Journal of Network and Computer Applications 44 (2014) 152–160

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Achieving optimal admission control with dynamic scheduling in energy constrained network systems

Weiwei Fang [a,b,*], Zhulin An [c], Lei Shu [d], Qingyu Liu [a], Yongjun Xu [c], Yuan An [b]

[a] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[b] State Key Lab of Astronautical Dynamics of China, Xi'an 710043, China
[c] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[d] Guangdong Provincial Key Lab of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form
19 April 2014
Accepted 19 May 2014
Available online 9 June 2014

Keywords:
Dynamic scheduling
Optimal control
Admission control
Energy efficiency
Renewal system

ABSTRACT

This paper considers optimization of time average admission rate in an energy-constrained network system with multiple classes of data flows. The system operates regularly over time intervals called frames, while each frame begins with a fixed-length active period and ends with a variable-length idle period. At the beginning of the frame, the system chooses a service mode from a collection of options that affect the class and the amount of data flow served as well as the energy incurred in the active period. After service, the system chooses an amount of time to remain idle. The optimization goal is to make decisions over time that maximizes a weighted sum of admitted data rates subject to constraints on queue stability and energy expenditure. However, conventional solutions suffer from a curse of dimensionality for systems with large state space. Therefore, using a generalized Lyapunov optimization technique, we design a new online control algorithm that solves the problem. The algorithm can push time average admission rate close to optimal, with a corresponding tradeoff in average queue backlog. Remarkably, it does not require any knowledge of the data arrival rates and is provably optimal.

1. Introduction

In this paper, we consider a type of network system that processes and transmits data information with a limited energy supply. There are totally N classes of data flows, while raw data of each class arrive randomly and are stored in separate queues to await service. The system operates over time intervals called frames. Each frame r begins with a fixed-length active period of size D and ends with a variable-length idle period of size $I[r]$. At the beginning of the rth frame, the system chooses a service mode $m[r]$ from a collection of available options. The mode $m[r]$ determines the class of data flow and the amount of raw data that it will serve in the active period of frame r, and the resulting energy expenditure. After data processing and transmission, the system chooses an amount of idle time $I[r]$ to be idle. When the (possibly 0) idle period ends, the system wakes up and a new frame is begun. The goal is to design an algorithm for dynamically making decisions to maximize time average system admission subject to stability requirements and energy constraints. This problem has arisen in many network communication scenarios where the users want to maintain the lifetime of resource-constrained node as long as possible, while processing and transmitting data as much as possible. Typical scenarios include sensor-to-sink data reporting in duty-cycled sensor networks (Keshavarzian et al., 2006; Zhang et al., 2013a), and non-real-time data communication in new-generation mobile networks (Deng and Balakrishnan, 2012; Niu et al., 2014).

The Renewal Theory (Gallager, 1996) is a conventional technique for solving such a problem. It can be shown that, based on the renewal-reward theorem (Gallager, 1996), there exists an optimal control algorithm that makes independent and identically distributed (i.i.d) decisions over frames. However, it is usually difficult to choose such a control policy in an online fashion, since this would require a priori knowledge of system statistics and incur excessive computational complexity for finding the i.i.d. policy (Neely, 2013). Furthermore, these solutions result in hard-to-implement systems, since significant re-computation might be incurred when statistics change.

In this paper, we develop a simple dynamic scheduling algorithm to solve the problem stated above based on the recently developed technique of Lyapunov optimization (Neely, 2013). In every frame, this algorithm observes the current queue status, and minimizes a bound on the drift-plus-penalty ratio by making control decisions on flow control and service scheduling. It can

* Corresponding author at: School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China.
E-mail addresses: wwfang@bjtu.edu.cn (W. Fang), anzhulin@ict.ac.cn (Z. An), lei.shu@lab.gdupt.edu.cn (L. Shu), spencerliuqingyu993@gmail.com (Q. Liu), xyj@ict.ac.cn (Y. Xu), 9112533@qq.com (Y. An).

http://dx.doi.org/10.1016/j.jnca.2014.05.009
1084-8045/

obtain a time average admission rate within a deviation of $\mathcal{O}(1/V)$ from the optimum, with an average queue size tradeoff that is $\mathcal{O}(V)$, where V is a non-negative parameter that weights the extent to which admission maximization is emphasized as compared to system stability. The algorithm does not require knowledge of the traffic arrival rates, and naturally adapts if these rates change. Moreover, it is computationally efficient and easy to implement in the practical systems. We thoroughly analyze the performance of this algorithm with rigorous theoretical analysis. We also carried out extensive simulation experiments to demonstrate its effectiveness and adaptiveness. To our knowledge, prior work has not explored such a problem for network systems with renewal frames, and our use of the Lyapunov framework for solving this issue is also novel.

The rest of this paper is organized as follows: Section 2 formulates the objective problem. Section 3 presents the dynamic scheduling algorithm, and Section 4 provides an analysis on performance bounds of our algorithm. Section 5 shows the performance evaluation results. Section 6 reviews some related work. Finally, Section 7 concludes the paper.

2. System model and problem formulation

2.1. System model

Consider a network system s with N classes of data flows, as shown in Fig. 1. Raw data of each class arrive randomly according to an i.i.d. arrival process with rates $\lambda_1, \ldots, \lambda_N$. We assume that there exists a finite constant λ^{max} such that $\lambda_n \le \lambda^{max}$ for all n. These data are stored in separate queues according to their classes. Let $Q_n(t)$ represents the amount of class n data queued at time t. The system operates in continuous time, so the time index t takes values in the set of non-negative real numbers. The value of $Q_n(t)$ is a non-negative real number for all $n \in \{1, \ldots, N\}$ and for all $t \ge 0$. Assume that the system is initially empty at time $t=0$, therefore $Q_n(0) = 0$ for all $n \in \{1, \ldots, N\}$.

The system s makes decisions over renewal frames, as shown in Fig. 2. The first frame is labelled as frame 0 and starts at time 0. At the beginning of each frame $r \in \{0, 1, 2, \ldots\}$, s chooses a variable $c[r] \in \{0, 1, \ldots, N\}$ that specifies which class of data flow will be served, where $c[r] = 0$ is a null choice that selects no data flow to serve and incurs little energy consumption. s also chooses a service mode $m[r]$ from a finite set $\mathcal{M}$ of mode options. The control decisions $c[r]$ and $m[r]$ determine values $\mu_n[r]$ for each $n \in \{1, \ldots, N\}$, representing the amount of class n data flow that can be served in the active period of frame r. They also determine the system processing energy $e[r]$ that is incurred. At the end of the processing time, s chooses an idle time $I[r]$ within an interval $[0, I_{max}]$ for some given non-negative value I_{max}. The energy expenditure in the idle state is often very low and even neglectable (Neely, 2010). Let $T[r] \in [D, D+I_{max}]$ be the total frame size, then all these above can be given by general functions as follows:

$$\mu_n[r] = \hat{\mu}_n[r] = \hat{\mu}_n(c[r], m[r]) \tag{1}$$

$$e[r] = \hat{e}[r] = \hat{e}(c[r], m[r], I[r]) \tag{2}$$

$$T[r] = \hat{T}[r] = D + I[r] \tag{3}$$

It is assumed that second moments of $\mu_n[r]$ and $e[r]$ are bounded by a finite constant d (Neely, 2012b, 2013), so that

$$\mathbb{E}[\hat{\mu}_n[r]^2] \le d \tag{4}$$

$$\mathbb{E}[\hat{e}[r]^2] \le d \tag{5}$$

where (4) and (5) holds regardless of the policy for any control decision.

Then, for each class of data flow $n \in \{1, \ldots, N\}$, we choose from the interval $[0, 1]$ a flow control variable $\gamma_n[r]$, which represents the probability of admitting new randomly arriving data of class n on frame r. This enables the system to decline new data of class n into the queue when Q_n cannot support to handle data in accordance with the raw arrival rate λ_n. It can easily be generalized to the case where arrivals that are not immediately accepted are stored in a buffer for future admission decision.

Finally, for each class of data flow $n \in \{1, \ldots, N\}$, we define $A_n[r]$ as the random number of new arrivals admitted on frame r, which depends on the total frame size $(D+I[r])$ and the admission probability $\gamma_n[r]$. Assume that the arrival vector $(A_1[r], \ldots, A_N[r])$ is conditionally independent of the past, and with expectations (Neely, 2012b):

$$\mathbb{E}[A_n[r]] = \lambda_n \gamma_n[r] \hat{T}[r] \tag{6}$$

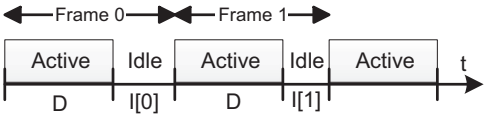

Fig. 2. A timeline illustrating the active and idle periods for each frame.

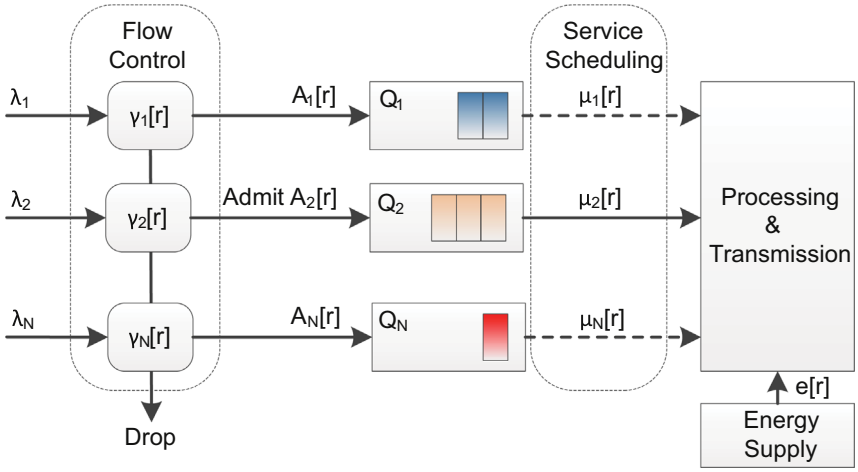

Fig. 1. A network system with random arrivals, flow control and service scheduling.

2.2. Problem formulation

Consider an algorithm that makes control decisions $c[r]$, $m[r]$, $I[r]$ and $\gamma_n[r]$ for each data flow $n \in \{1,2,\ldots,N\}$ in the network system on each frame $r \in \{0,1,2,\ldots\}$. For simplicity, it is assumed that, with probability 1, the algorithm yields well defined frame averages $\overline{\mu}_n$, $\overline{A}_n$, $\overline{T}$, $\overline{e}$, as defined below:

$$\overline{\mu}_n \triangleq \lim_{R\to\infty}\frac{1}{R}\sum_{r=0}^{R-1}\mu_n[r] \tag{7}$$

$$\overline{A}_n \triangleq \lim_{R\to\infty}\frac{1}{R}\sum_{r=0}^{R-1}A_n[r] \tag{8}$$

$$\overline{T} \triangleq \overline{D}+\overline{I} = D+\lim_{R\to\infty}\frac{1}{R}\sum_{r=0}^{R-1}I[r] \tag{9}$$

$$\overline{e} \triangleq \lim_{R\to\infty}\frac{1}{R}\sum_{r=0}^{R-1}e[r] \tag{10}$$

The time average rate that s admits class n data (in data amount per unit time) is equal to

$$\lim_{R\to\infty}\frac{\sum_{r=0}^{R-1}A_n[r]}{\sum_{r=0}^{R-1}T[r]} = \lim_{R\to\infty}\frac{\frac{1}{R}\sum_{r=0}^{R-1}A_n[r]}{\frac{1}{R}\sum_{r=0}^{R-1}T[r]} = \frac{\overline{A}_n}{\overline{T}} \tag{11}$$

Because all queues must be stabilized, the time average admitting rate of each queue n must be no larger than that of service, which is defined as follows:

$$\lim_{R\to\infty}\frac{\sum_{r=0}^{R-1}\mu_n[r]}{\sum_{r=0}^{R-1}T[r]} = \lim_{R\to\infty}\frac{\frac{1}{R}\sum_{r=0}^{R-1}\mu_n[r]}{\frac{1}{R}\sum_{r=0}^{R-1}T[r]} = \frac{\overline{\mu}_n}{\overline{T}} \tag{12}$$

Note that $\overline{e}$ does not represent the time average power consumption, since it does not take the amount of time spent in each frame into consideration. The time average power (energy per unit time) is written as follows:

$$\lim_{R\to\infty}\frac{\sum_{r=0}^{R-1}e[r]}{\sum_{r=0}^{R-1}T[r]} = \lim_{R\to\infty}\frac{\frac{1}{R}\sum_{r=0}^{R-1}e[r]}{\frac{1}{R}\sum_{r=0}^{R-1}T[r]} = \frac{\overline{e}}{\overline{T}} \tag{13}$$

Our objective is to maximize a weighted sum of admission rates (Wang and Agrawal, 2004; Neely, 2012b), subject to supporting all of the admitted data flows, and to maintaining average power to within a given positive constant P:

$$\begin{aligned}
&\text{Maximize} \quad \frac{\sum_{n=1}^{N}w_n\overline{A}_n}{\overline{T}}\\
&\text{Subject to} \quad \frac{\overline{A}_n}{\overline{T}} \le \frac{\overline{\mu}_n}{\overline{T}} \quad \forall n\in\{1,\ldots,N\}\\
&\qquad \frac{\overline{e}}{\overline{T}} \le P\\
&\qquad c[r]\in\{0,1,\ldots,N\} \quad \forall r\in\{0,1,\ldots\}\\
&\qquad m[r]\in\mathcal{M} \quad \forall r\in\{0,1,\ldots\}\\
&\qquad 0\le I[r]\le I_{max} \quad \forall r\in\{0,1,\ldots\}\\
&\qquad 0\le \gamma_n[r]\le 1 \quad \forall r\in\{0,1,\ldots\}, \forall n\in\{1,\ldots,N\}
\end{aligned} \tag{14}$$

where $(w_1,\ldots,w_N)$ are a set of positive weights that prioritize the different classes of data flows in the optimization objective.

In principle, we can find the optimal i.i.d. algorithm for (14) by solving a transformed linear fractional programming problem (Neely, 2013). However, neither offline solutions (Boyd and Vandenberghe, 2004) nor online solutions (Neely, 2012b) for linear fractional program are practical for solving such a problem due to their requirements on pre-knowledge of system dynamics and high computational complexity (Neely, 2012a, 2013). The algorithm developed in this paper overcomes these challenges.

3. The dynamic scheduling algorithm

This section develops a dynamic scheduling algorithm to solve the problem (14).

3.1. Virtual queues

To treat the constraints $\overline{A}_n/\overline{T} \le \overline{\mu}_n/\overline{T}$, define virtual queues $Q_n[r]$ for $n\in\{1,\ldots,N\}$. The queues have initial condition $Q_n[r]=0$ and update equation as follows:

$$Q_n[r+1] = \max[Q_n[r]+A_n[r]-\mu_n[r],0] \tag{15}$$

The intuition is that if we stabilize the queue $Q_n[r]$, then the average of the admission process $A_n[r]$ is less than or equal to that of the service process $\mu_n[r]$, i.e., $\overline{A}_n/\overline{T} \le \overline{\mu}_n/\overline{T}$.

Lemma 1. *If $Q_n[0]=0$ and $Q_n[r]$ satisfies (15) for all $r\in\{0,1,2,\ldots\}$, then for all integers $R>0$:*

$$\frac{1}{R}\sum_{r=0}^{R-1}(A_n[r]-\mu_n[r]) \le \frac{Q_n[R]}{R} \tag{16}$$

and hence

$$\overline{A}_n-\overline{\mu}_n \le \frac{\mathbb{E}[Q_n[R]]}{R} \tag{17}$$

Proof. From (15) we have

$$Q_n[r+1] \ge Q_n[r]+A_n[r]-\mu_n[r]$$

Summing over $r\in 0,1,\ldots,R-1$ gives

$$Q_n[R]-Q_n[0] \ge \sum_{r=0}^{R-1}A_n[r]-\sum_{r=0}^{R-1}\mu_n[r]$$

Dividing by R and using $Q_n[0]=0$ prove (16). Taking expectations proves (17). □

Similarly, to enforce $\overline{e}/\overline{T}\le P$, define a virtual queue $Z[r]$. This queue has initial condition $Z[r]=0$ and update equation as follows:

$$Z[r+1] = \max[Z[r]+e[r]-PT[r],0] \tag{18}$$

Lemma 2. *If $Z[0]=0$ and $Z[r]$ satisfies (18) for all $r\in\{0,1,2,\ldots\}$, then for all integers $R>0$:*

$$\frac{1}{R}\sum_{r=0}^{R-1}(e[r]-PT[r]) \le \frac{Z[R]}{R} \tag{19}$$

and hence

$$\overline{e}-P\overline{T} \le \frac{\mathbb{E}[Z[R]]}{R} \tag{20}$$

Proof. The proof is similar to that of Lemma 1, and therefore omitted here. □

It follows that if $\lim_{R\to\infty}\mathbb{E}[Q_n[R]]/R=0$ and $\lim_{R\to\infty}\mathbb{E}[Z[R]]/R=0$ for all $n\in\{1,\ldots,N\}$ (called mean rate stability, Neely, 2010), then all desired constraints are satisfied.

3.2. Lyapunov drift

Let $\mathbf{\Theta}[r]=(\mathbf{Q}[r],\mathbf{Z}[r])$ be a concatenated vector of all $Q_n[r]$ and $Z[r]$ queues. As a scalar measure of all the queue backlogs, we

define a quadratic Lyapunov function (Neely, 2010) as follows:

$$L[r] \triangleq \frac{1}{2}\sum_{n=1}^{N} Q_n[r]^2 + \frac{1}{2}Z[r]^2$$

The intuition is that we can take actions to consistently push this value down so that the queues will be stabilized. Define $\Delta[r]$ as the drift in the Lyapunov function from one frame to the next:

$$\Delta[r] \triangleq L[r+1] - L[r]$$

Taking actions to minimize $\Delta[r]$ every frame can be shown to ensure that the desired constraints on queue stability are satisfied when it is possible, but does not take the maximization objective and the frame length into consideration. To incorporate this, for every frame r we observe the current queue vector $\mathbf{\Theta}[r]$ and choose the control actions $c[r], m[r], I[r]$ and $\gamma_n[r]$ to minimize a bound on the following drift-plus-penalty ratio (Neely, 2010, 2012b):

$$\frac{\mathbb{E}[\Delta[r] - V\sum_{n=1}^{N} w_n A_n[r] | \mathbf{\Theta}[r]]}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]} \tag{21}$$

where V is a non-negative parameter that weights the extent to which admission maximization is emphasize. Note that the sign of the penalty expression is negative, since the problem in (14) has a maximum objective.

To construct an explicit algorithm, we first bound the ratio in (21).

Theorem 1. *For all frames $r \in \{0,1,2,\ldots\}$ and all classes $n \in \{1,\ldots,N\}$, all possible $\mathbf{\Theta}[r]$, and under any decisions for $c[r], m[r], I[r]$ and $\gamma_n[r]$, we have*

$$\begin{aligned}\frac{\mathbb{E}[\Delta[r] - V\sum_{n=1}^{N} w_n A_n[r] | \mathbf{\Theta}[r]]}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]} &\le \frac{B}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]} + \frac{\sum_{n=1}^{N} Q_n[r]\mathbb{E}[\lambda_n\gamma_n[r]\hat{T}[r] - \hat{\mu}_n[r] | \mathbf{\Theta}[r]]}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]} \\ &+ \frac{Z[r]\mathbb{E}[\hat{e}[r] - P\hat{T}[r] | \mathbf{\Theta}[r]]}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]} - \frac{V\mathbb{E}[\sum_{n=1}^{N} w_n\lambda_n\gamma_n[r]\hat{T}[r] | \mathbf{\Theta}[r]]}{\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]}\end{aligned} \tag{22}$$

where B is a constant that satisfies the following for all possible $\mathbf{\Theta}[r]$ and all policies:

$$B \ge \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[(\lambda_n\gamma_n[r]\hat{T}[r] - \hat{\mu}_n[r])^2 | \mathbf{\Theta}[r]] + \frac{1}{2}\mathbb{E}[(\hat{e}[r] - P\hat{T}[r])^2 | \mathbf{\Theta}[r]]$$

Such a constant B exists by the boundedness assumptions (4) and (5). In particular, we can use

$$B = \frac{N[(D+I_{max})\lambda^{max}]^2 + (N+1)d + [(D+I_{max})P]^2}{2}$$

Proof. Squaring (15) and noting that $\max[x,0]^2 \le x^2$, we have for each n,

$$Q_n[r+1]^2 \le Q_n[r]^2 + (A_n[r] - \mu_n[r])^2 + 2Q_n[r](A_n[r] - \mu_n[r])$$

Summing the above over $n = 1,\ldots,N$, we have

$$\begin{aligned}&\frac{1}{2}\sum_{n=1}^{N} Q_n[r+1]^2 - \frac{1}{2}\sum_{n=1}^{N} Q_n[r]^2 \\ &\le \frac{1}{2}\sum_{n=1}^{N} (A_n[r] - \mu_n[r])^2 + \sum_{n=1}^{N} Q_n[r](A_n[r] - \mu_n[r])\end{aligned}$$

Similarly, we have

$$\frac{1}{2}Z[r+1]^2 - \frac{1}{2}Z[r]^2 \le \frac{1}{2}(e[r] - PT[r])^2 + Z[r](e[r] - PT[r])$$

Combining these two bounds together, using the facts (1), (2) and (6), and taking the expectation with respect to $\Theta[r]$ on both sides, we have

$$\begin{aligned}\mathbb{E}[\Delta[r] | \mathbf{\Theta}[r]] \le{}& \frac{1}{2}\sum_{n=1}^{N} \mathbb{E}[(\lambda_n\gamma_n[r]\hat{T}[r] - \hat{\mu}_n[r])^2 | \mathbf{\Theta}[r]] \\ &+ \frac{1}{2}\mathbb{E}[(\hat{e}[r] - P\hat{T}[r])^2 | \mathbf{\Theta}[r]] \\ &+ \sum_{n=1}^{N} Q_n[r]\mathbb{E}[\lambda_n\gamma_n[r]\hat{T}[r] - \hat{\mu}_n[r] | \mathbf{\Theta}[r]] \\ &+ Z[r]\mathbb{E}[\hat{e}[r] - P\hat{T}[r] | \mathbf{\Theta}[r]]\end{aligned}$$

Now adding to both sides the penalty expression, i.e., the term $-V\mathbb{E}[\sum_{n=1}^{N} w_n\lambda_n\gamma_n[r]\hat{T}[r] | \mathbf{\Theta}[r]]$, and dividing both sides with $\mathbb{E}[\hat{T}[r] | \mathbf{\Theta}[r]]$, we prove the theorem. □

Hence, rather than directly minimizing the drift-plus-penalty expression every frame, we can actually seek to minimize the bound given in the right-hand side of (22). Therefore, we can design the dynamic scheduling algorithm as in Algorithm 1.

Algorithm 1. The dynamic scheduling algorithm.

On each time frame $r \in \{0,1,2,\ldots\}$, observe the vector of current queue states $\mathbf{\Theta}[r]$ and perform the following:

1: **Flow Control**: For each $n \in \{1,\ldots,N\}$, choose $\gamma_n[r]$ as

$$\gamma_n[r] = \begin{cases} 1 & \text{if } Q_n[r] \le Vw_n \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

2: **Service Scheduling**: Choose $c[r] \in \{0,1,\ldots,N\}$, $m[r] \in \mathcal{M}$, $I[r] \in [0, I_{max}]$ to minimize

$$\begin{aligned}&\frac{Z[r]\hat{e}[r] - \sum_{n=1}^{N} Q_n[r]\hat{\mu}_n[r]}{\hat{T}[r]} \\ &= \frac{Z[r]\hat{e}(c[r], m[r], I[r]) - \sum_{n=1}^{N} Q_n[r]\hat{\mu}_n(c[r], m[r])}{D + I[r]}\end{aligned} \tag{24}$$

3: **Queue Update**: Observe the resulting $\gamma_n[r]$, $\mu_n[r]$, $e[r]$ values to update $Q_n[r]$ for each $n \in \{1,\ldots,N\}$ by (15) and update $Z[r]$ by (18).

4. Algorithm analysis

In this section, we analyze the performance bound of our dynamic scheduling algorithm.

Theorem 2. *Suppose $Q_n[0] = 0$ and $Z[0] = 0$ for all $n \in \{1,\ldots,N\}$, and that the problem (14) is feasible. Then under our dynamic scheduling algorithm,*

(1) *For all frames $R \in \{1,2,3,\ldots\}$, we have*

$$\frac{\sum_{n=1}^{N} w_n\overline{A}_n[R]}{\overline{T}[R]} \ge OPT - \frac{B}{V\overline{T}[R]} \tag{25}$$

where B is defined in Theorem 1, OPT is the maximization admission solution for the problem (14), and $\overline{T}[R]$, $\overline{A}_n[R]$ are defined by

$$\overline{T}[R] \triangleq \frac{1}{R}\sum_{r=0}^{R-1} \mathbb{E}[T[r]], \quad \overline{A}_n[R] \triangleq \frac{1}{R}\sum_{r=0}^{R-1} \mathbb{E}[A_n[r]]$$

(2) *For all frames $R \in \{1,2,3,\ldots\}$ and all $n \in \{1,\ldots,N\}$, we have*

$$Q_n[R] \le Vw_n + \lambda^{max}(D + I_{max}) \tag{26}$$

(3) *For all $n \in \{1,\ldots,N\}$, we have*

$$\lim_{R\to\infty} \sup \frac{\overline{A}_n[R] - \overline{\mu}_n[R]}{\overline{T}[R]} \le 0$$

and

$$\lim_{R\to\infty}\sup\frac{\overline{e}[R]}{\overline{T}[R]}\le P$$

Proof. (1) Given $\Theta[r]$ for frame r, our control decisions minimize the last three terms in the right-hand side of the drift-ratio bound (22), and hence

$$\begin{aligned}
&\mathbb{E}\left[\Delta[r]-V\sum_{n=1}^{N}w_nA_n[r]|\mathbf{\Theta}[r]\right]\\
&\quad\le B+\sum_{n=1}^{N}Q_n[r]\mathbb{E}[\lambda_n\gamma_n[r]\hat{T}[r]-\hat{\mu}_n[r]\big|\mathbf{\Theta}[r]]+Z[r]\mathbb{E}[\hat{e}[r]-P\hat{T}[r]\big|\mathbf{\Theta}[r]]\\
&\qquad-V\mathbb{E}\left[\sum_{n=1}^{N}w_n\lambda_n\gamma_n[r]\hat{T}[r]|\mathbf{\Theta}[r]\right]\\
&\quad\le B+\mathbb{E}[\hat{T}[r]|\mathbf{\Theta}[r]]\frac{\sum_{n=1}^{N}Q_n[r]\mathbb{E}[\lambda_n\gamma_n^*[r](D+\hat{I}^*[r])-\hat{\mu}_n(c^*[r],m^*[r])]}{\mathbb{E}[(D+\hat{I}^*[r])]}\\
&\quad+\mathbb{E}[\hat{T}[r]|\mathbf{\Theta}[r]]\frac{Z[r]\mathbb{E}[\hat{e}(c^*[r],m^*[r],I^*[r])-P(D+\hat{I}^*[r])]}{\mathbb{E}[(D+\hat{I}^*[r])]}\\
&\qquad-\mathbb{E}[\hat{T}[r]|\mathbf{\Theta}[r]]\frac{V\mathbb{E}\left[\sum_{n=1}^{N}w_n\lambda_n\gamma_n^*[r](D+\hat{I}^*[r])\right]}{\mathbb{E}[(D+\hat{I}^*[r])]}
\end{aligned}$$

where $c^*[r]$, $m^*[r]$, $I^*[r]$, and $\gamma^*[r]$ are from any alternative (possibly randomized) decisions that can be made on frame r. According to Neely (2010), there exists for the algorithm making these decisions a $\delta>0$ that satisfies

$$\begin{aligned}
&\frac{\mathbb{E}[\lambda_n\gamma_n^*[r](D+\hat{I}^*[r])-\hat{\mu}_n(c^*[r],m^*[r])]}{\mathbb{E}[(D+\hat{I}^*[r])]}\le\delta\\
&\frac{\mathbb{E}[\hat{e}(c^*[r],m^*[r],I^*[r])]}{\mathbb{E}[(D+\hat{I}^*[r])]}\le P+\delta\\
&\frac{\mathbb{E}\left[\sum_{n=1}^{N}w_n\lambda_n\gamma_n^*[r](D+\hat{I}^*[r])\right]}{\mathbb{E}[(D+\hat{I}^*[r])]}\le OPT+\delta
\end{aligned}$$

Substituting these into the right-hand side of the previous drift expression yields

$$\begin{aligned}
&\mathbb{E}\left[\Delta[r]-V\sum_{n=1}^{N}w_nA_n[r]|\mathbf{\Theta}[r]\right]\\
&\quad\le B+\mathbb{E}[\hat{T}[r]|\mathbf{\Theta}[r]\left[\sum_{n=1}^{N}Q_n[r]\delta+Z[r]\delta-V(OPT+\delta)\right]
\end{aligned}$$

Taking a limit as $\delta\to0$ in the above yields

$$\mathbb{E}\left[\Delta[r]-V\sum_{n=1}^{N}w_nA_n[r]|\mathbf{\Theta}[r]\right]\le B-\mathbb{E}[\hat{T}[r]|\mathbf{\Theta}[r]]\times V\times OPT$$

Taking expectations of the above yields

$$\begin{aligned}
&\mathbb{E}\left[\Delta[r]-V\sum_{n=1}^{N}w_nA_n[r]\right]=\mathbb{E}[L[r+1]]-\mathbb{E}[L[r]]-V\mathbb{E}\left[\sum_{n=1}^{N}w_nA_n[r]\right]\\
&\quad\le B-\mathbb{E}[\hat{T}[r]]\times V\times OPT
\end{aligned}\tag{27}$$

Summing the above over $r\in0,\ldots,R-1$ for some integer $R>0$ and dividing by R yield

$$\frac{\mathbb{E}[L[R]]-\mathbb{E}[L[0]]}{R}-V\sum_{n=1}^{N}w_n\overline{A}_n[R]\le B-\overline{T}[R]\times V\times OPT$$

Rearranging terms in the above and using the facts that $\mathbb{E}[L[R]]\ge0$ and $\mathbb{E}[L[0]]=0$ yield the result of part (1).

(2) Suppose this is true for a particular frame R, we show that it also holds for $R+1$. According to (23), if $Q_n[R]\le Vw_n$, then $Q_n[R+1]\le Q_n[R]+A_n^{max}=Q_n[R]+\lambda^{max}(D+I_{max})$, because it can increase by at most $A_n^{max}=\lambda^{max}(D+I_{max})$ in one frame when $\gamma_n[R]=1$. If $Q_n[R]>Vw_n$, then $\gamma_n[R]=0$ and Q_n cannot increase in the frame R, and we have $Q_n[R+1]\le Q_n[R]$. This proves the bound of part (2).

(3) Note that (27) implies that for all r we have $\mathbb{E}[\Delta[r]]\le B-D\times V\times OPT+V\lambda^{max}(D+I_{max})\sum_{n=1}^{N}w_n$. This proves that all queues are mean rate stable by Theorem 4.1 in Neely (2010), so we have

$$\lim_{R\to\infty}\frac{\mathbb{E}[Q_n[R]]}{R}=0$$

$$\lim_{R\to\infty}\frac{\mathbb{E}[Z[R]]}{R}=0$$

Part (3) then follows from Lemmas 1 and 2. □

5. Performance evaluation

In this section, we evaluate the performance of our dynamic scheduling algorithm using simulations. Similar to those in Neely (2012b), we consider a network system with 10 classes of data flows and two processing modes. Traffic arrivals are from independent Bernoulli processes for each class $n\in\{1,\ldots,10\}$. We use weights $w_n=1$ for all n, so that the objective is to maximize total throughput, and $P=0.5$ W. The active period $D=50$ s, and the idle period $I\in[0,10]$ s. The energy consumption of the system in idle state is assumed to be neglectable, therefore $\hat{e}(c[r],m[r],I[r])=\hat{e}(c[r],m[r])$. According to Neely (2012b), the specific parameter settings on λ_n (Mbps), $\hat{\mu}_n$ (Mb) and $\hat{e}_n$ (J) are listed in Table 1.

In the first set of experiments, we simulate the dynamic scheduling algorithm for different values of the control parameter V over 1000 frames. In Fig. 3, we plot the time average admission rate achieved by the system over this period. It can be seen that the admission rate increases with V and converges to the optimal value (i.e., raw arrival rate in total) 1.56212 Mbps, with the difference exhibiting a $\mathcal{O}(1/V)$ behavior as predicted by Theorem 2. For comparison, we also simulate another random scheduling algorithm, which randomly makes the control decisions from available options. According to Theorem 2.4 in Neely (2010), we use a simple admission policy to enforce that $A_n[r]=0$ when $\overline{A}_n>\overline{\mu}_n$, so as to guarantee the rate stability of queue Q_n. We simulate out this random scheduling algorithm for 10 million times (to try out as many combinations of control decisions as possible), and finally find that the admission rates range between 1.18 Mbps and 1.47 Mbps, i.e., less than the results of our algorithm when $V>240$. The results show that when V is large enough, our

Table 1
Simulation parameters for λ_n, $\hat{\mu}_n$ and $\hat{e}_n$.

n	λ_n	Mode 1		Mode 2	
		$\hat{\mu}_n$	$\hat{e}_n$	$\hat{\mu}_n$	$\hat{e}_n$
1	0.5333	200	10	333	20
2	0.2667	100	9	166	18
3	0.1778	66	8	111	16
4	0.1333	50	7	83	14
5	0.1067	40	6	66	12
6	0.0889	33	5	55	10
7	0.0762	28	4	47	8
8	0.0667	25	3	41	6
9	0.0593	22	2	37	4
10	0.0533	20	1	33	2

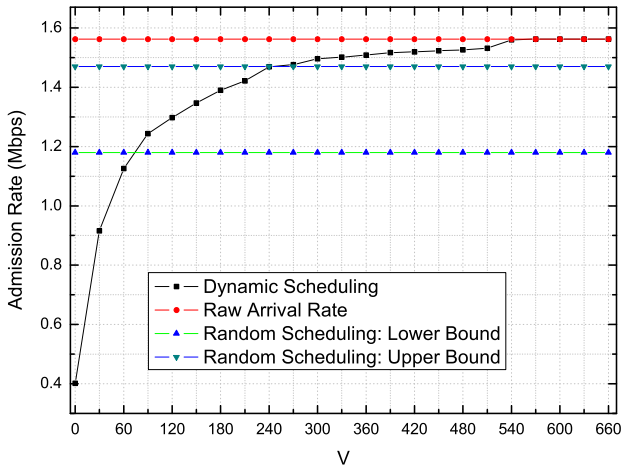

Fig. 3. Total admission rate versus V.

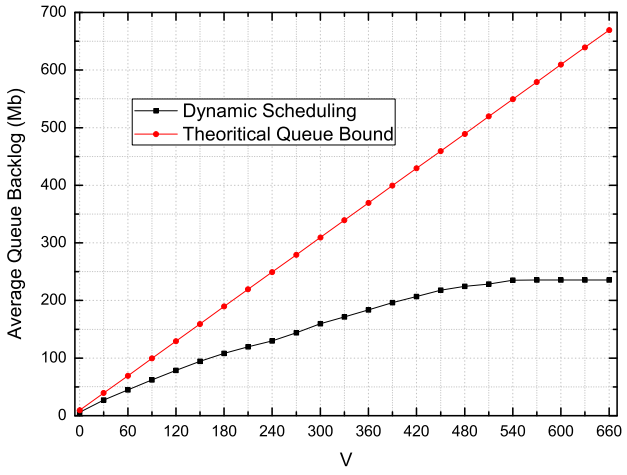

Fig. 4. Average queue backlog versus V.

algorithm can reliably guarantee a better performance on data admission through Lyapunov optimization techniques. In Fig. 4, we plot the average queue backlog of the system over 1000 frames, together with the theoretical bound given in Theorem 2. The average queue backlog grows linearly with V when $V<570$, as predicted by Theorem 2. When $V\geq 570$, the queues saturate by admitting everything. The saturation value is the average queue backlog associated with admitting the raw arrival rates directly. These results show that, by appropriately tuning the parameter V, we can achieve a desired $[\mathcal{O}(1/V), \mathcal{O}(V)]$ tradeoff between the admission goal and the queue lengths.

In the second set of experiments, we examine whether the constraint on energy consumption can be guaranteed in the simulation. According to Fig. 5, it is obvious that our algorithm can strictly guarantee that the average power is well below P on any frame. Besides, the average power is much lower when V is relatively larger and system optimization is emphasized more by our algorithm. That is because the power constraint in (14) is satisfied by stabilizing queue Z in our algorithm, i.e., choosing a relatively longer idle time $I[r]$ when $Z[r]>0$. When V is large, Z becomes less stable while large values of $I[r]$ would be chosen by the algorithm more frequently. Since the idle energy is neglected in energy calculation, the average power will decrease with the growth of V and $I[r]$.

In the third set of experiments, we illustrate that the dynamic scheduling algorithm is robust to abrupt rate changes. The simulation is carried out over 1500 frames with $V=570$. However, the total timeline is broken into three phases. During the course of the simulation, we double the rates λ_n for each n after the first 350 frames and then again to the original rates after the first 700 frames. In Figs. 6 and 7, we plot the moving average (over 100 frames) of the total admission rate and the average queue backlog, compared with those of unchanged rates. These results show that our algorithm automatically adapts to the changes in λ_n. For Fig. 6, we carefully examine the simulation traces, and find that: when $r<350$, the data flows are all admitted into the system. However, when $r>350$ and all λ_n are doubled, the abrupt rate changes significantly increase the backlog of queues. Thus, the system frequently declines data admission of the queues with lower service rate (e.g., Q_8, Q_9 and Q_{10}) to avoid congestion, which makes the average admitted rates 16.35–22.04% lower than those that should be in a system with doubled loading always. However, when $r>700$ and all λ_n falls back, some congested queues (i.e., Q_9 and Q_{10}) still have few chance to be served and then continue to decline any arrival, which makes the average admitted rates 3.41–22.04% lower than the original value for about 180 frames. In Fig. 7, we can notice that the deterministic queue bound is 579.373 Mb, which holds for all frames, regardless of the raw arrival rates.

In the fourth set of experiments, we investigate how the weights w_n for each $n\in\{1,\ldots,N\}$ could impact the control decisions and optimization results. To make the results more comparable and clear, we only start two data flows (i.e., $n=1$ and $n=2$).

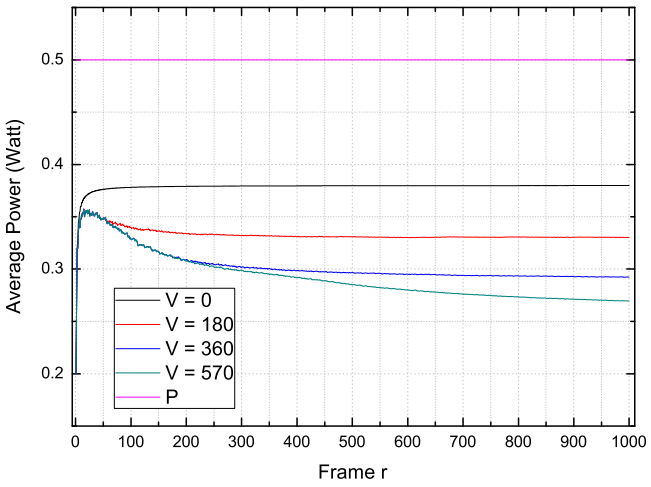

Fig. 5. Average power versus frame index r under different V.

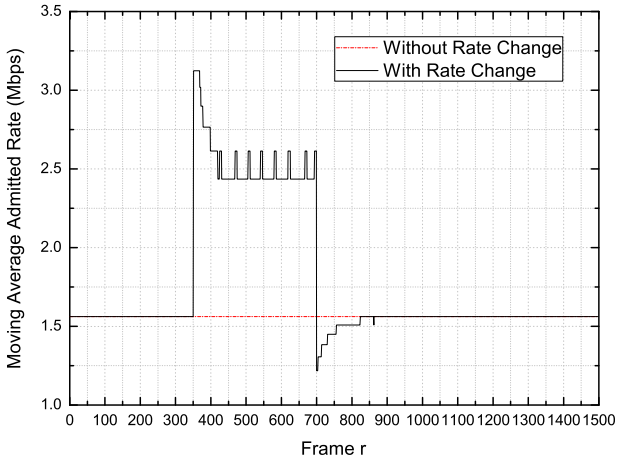

Fig. 6. Moving average admission rate versus frame index r for the system with abrupt rate changes.

Both of them adopt the parameter settings on λ_n, $\hat{\mu}_n$ and $\hat{e}_n$ of class 2 in Table 1. However, they have different weights, i.e., $w_2 = 2w_1$. Thus, the algorithm can only choose to serve either one of them. From Fig. 8, we know that w_n will have impacts on the control decisions and the individual admissions when V is relatively larger (e.g., V=20 in our simulations), and the queues will be served differentially according to their weights. That is because V is a parameter that weights the extent to which optimization goal is emphasized as compared to system stability.

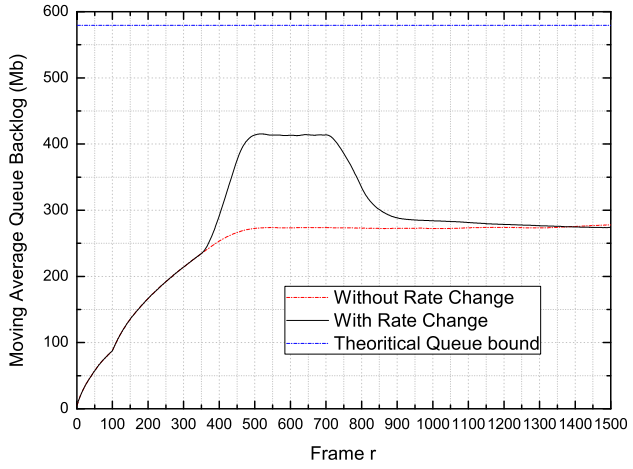

Fig. 7. Moving average queue backlog versus frame index r for the system with abrupt rate changes.

6. Related work

6.1. Admission control

Admission control is an important validation process for effective and efficient resource utilization in communication systems. A most well-known application is the call admission control used in VoIP networks to avert traffic congestion and ensures that there is enough bandwidth for authorized flows (Naghshineh and Schwartz, 1996). Earlier researches mainly concentrate on how to characterize the traffic to design efficient control approach for single-class, single-queue ATM networks (Perros and Elsayed, 1996). With the growing interest of data and multimedia services, research interests have been shifted from single-class schemes to multiple-class schemes (Ahmed, 2005). The design of multiple-class schemes is more challenging because some critical issues, such as service prioritization (Haung and Ho, 2002), fairness guarantee (Mosharaf et al., 2005), power control (Saad, 2011) and mobile characteristics (Both et al., 2012; Haider et al., 2013), must be considered. Some recent admission schemes have taken energy consumption into consideration. In Wang et al. (2005), an energy-based connection admission

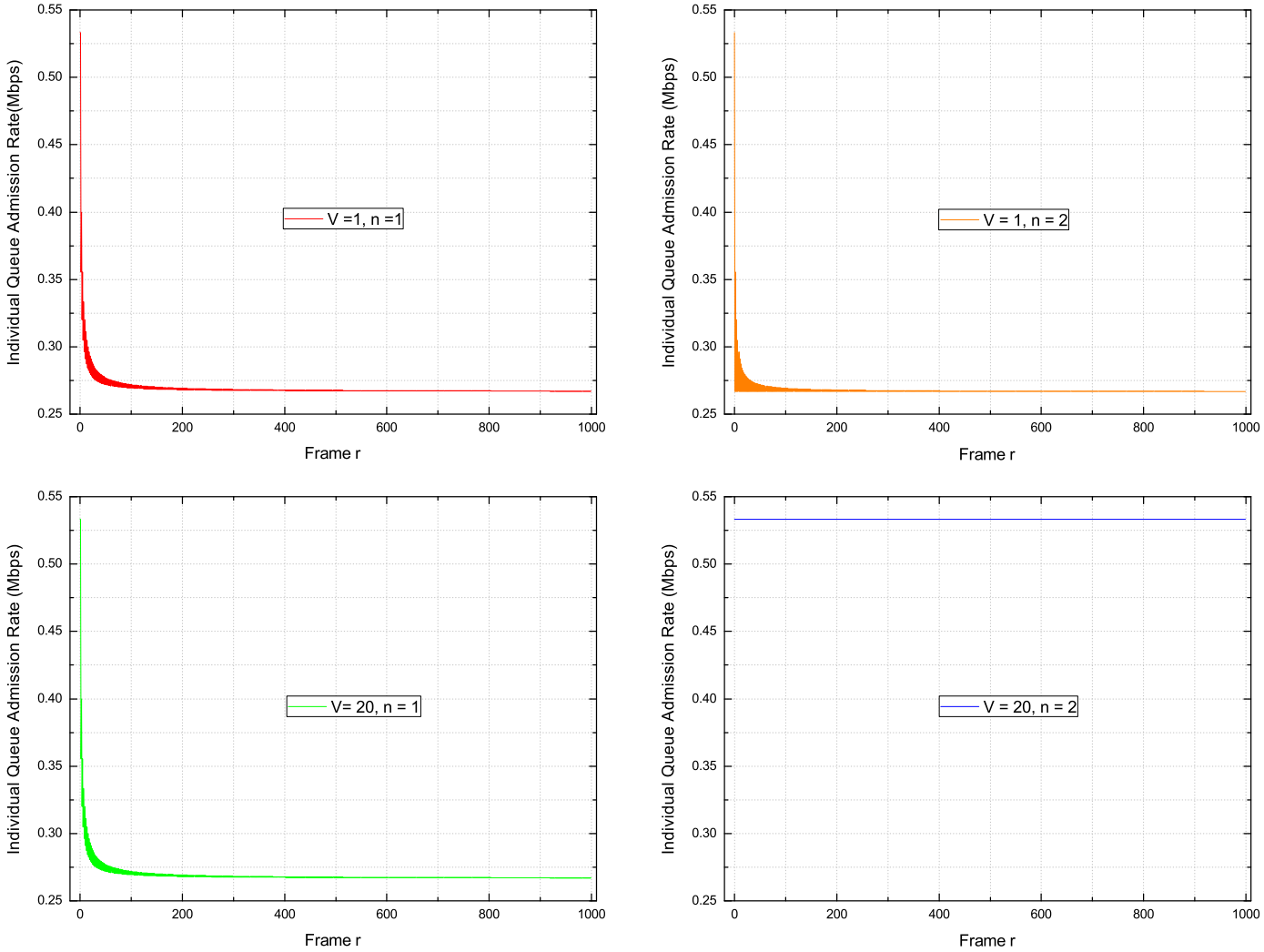

Fig. 8. Individual queue admission rate for queues with different weights w_n.

control scheme is proposed to adjust bandwidth utilization and minimize energy consumption of mobile terminals. This work introduces a new parameter (energy consumption rate), and proposes an adaptive admission control scheme for rate reduction, bandwidth reallocation and QoS guarantee based on energy consumption. However, this scheme requires statistical knowledge of system dynamics. Besides, it did not take sleep scheduling for energy saving into account. In Zuger and Labit (2013), the authors propose to send and admit packets in the optimal period during which the link energy consumption is small and the waiting delay added is acceptable. However, this scheme can only be used in the energy-efficient ethernet with Adaptive Link Rate (ALR) functions, which has not been mature yet by far. In Morfopoulou et al. (2013), admission control is performed in wired networks based on the energy demanding of user's tasks. It not only delays the users potentially having large energy demanding, but also turns off idle machines into sleep state for energy saving. However, the authors did not take QoS issues into consideration. As far as we know, no previous work has comprehensively considered optimizing admission rates in renewal network systems (Neely, 2012b), e.g., duty-cycled sensor networks (Keshavarzian et al., 2006).

6.2. Lyapunov optimization

Lyapunov optimization (Neely, 2010) is a newly developed technique for solving problems of joint system stability and performance optimization in queuing systems and stochastic networks. The basic idea is to make control actions that greedily minimize a bound on the drift-plus-penalty expression over fixed-length time slots, so as to optimize the time averages of certain quantities. It does not require knowledge of the statistics of related stochastic models, and has a good computational complexity. By now, this new technique has been applied in solving many stochastic network optimization problems, including workload/resource scheduling among data centers (Yao et al., 2012; Zhou et al., 2013), power management in smart grid (Urgaonkar et al., 2011; Guo et al., 2012), and energy/throughput optimization for wireless systems (Ra et al., 2010; Neely, 2011). However, the drift-plus-penalty rule cannot be used for renewal systems with variable frame lengths. The work in Neely (2012b) extends this to a generalization, called drift-plus-penalty ratio rule, to allow optimization over renewal systems. This new rule tries to optimize the ratio of the drift-plus-penalty expression and the expected frame size. It has already been formally used in for solving problems in energy-constrained wireless systems, e.g., RF chain sleeping scheduling for energy saving in multi-user MIMO systems (Zhang et al., 2013b) and intelligent cooperation for throughput maximization of secondary users in cognitive femtocell networks (Urgaonkar and Neely, 2012). Note that these problems have different optimization objectives and constraints as compared to ours.

7. Conclusion

In this paper, we studied the problem of optimal admission control for an energy-constrained network system with multiple data queues. Traditional solutions either require extensive knowledge of the system dynamics or suffer from large convergence times. However, using the Lyapunov optimization technique for renewal systems, we designed a new online control algorithm that overcomes these challenges. Both mathematical analyses and simulation evaluations demonstrated the optimality, stability and robustness of this algorithm. Especially, it can approach the offline optimal admission rate within a diminishing gap of $\mathcal{O}(1/V)$ while bounding the queue backlog by $\mathcal{O}(V)$, where V is a tunable control parameter. This algorithm can be easily applied to a variety of network systems with renewal characteristics, including duty-cycled sensor networks and new-generation mobile networks.

In the future, we would like to investigate the robustness of our algorithm to queue backlog estimation errors both analytically as well as through experiments (Yao et al., 2012). We also plan to solve the extended problem of optimizing a sum of concave utility functions of the time average admission rates. Another important work is to further evaluate the algorithm performance in our sensor network test-bed.

Acknowledgments

This work was supported in part by the National Science Foundation of China (NSFC) under Grants 61202430 and 61303245, the State Key Lab of Astronautical Dynamics of China under Grant 2014ADL-DW0401, and the Science and Technology Foundation of Beijing Jiaotong University under Grant 2012RC040.

References

Ahmed M. Call admission control in wireless networks: a comprehensive survey. IEEE Commun Surv Tutor 2005;7(1):49–68.
Both CB, Marquezan CC, Kunst R, Granville LZ, Rochol J. A self-adapting connection admission control solution for mobile wimax: enabling dynamic switching of admission control algorithms based on predominant network usage profiles. J Netw Comput Appl 2012;35(5):1392–401.
Boyd S, Vandenberghe L. Convex optimization. New York, NY, USA: Cambridge University Press; 2004.
Deng S, Balakrishnan H. Traffic-aware techniques to reduce 3g/lte wireless energy consumption. In: Proceedings of the eighth international conference on emerging networking experiments and technologies (CoNEXT'12). New York, NY, USA: ACM; 2012. p. 181–92.
Gallager RG. Discrete stochastic processes. Boston: Kluwer Academic Publishers; 1996.
Guo Y, Pan M, Fang Y. Optimal power management of residential customers in the smart grid. IEEE Trans Parallel Distrib Syst 2012;23(9):1593–606.
Haider A, Gondal I, Kamruzzaman J. Social-connectivity-aware vertical handover for heterogeneous wireless networks. J Netw Comput Appl 2013;36(4):1131–9.
Haung YR, Ho JM. Distributed call admission control for a heterogeneous pcs network. IEEE Trans Comput 2002;51(12):1400–9.
Keshavarzian A, Lee H, Venkatraman L. Wakeup scheduling in wireless sensor networks. In: Proceedings of the seventh ACM international symposium on mobile ad hoc networking and computing (MobiHoc '06). New York, NY, USA: ACM; 2006. p. 322–33.
Morfopoulou C, Sakellari G, Gelenbe E. Energy-aware admission control for wired networks. In: Proceedings of the twenty-eighth international symposium on computer and information sciences. Paris, France: Springer International Publishing; 2013. p. 117–25.
Mosharaf K, Lambadaris I, Talim J. A call admission control for service differentiation and fairness management in WDM grooming networks. Opt Switch Netw 2005;2(2):113–26.
Naghshineh M, Schwartz M. Distributed call admission control in mobile/wireless networks. IEEE J Sel Areas Commun 1996;14(4):711–7.
Neely M. Opportunistic scheduling with worst case delay guarantees in single and multi-hop networks. In: Proceedings of IEEE INFOCOM 2011, Shanghai, China; 2011. p. 1728–36.
Neely M. Asynchronous scheduling for energy optimality in systems with multiple servers. In: 2012 46th annual conference on information sciences and systems (CISS), Princeton, NJ, USA: 2012a. p. 1–6.
Neely M. Green communications and networking. Boca Raton, FL, USA: CRC Press; 2012b. p. 231–72.
Neely M. Dynamic optimization and learning for renewal systems. IEEE Trans Autom Control 2013;58(1):32–46.
Neely MJ. Stochastic network optimization with application to communication and queueing systems. CA, USA: Morgan & Claypool; 2010.
Niu J, Song W, Atiquzzaman M. Bandwidth-adaptive partitioning for distributed execution optimization of mobile applications. J Netw Comput Appl 2014;37:334–47.
Perros H, Elsayed K. Call admission control schemes: a review. IEEE Commun Mag 1996;34(11):82–91.
Ra MR, Paek J, Sharma AB, Govindan R, Krieger MH, Neely MJ. Energy-delay tradeoffs in smartphone applications. In: Proceedings of the eighth international conference on mobile systems, applications, and services, New York, NY, USA; 2010. p. 255–70.
Saad M. Joint admission and power control for quality-of-service in the wireless downlink. J Netw Comput Appl 2011;34(2):644–52.

Urgaonkar R, Neely M. Opportunistic cooperation in cognitive femtocell networks. IEEE J Sel Areas Commun 2012;30(3):607–16.

Urgaonkar R, Urgaonkar B, Neely MJ, Sivasubramaniam A. Optimal power cost management using stored energy in data centers. In: Proceedings of the ACM SIGMETRICS joint international conference on measurement and modeling of computer systems, New York, NY, USA; 2011. p. 221–32.

Wang H, Agrawal DP. A weight-based adaptive call admission control scheme for integrated multimedia traffic in mobile wireless networks. In: Proceedings of the 2004 international conference on parallel processing workshops (ICPPW'04). Washington, DC, USA: IEEE Computer Society; 2004. p. 308–13.

Wang W, Wang X, Nilsson A. A new admission control scheme under energy and qos constraints for wireless networks. In: 24th Annual joint conference of the IEEE computer and communications societies (INFOCOM 2005). Proceedings IEEE, vol. 2; 2005. p. 1283–94.

Yao Y, Huang L, Sharma A, Golubchik L, Neely M. Data centers power reduction: a two time scale approach for delay tolerant workloads. In: Proceedings IEEE INFOCOM 2012, Orlando, FL, USA; 2012. p. 1431–9.

Zhang H, Shu L, Rodrigues JJ, Chao Hc. Solving network isolation problem in duty-cycled wireless sensor networks. In: Proceeding of the 11th annual international conference on mobile systems, applications, and services (MobiSys'13). New York, NY, USA: ACM; 2013a. p. 543–4.

Zhang X, Zhou S, Niu Z, Lin X. An energy-efficient user scheduling scheme for multiuser mimo systems with RF chain sleeping. In: Wireless communications and networking conference (WCNC), 2013. Shanghai, China: IEEE; 2013b. p. 169–74.

Zhou Z, Liu F, Jin H, Li B, Li B, Jiang H. On arbitrating the power-performance tradeoff in saas clouds. In: Proceedings of IEEE INFOCOM 2013, Turin, Italy; 2013. p. 872–80.

Zuger H, Labit Y. An energy aware admission control with traffic class differentiation: from theory to ns-2 simulation. In: Proceedings of the seventh international conference on performance evaluation methodologies and tools, Turin, Italy; 2013. p. 1–8.