E}\{Z(K)\} \leq \mathbb{E}\{|Z(K)|\}$ . Thus, if  $Z(t)$  is mean rate stable, i.e.,  $\lim_{K \rightarrow \infty} \frac{\mathbb{E}\{Z(K)\}}{K} = 0$ , we have  $\lim_{K \rightarrow \infty} \frac{\mathbb{E}\{Z(K)\}}{K} = 0$ . So (50) becomes

$$\eta_{EE}^{\text{req}} \bar{P}_{\text{Tot}} \leq \bar{R}_{\text{Tot}}. \quad (51)$$

That is to say,  $\frac{\bar{R}_{\text{Tot}}}{\bar{P}_{\text{Tot}}} \geq \eta_{EE}^{\text{req}}$  holds. We complete the proof of Lemma 1.

### APPENDIX B PROOF OF LEMMA 2

Since  $\{\max[Q - R, 0] + A\}^2 \leq Q^2 + R^2 + A^2 - 2Q(R - A)$ , squaring both sides of (7) results in

$$\begin{aligned} Q_m(t+1)^2 & \leq Q_m(t)^2 + R_m(t)^2 + C_m(t)^2 \\ & \quad - 2Q_m(t)(R_m(t) - C_m(t)). \end{aligned} \quad (52)$$

Summing over all  $m \in \{1, 2, \dots, M\}$  yields

$$\begin{aligned} & \sum_{m=1}^M \frac{Q_m(t+1)^2 - Q_m(t)^2}{2} \\ & \leq \sum_{m=1}^M \frac{R_m(t)^2 + C_m(t)^2}{2} - \sum_{m=1}^M Q_m(t) (R_m(t) - C_m(t)). \end{aligned} \quad (53)$$

Similarly, for  $Z(t)$ , we have

$$\frac{Z(t+1)^2 - Z(t)^2}{2} \leq \frac{y(t)^2}{2} + Z(t)y(t). \quad (54)$$

Summing (53) and (54) generates

$$\begin{aligned} L(\Theta(t+1)) - L(\Theta(t)) & \leq \sum_{m=1}^M \frac{R_m(t)^2 + C_m(t)^2}{2} + \frac{y(t)^2}{2} \\ & \quad + \sum_{n=1}^N Z(t)y(t) - \sum_{m=1}^M Q_m(t) (R_m(t) - C_m(t)). \end{aligned} \quad (55)$$

Taking conditional expectations of the above inequalities, we further obtain

$$\begin{aligned} & \Delta(\Theta(t)) - V\mathbb{E}\{\text{Thr}(t)|\Theta(t)\} \\ & \leq B - V\mathbb{E}\{\text{Thr}(t)|\Theta(t)\} + Z(t)\mathbb{E}\{y(t)|\Theta(t)\} \\ & \quad + \sum_{m=1}^M Q_m(t)\mathbb{E}\{C_m(t) - R_m(t)|\Theta(t)\} \end{aligned} \quad (56)$$

where

$$\begin{aligned} y(t) & = \eta_{\text{EE}}^{\text{req}} P_{\text{tot}}(t) - R_{\text{tot}}(t) \\ B & \geq \frac{1}{2} \sum_{m=1}^M \mathbb{E}\{C_m(t)^2 + R_m(t)^2 | \Theta(t)\} + \frac{1}{2} \mathbb{E}\{y(t)^2 | \Theta(t)\}. \end{aligned} \quad (57)$$

This completes the proof of Lemma 2.

## REFERENCES

- [1] E. M. Yeh, "Fundamental performance limits in cross-layer wireless optimization: Throughput, delay, energy," *Found. Trends Commun. Inf. Theory*, vol. 9, no. 1, pp. 1–112, 2012.
- [2] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5 G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.
- [3] The 1000x data challenge. [Online]. Available: <http://www.qualcomm.com/solutions/wireless-networks/technologies/1000x-data>
- [4] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, White paper, Feb. 2013, 2012–2017.
- [5] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55–62, Aug. 2011.
- [6] D. Feng *et al.*, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 2013.
- [7] H. Ju, B. Liang, J. Li, and X. Yang, "Dynamic power allocation for throughput utility maximization in interference-limited networks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 22–25, Feb. 2013.
- [8] M. J. Neely, "Energy optimal control for time-varying wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2915–2934, Jul. 2006.
- [9] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
- [10] X. Hong, J. Wang, C.-X. Wang, and J. Shi, "Cognitive radio in 5 G: A perspective on energy-spectral efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 46–53, Jul. 2014.
- [11] C. He, B. Sheng, P. Zhu, X. You, and G. Y. Li, "Energy- and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 894–902, May 2013.
- [12] Y. Li, M. Sheng, C. Yang, and X. Wang, "Energy efficiency and spectral efficiency tradeoff in interference-limited wireless networks," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1924–1927, Oct. 2013.
- [13] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.
- [14] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.
- [15] I. Ku, C.-X. Wang, and J. Thompson, "Spectral-energy efficiency tradeoff in relay-aided cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4970–4982, Oct. 2013.
- [16] I. Ku, C.-X. Wang, and J. Thompson, "Spectral, energy and economic efficiency of relay-aided cellular networks," *IET Commun.*, vol. 7, no. 14, pp. 1476–1486, Sep. 2013.
- [17] X. Hong, Y. Jie, C.-X. Wang, J. Shi, and X. Ge, "Energy-spectral efficiency trade-off in virtual MIMO cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2128–2140, Oct. 2013.
- [18] M.-R. Ra *et al.*, "Energy-delay tradeoffs in smartphone applications," in *Proc. MobiSys*, New York, NY, USA, Jun. 2010, pp. 255–270.
- [19] P. Shu *et al.*, "eTime: Energy-efficient transmission between cloud and mobile devices," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 14–19.
- [20] M. J. Neely, "Low power dynamic scheduling for computing systems," in *Green Communications and Networking*, F. R. Yu, X. Zhang, and V. C. M. Leung, Eds. Boca Raton, FL, USA: CRC, 2012, pp. 219–259.
- [21] Y. Li, M. Sheng, C. Yang, X. Ma, and Z. Li, "Energy-delay tradeoffs for multi-user multi-antenna downlink communication systems," in *Proc. ICT*, Casablanca, Morocco, May 2013, pp. 1–5.
- [22] J. Wu, S. Zhou, and Z. Niu, "Traffic-aware base station sleeping control and power matching for energy-delay tradeoffs in green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4196–4209, Aug. 2013.
- [23] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, Sep. 2011.
- [24] F. Meshkati, H. Poor, and S. Schwartz, "Energy efficiency-delay tradeoffs in CDMA networks: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3220–3228, Jul. 2009.
- [25] R. Heath, S. Peters, Y. Wang, and J. Zhang, "A current perspective on distributed antenna systems for the downlink of cellular systems," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 161–167, Apr. 2013.
- [26] Y. Li, M. Sheng, Y. Zhang, X. Wang, and J. Wen, "Energy-efficient antenna selection and power allocation in downlink distributed antenna systems: A stochastic optimization approach," in *Proc. IEEE ICC*, Sydney, NSW, Australia, Jun. 2014, pp. 4963–4968.
- [27] Y. Li, M. Sheng, X. Wang, Y. Shi, and Y. Zhang, "Globally optimal antenna selection and power allocation for energy efficiency maximization in downlink distributed antenna systems," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, Dec. 2014, to be published.
- [28] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [29] Y. Li, M. Sheng, X. Wang, Y. Zhang, and J. Wen, "Max-min energy-efficient power allocation in interference-limited wireless networks," *IEEE Trans. Veh. Technol.*, 2014, to be published.
- [30] F. Haider, C.-X. Wang, H. Haas, E. Hepsaydir, X. Ge, and Y. D., "Spectral and energy efficiency for heterogeneity analysis for heterogeneous cognitive radio networks," *IEEE Trans. Wireless Commun.*, 2014, to be published.
- [31] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. San Rafael, CA, USA: Morgan & Claypool, 2010.
- [32] Y. Li, M. Sheng, Y. Shi, X. Ma, and W. Jiao, "Energy efficiency and delay tradeoff for time-varying and interference-free wireless networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 5921–5931, Nov. 2014.
- [33] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.
- [34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

- [35] M. Grant, S. Boyd, and Y. Ye, *Cvx: Matlab Software for Disciplined Convex Programming*, version 2.0 beta, Sep. 2012. [Online]. Available: <http://cvxr.com/cvx>
- [36] H. Kim, S.-R. Lee, C. Song, and I. Lee, "Optimal power allocation for energy efficiency maximization in distributed antenna systems," in *Proc. IEEE ICC*, Budapest, Hungary, Jun. 2013, pp. 5769-5773.
- [37] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726-2737, Nov. 2005.



**Yuzhou Li** received the B.Eng. degree in electronic and information engineering from Jilin University, Changchun, China, in 2009. He is currently working toward the Ph.D. degree with the State Key Laboratory of ISN and the School of Telecommunications Engineering, Xidian University, Shaanxi, China. His research interests include green communications, wireless resource allocation, and convex optimization and stochastic network optimization and their applications in wireless communications.



**Min Sheng** (M'03) received the M. Eng. and Ph.D. degrees in communication and information systems from Xidian University, Shaanxi, China, in 1997 and 2000, respectively. She is currently a Full Professor with the Broadband Wireless Communications Laboratory, School of Telecommunication Engineering, Xidian University. She is the author of two books and over 50 papers in refereed journals and conference proceedings. Her general research interests include mobile ad hoc networks, wireless sensor networks, wireless mesh networks, third- and fourth-

generation mobile communication systems, dynamic radio resource management for integrated services, cross-layer algorithm design and performance evaluation, cognitive radio and networks, cooperative communications, and medium access control protocols. She was the New Century Excellent Talents in University by the Ministry of Education of China and obtained the Young Teachers Award from the Fok Ying-Tong Education Foundation, China, in 2008.



**Cheng-Xiang Wang** (S'01-M'05-SM'08) received the B.Sc. and M.Eng. degrees in communication and information systems from Shandong University, Jinan, China, in 1997 and 2000, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2004. From 2000 to 2001, he was a Research Assistant with the Hamburg University of Technology, Hamburg, Germany. In 2004, he was a Visiting Researcher with Siemens AG-Mobile Phones, Munich, Germany. From 2001 to 2005, he was a Research Fellow with

the University of Agder, Grimstad, Norway. Since 2005, he has been with Heriot-Watt University, Edinburgh, U.K., where he was also promoted as a Professor in 2011. He is also an Honorary Fellow of the University of Edinburgh, Edinburgh, and a Chair/Guest Professor with Shandong University, and with Southeast University, Nanjing, China. He is the Editor of one book. He has published one book chapter and over 200 papers in refereed journals and conference proceedings. His research interests include wireless channel modeling and simulation, green communications, cognitive radio networks, vehicular communication networks, massive multiple-input-multiple-output systems, and fifth-generation wireless communications.

Prof. Wang is a Fellow of the Institution of Engineering and Technology and the HEA and a member of the Engineering and Physical Research Council Peer Review College. He has served as an Editor for eight international journals, including the *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY* (since 2011) and the *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS* (2007-2009). He was the lead Guest Editor for the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, Special Issue on vehicular Communications and Networks. He has served as a Technical Program Committee (TPC) Member, a TPC Chair, and a General Chair for more than 70 international conferences. He received the Best Paper Awards at *IEEE Globecom* 2010, *IEEE ICCT* 2011, *ITST* 2012, and *IEEE VTC* 2013-Spring.



**Xijun Wang** (M'12) received the B.S. degree (with distinction) in telecommunications engineering from Xidian University, Xi'an, Shaanxi, China, in 2005. He received the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in January 2012. Since 2012, he has been with School of Telecommunications Engineering, Xidian University, where he is currently an Assistant Professor. His research interests include wireless communications, cognitive radios and interference management.

He served as Publicity Chair of *IEEE/CIC ICC* 2013. He received the 2005 "Outstanding Graduate of Shaanxi Province" Award, the Excellent Paper Award at the Sixth International Student Conference on Advanced Science and Technology in 2011, and the Best Paper Award at *IEEE/CIC ICC* 2013.



**Yan Shi** (M'10) received the Ph.D. degree from Xidian University, Shaanxi, China, in 2005. He is currently an Associate Professor with the State Key Laboratory of ISN, Xidian University. His research interests include cognitive networks, modern switching technologies, and distributed wireless networking.



**Jiandong Li** (SM'05) received the B.E., M.S., and Ph.D. degrees in communications engineering from Xidian University, Xi'an, China, in 1982, 1985 and 1991 respectively. Since 1985, he has been a faculty member with the School of Telecommunications Engineering, Xidian University, where he is currently a Professor and the Vice Director of the academic committee of State Key Laboratory of Integrated Service Networks. From 2002 to 2003, he was a Visiting Professor with the Department of Electrical and Computer Engineering, Cornell University, Ithaca,

NY, USA. His main research interests include communications and information systems, cognitive radio, and signal processing. He served as the General Vice Chair for *ChinaCom* 2009 and TPC Chair of *IEEE ICC* 2013. He received the Distinguished Young Researcher Award from NSFC and the Changjiang Scholarship the from Ministry of Education of China.