Wireless Scheduling for Network Utility Maximization with Optimal Convergence Speed

Bin Li, Atilla Eryilmaz and Ruogu Li

Abstract—**In this paper, we study the design of joint flow rate control and scheduling policies in multi-hop wireless networks for achieving maximum network utility *with provably optimal convergence speed*. Fast convergence is especially important in wireless networks which are dominated by the dynamics of incoming and outgoing flows as well as the time sensitive applications. Yet, the design of fast converging policies in wireless networks is complicated by: (i) the interference-constrained communication capabilities, and (ii) the *finite* set of transmission rates to select from due to operational and physical-layer constraints.**

We tackle these challenges by explicitly incorporating such *discrete* constraints to understand their impact on the convergence speed at which the running average of the received service rates and the network utility converges to their limits. In particular, we establish a fundamental fact that the convergence speed of any feasible policy cannot be faster than $\Omega\left(\frac{1}{T}\right)$ under both the rate and utility metrics. Then, we develop an algorithm that achieves this optimal convergence speed in both metrics. We also show that the well-known dual algorithm can achieve the optimal convergence speed in terms of its utility value.

These results reveal the interesting fact that the convergence speed of rates and utilities in wireless networks is dominated by the discrete choices of scheduling and transmission rates, which also implies that the use of higher-order flow rate controllers with fast convergence guarantees cannot overcome the aforementioned fundamental limitation.

I. Introduction

Wireless networks are expected to serve users in an efficient and fair way, which requires careful flow control and interference management among simultaneous transmissions. These design goals can be achieved by solving the Network Utility Maximization (NUM) problem, where the utility of *long-term average* flow rates is maximized under stability constraints. Moreover, it is desired that these optimal solutions are reached rapidly due to the dynamic nature of flows and the time-sensitive nature of many wireless applications.

However, the design of controllers with *fast* convergence speed in most wireless networks is complicated by two natural constraints: (i) interference constraints leading to *discrete* link scheduling choices; and (ii) a *finite* set of choices for the transmission rate selection over the scheduled links. The latter constraint is caused by both digital communication (e.g., modulation, coding, etc.) and hardware design principles. For example, in IEEE 802.11b standard, there are four transmission rates: 1Mpbs, 2Mpbs, 5.5Mpbs and 11Mbps.

Bin Li, Atilla Eryilmaz, and Ruogu Li ({lib, eryilmaz, lir}@ece.osu.edu) are with the Department of Electrical and Computer Engineering at The Ohio State University, Columbus, Ohio 43210 USA.

This work is supported by the NSF grants: CAREER-CNS-0953515 and CCF-0916664. Also, the work of A. Eryilmaz was in part supported by the QNRF grant number NPRP 09-1168-2-455

Previous works mainly focus on the design and analysis of policies with optimal *limiting* behavior. A large body of works (e.g. [1], [2], [3], [4], [5], [6]) has utilized *dual* and *primal-dual* methods to develop cross-layer policies with long-term optimality guarantees. Such solutions are amenable to distributed implementation due to their natural decomposition into loosely coupled components. However, being *first-order* methods, they suffer from the slow convergence speed shared by all such methods (e.g. [7], [8], [9]).

This speed deficiency of dual methods has recently spurred an exciting thread of research activity in the design of distributed Interior Point (e.g. [10]) and Newton's (e.g. [11], [12], [13]) methods for network utility maximization. However, these works do not incorporate two aforementioned features of wireless networks, namely the discreteness in the scheduling and transmission rate selections. We explicitly incorporate these intrinsic characteristics of wireless networks in our analysis and algorithm design. To the best of our knowledge, this is the first work that systematically analyzes and designs algorithms in terms of their converge speed in wireless networks with such discrete constraints. Next, we list our main contributions, along with references on where they appear in the text.

- We show that the convergence speed[1] at which the running average of the received service rates (see Section III-A) and their utility (see Section III-A) over T time slots cannot be faster than $\Omega\left(\frac{1}{T}\right)$. This fundamental limitation on the convergence speed is due to the discreteness of the transmission rates.
- We develop a generic algorithm that can work with a range of flow rate controllers, and achieves the optimal convergence speed in both rate (see Section III-B) and utility (see Section IV-B) metrics.
- Somewhat surprisingly, we also show that even a first-order method such as the well-known dual algorithm can achieve the aforementioned optimal convergence speed in terms of its utility value (see Section IV-C).
- These results collectively reveal that, under wireless networks subject to discrete scheduling and rate constraints, the convergence speed of cross-layer algorithms is dominated by the convergence speed of the *scheduling* component, and not the flow rate controller. As such, the speed improvements in the flow rate convergence, unfortunately, cannot extend to the received service rates or utilities in wireless networks. On the bright side, however, with careful design we can achieve the optimal convergence speed under such constraints.

[1]The following standard notations are used to describe the rates of growth of two real-valued sequences $\{a_n\}$ and $\{b_n\}$:$a_n = O(b_n)$ if $\exists c > 0$ such that $|a_n| \le c|b_n|$; $a_n = \Omega(b_n)$ if $b_n = O(a_n)$.

II. System Model

We consider a multi-hop fading wireless network with a set $\mathcal{L} = \{1, 2, ..., L\}$ of links that operates in a time-slotted fashion, where all links transmit at the beginning of each time slot subject to interference constraints. Due to modulation, coding, as well as other practical constraints, each link has to transmit at one of a **finite** set of rates[2]. We use $\mathbf{S}[t] = (S_l[t])_{l=1}^L$ to denote the service rate vector offered to the links in slot t, which must be selected from a **feasible** set of transmission rates at the time. The feasible set, in turn, depends on the *network fading state* and the interference constraints amongst the links. Using $\mathcal{J}$ to denote the set of global channel states (with finite cardinality), we let $\mathcal{S}^j$ denote the set of *feasible service rate vectors* when the channel is in state $j \in \mathcal{J}$. We assume that the fading process is stationary and ergodic with π_j denoting the stationary probability of the channel state being in state j. Then, the *capacity region* can be defined as

$$\mathcal{R} \triangleq \sum_{j \in \mathcal{J}} \pi_j \cdot \text{CH}\{\mathcal{S}^j\}, \tag{1}$$

where $\text{CH}\{\mathcal{A}\}$ denotes the convex hull of the set $\mathcal{A}$. We note that $\mathcal{R}$ is a polyhedron due to the discreteness of the transmission rate choices, and hence can be written as $\mathcal{R} = \{\mathbf{y} \geq 0 : \mathbf{H}\mathbf{y} \leq \mathbf{b}\}$, where $\mathbf{y} \in \mathbb{R}^L$ and $\mathbf{H}$ is some positive matrix. Note that $\mathbf{H}$ has L columns and the number of rows in $\mathbf{H}$ is equal to the dimension of $\mathbf{b}$ associated with the number of interference constraints. As a special case, when $|\mathcal{J}| = 1$ we obtain the *non-fading* scenario.

To capture the heterogeneous and potentially inter-dependent preferences of users, we define a utility function $U : \mathbb{R}_+^L \to \mathbb{R}_+$ that measures the total network utility when link l receives an average service rate of r_l, where $\mathbf{r} = (r_l)_{l=1}^L$. We assume that $U(\mathbf{r})$ to be a strictly concave function that is non-decreasing in each coordinate. The objective of Network Utility Maximization (NUM), then, is to design a congestion control and scheduling algorithm such that the average service rate vector $\mathbf{r}$ solves the following optimization problem:

Definition 1: (**Network Utility Maximization (NUM)**)

$$\max_{\mathbf{r}=(r_l)_{l=1}^L} \quad U(\mathbf{r}) \tag{2}$$

$$\text{Subject to} \quad \mathbf{r} \in \mathcal{R}, \tag{3}$$

where $\mathcal{R}$ is defined in (1). ⋄

The strict concavity of $U(\cdot)$ together with the convexity of $\mathcal{R}$ guarantees the uniqueness of the solution of NUM, which is denoted as $\mathbf{r}^* = (r_l^*)_{l=1}^L$. Also, due to the non-decreasing nature of $U(\cdot)$, $\mathbf{r}^*$ must lie on the boundary of $\mathcal{R}$.

It is important to note that $\mathbf{r}^*$ is the optimal average **offered service rate** to the links. The purpose of the flow rate controller, however, is to determine the optimal **injection rate** of traffic into the network while maintaining network stability. To define network stability more rigorously, let $Q_l[t]$ denote the queue-length at link $l \in \mathcal{L}$ at the beginning of slot t, let $X_l[t]$ denote the amount of injected data into Queue-l in slot t under a given *flow rate controller*, and recall that $S_l[t]$ denotes the service rate offered to link l in slot t under a given *scheduler*. Then, the evolution of Q_l can be expressed as

$$Q_l[t+1] \quad = \quad (Q_l[t] + X_l[t] - S_l[t])^+, \qquad t \geq 1,$$

where $(y)^+ \triangleq \max\{0, y\}$, and Queue-$l$ is said to be *stable* if

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Q_l[t]] < \infty, \tag{4}$$

and the *network is stable* if all queues are stable.

In this work, we are interested in the convergence speed of a broad class of joint flow rate control and scheduling policies $\mathcal{P}$ that are both **stabilizing** and **asymptotically rate optimal**. To define this class of policies abstractly, we introduce the parameter $\epsilon > 0$ as a generic term to characterize the performance of the joint policy under specific design choices. Accordingly, the average injection rate of a given policy under parameter ϵ lies in a set $\overline{\mathcal{X}}^{(\epsilon)}$. Similarly, we will use the superscript $\cdot^{(\epsilon)}$ over $(Q_l[t])_l, (X_l[t])_l, (S_l[t])_l$, etc. to express the queue-lengths, injections, offered service rates, etc. under the policy with parameter ϵ. The *stability condition* requires that $\overline{\mathcal{X}}^{(\epsilon)} \subset \mathcal{R}$ for all $\epsilon > 0$, and the *asymptotic rate optimality condition* requires that $\lim_{\epsilon \downarrow 0} \overline{\mathcal{X}}^{(\epsilon)} = \{\mathbf{r}^*\}$, i.e., the asymptotically optimal policy achieves the optimal service rate vector in the limit. Thus, the parameter ϵ captures the closeness of the injection rate to the optimal service rate $\mathbf{r}^*$ under the class of joint policies parametrized by ϵ. We note that this abstraction includes a wide range of joint control and scheduling policies in the literature. For example, in the well-known subgradient-based designs (e.g., [1], [2], [3], [4], [5]) the generic term ϵ maps to the particular design parameter that corresponds to the step-size on the subgradient iteration.

The stability condition of the joint flow rate control and scheduling policies in $\mathcal{P}$ implies that the running average of **departures** over time must also converge to the set[3] $\overline{\mathcal{X}}^{(\epsilon)}$. Since the running average of departures[4] up to time T is the real measure of **received** service until that time, we are interested in its convergence speed to $\overline{\mathcal{X}}^{(\epsilon)}$. To be more precise, for the policy with parameter ϵ we use $D_l^{(\epsilon)}[t] \triangleq \min(S_l^{(\epsilon)}[t], Q_l^{(\epsilon)}[t])$ to denote the departures in slot t for link $l \in \mathcal{L}$, and define its *running average* until $T \geq 1$ as

$$\overline{d}_l^{(\epsilon)}[T] \triangleq \frac{1}{T} \sum_{t=1}^{T} D_l^{(\epsilon)}[t], \qquad \forall l \in \mathcal{L}, \tag{5}$$

and use $\overline{\mathbf{d}}^{(\epsilon)}[T] \triangleq (\overline{d}_l^{(\epsilon)}[T])_l$. Next, we introduce the metrics of interest in our study of convergence speed, both in the running

[2]For example, IEEE 802.11a standard uses OFDM transmission technique and can support rates in Mega bits per second selected from the finite set {6, 9, 12, 18, 24, 36, 48, 54}; In CDMA2000 1xEV-DO specification, the forward link transmission rate in kilo bits per second is chosen from the finite set $\{38.4, 76.8, 153.6, 307.2, 614.4, 921.6, 1228.8, 1843.2, 2457.6\}$.

[3]Note that the convergence of a sequence to a set is the convergence of its minimum distance to the set.

[4]Due to the discreteness of transmission rate choices, it is unlikely that the departure rate in slot T converges to the set $\overline{\mathcal{X}}^{(\epsilon)}$, as T increases.

average departure rate and its corresponding utility value.

Definition 2: (**Metrics of Interest**) For any policy in $\mathcal{P}$ with parameter ϵ, we define the *rate deviation* $\phi(\overline{\mathbf{d}}^{(\epsilon)}[T], \overline{\mathcal{X}}^{(\epsilon)})$ between $\overline{\mathbf{d}}^{(\epsilon)}[T]$ and the set $\overline{\mathcal{X}}^{(\epsilon)}$ at time T as

$$\phi(\overline{\mathbf{d}}^{(\epsilon)}[T], \overline{\mathcal{X}}^{(\epsilon)}) \triangleq \inf_{\overline{\mathbf{x}}^{(\epsilon)} \in \overline{\mathcal{X}}^{(\epsilon)}} \left\| \overline{\mathbf{d}}^{(\epsilon)}[T] - \overline{\mathbf{x}}^{(\epsilon)} \right\|, \qquad (6)$$

and the *utility benefit received until time* T as $U(\overline{\mathbf{d}}^{(\epsilon)}[T])$, where $\|\mathbf{y}\|$ is the l_2 norm of the vector $\mathbf{y}$. ◇

In the rest of paper, we will: (i) establish fundamental limits on the speed at which $\mathbb{E}[\phi(\overline{\mathbf{d}}^{(\epsilon)}[T], \overline{\mathcal{X}}^{(\epsilon)})]$ converges to zero as T increases (Section III-A); (ii) develop joint flow control and scheduling policy with provably optimal convergence speed in terms of rate deviation (Section III-B); (iii) derive fundamental limits on the speed at which the utility benefit converges to the optimal utility value of NUM when sources of randomness are eliminated (Section IV-A); and finally (iv) show that our proposed algorithm, as well as the well-known dual algorithm, achieves the optimal convergence speed in terms of utility benefit (Section IV-B,IV-C).

III. Convergence Speed in Rate Deviation

In this section, we study the optimal convergence speed in terms of rate deviation over wireless fading channels. To that end, we first give the fundamental lower bound on the expected rate deviation for any algorithm. Then, we provide an algorithm that can achieve this lower bound and establish the optimality of the proposed algorithm.

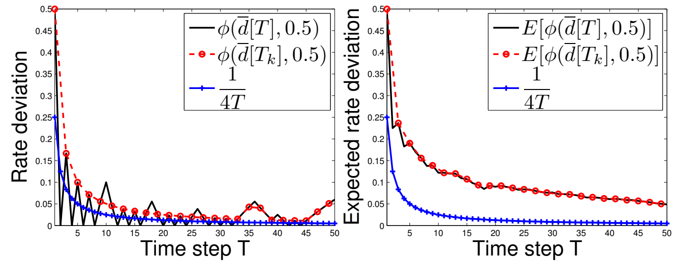
A. A lower bound on the expectation of rate deviation

In this subsection, we show that for any policy in $\mathcal{P}$, the convergence speed of expected rate deviation is $\Omega\left(\frac{1}{T}\right)$. To that end, we need the following integer assumption on the transmission rate, which measures the number of packets that can be transmitted in one time unit.

Assumption 1: Each feasible service rate vector $\mathbf{S}[t]$ is nonnegative-integer-valued and bounded with $S_l[t] \leq B$ for some positive integer number B, that is, each link can service i packets depending on the channel fading condition and the interference constraints, where $i = 0, 1, 2, ..., B$. ◇

Before introducing our main results, we provide an example that will give us some insights. Consider an independently and identically distributed (i.i.d.) Bernoulli random departure rate sequence with mean 0.5. Figure 1a shows one realization of this random departure rate sequence. From this figure, we can see that $\phi(\overline{d}[T], 0.5)$ hits 0 for some T, and is always non-zero when T is odd. The subsequence $\{\phi(\overline{d}[T_k], 0.5) : T_k \text{ is odd}\}$ is always lower-bounded by $\frac{1}{4T_k}$. In fact, all sample paths of subsequence $\{\phi(\overline{d}[T_k], 0.5) : T_k \text{ is odd}\}$ are always bounded below by $\frac{1}{4T_k}$, as we will shown in Proposition 1. This key fact also implies that the convergence speed of the subsequence $\{\mathbb{E}[\phi(\overline{d}[T_k], 0.5)] : T_k \text{ is odd}\}$ cannot be faster than $\frac{1}{4T_k}$, as shown in Figure 1b.

This simple example suggests that the discreteness in the choice of elements in the sequence exerts a fundamental limitation on speed with which its running average over time can approach its limit. In what follows, we will show that this observation indeed holds even in the wider context of a multi-hop fading wireless network with a finite selection of transmission rates.



(a) Convergence for one sample path (b) Convergence in the mean sense

Fig. 1: The convergence speed of an i.i.d Bernoulli sequence.

Next, we give the following key lemma, which will also be useful in the later section.

Lemma 1: Let $\mathcal{I} \triangleq \{a_1, a_2, ..., a_K\}$, where $0 \leq a_1 < a_2 < ... < a_K$ and K is some positive integer. If $b \in (a_i, a_{i+1})$ for some $i = 1, ..., K-1$, then for any sequence $\{I[t] : I[t] \in \mathcal{I}\}_{t \geq 1}$, there exists a constant $c_b \in (0, \min\{\frac{b-a_i}{2}, \frac{a_{i+1}-b}{2}\})$ such that if $|\frac{1}{T}\sum_{t=1}^{T} I[t] - b| \leq \frac{c_b}{T}$, then

$$\left| \frac{1}{T+1} \sum_{t=1}^{T+1} I[t] - b \right| \geq \frac{c_b}{T+1}. \qquad (7)$$

Remark: Note that K can be as large as ∞.

Proof: See Appendix A for the proof. ■

Proposition 1: Under Assumption 1, for any policy in $\mathcal{P}$ with parameter ϵ, if the closure of set $\overline{\mathcal{X}}^{(\epsilon)}$ does not contain a vector with all integer-valued coordinates, then the convergence speed of the expected rate deviation to zero is $\Omega\left(\frac{1}{T}\right)$, i.e., there exists a strictly positive constant c and a positive integer-valued sequence $\{T_k\}_{k=1}^{\infty}$ such that

$$\phi(\overline{\mathbf{d}}^{(\epsilon)}[T_k], \overline{\mathcal{X}}^{(\epsilon)}) \geq \frac{c}{T_k}, \qquad \forall k \geq 1, \qquad (8)$$

holds for any sample path of departure rate vector sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t \geq 1}$, which also implies that

$$\mathbb{E}[\phi(\overline{\mathbf{d}}^{(\epsilon)}[T_k], \overline{\mathcal{X}}^{(\epsilon)})] \geq \frac{c}{T_k}, \qquad \forall k \geq 1. \qquad (9)$$

Remark: If $\mathbf{r}^*$ has at least one non-integer-valued coordinate, then the condition for Proposition 1 holds when ϵ is sufficiently small. Moreover, since the region $\mathcal{R}$ is compact, there are finitely many rate vectors with all coordinates being integer in $\mathcal{R}$. Thus, Proposition 1 holds in almost all cases.

Proof: See Appendix B for the proof. ■

Proposition 1 indicates that the discrete structure of the transmission rates intrinsically limits the convergence speed for any algorithm in class $\mathcal{P}$. Thus, the search for higher-order numerical optimization methods cannot overcome this fundamental limitation in wireless networks. Despite pessimism of this observation, we are still interested in designing an algorithm that can achieve this fundamental bound and establish the optimality of this algorithm in terms of its convergence speed. To that end, we define the rate deviation optimality for an algorithm in class $\mathcal{P}$.

Definition 3: (**Rate Deviation Optimality**) An algorithm in class $\mathcal{P}$ with parameter ϵ is called *rate deviation optimal*, if its departure rate vector sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\geq 1}$ satisfies

$$\mathbb{E}[\phi(\overline{\mathbf{d}}^{(\epsilon)}[T], \overline{\mathcal{X}}^{(\epsilon)})] \leq \frac{F_1^{(\epsilon)}}{T}, \qquad \forall T \geq 1, \tag{10}$$

where $F_1^{(\epsilon)}$ is a positive constant and $\overline{\mathbf{d}}^{(\epsilon)}[T]$ is defined in (5).

Next, we propose an algorithm with rate deviation optimality.

B. A Rate Deviation Optimal Policy

In this subsection, we propose a rate deviation optimal algorithm with parameter $\epsilon > 0$ that converges to the injection rate $\overline{\mathbf{x}}^{(\epsilon)}$ solving the following optimization problem.

Definition 4: (ϵ-NUM)

$$\max_{\mathbf{r}=(r_l)_{l=1}^L} \quad U(\mathbf{r}) \tag{11}$$

$$\text{Subject to} \quad \mathbf{r} \in \mathcal{R}^{(\epsilon)}, \tag{12}$$

where $\mathcal{R}^{(\epsilon)} \triangleq \{\mathbf{y} \geq 0 : \mathbf{H}\mathbf{y} \leq \mathbf{b} - \epsilon\}$.

Since U is strictly concave and $\mathcal{R}^{(\epsilon)}$ is convex, $\overline{\mathbf{x}}^{(\epsilon)}$ is unique. In addition, as $\epsilon \to 0$, $\overline{\mathbf{x}}^{(\epsilon)}$ converges to the optimal rate vector $\mathbf{r}^*$. Without loss of generality, we assume $\|\overline{\mathbf{x}}^{(\epsilon)} - \mathbf{r}^*\| \leq \rho^{(\epsilon)}$, where $\lim_{\epsilon \downarrow 0} \rho^{(\epsilon)} = 0$. The relationship between $\mathcal{R}^{(\epsilon)}$ and $\mathcal{R}$ in two-dimensional case is shown in Figure 2.

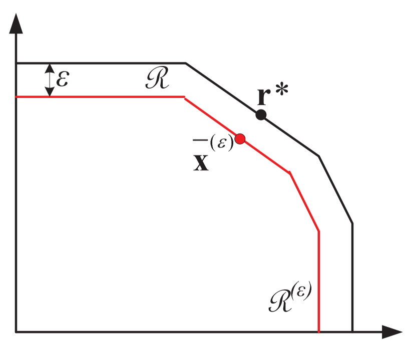


Fig. 2: The relationship between $\mathcal{R}^{(\epsilon)}$ and $\mathcal{R}$.

Each link maintains a data queue and a virtual queue. Let $Y_l^{(\epsilon)}[t]$ denote the virtual queue length at link l in each slot t.

Algorithm 1: (**Rate Deviation Optimal (RDO) Algorithm with parameter** ϵ): At each time slot t,

Flow control: $\{\mathbf{x}^{(\epsilon)}[t] = (x_l^{(\epsilon)}[t])_l\}_{t\geq 1}$ is a sequence generated by a numerical optimization algorithm solving ϵ-NUM. Note that $\mathbf{x}^{(\epsilon)}[t] \in \mathcal{R}^{(\epsilon)}, \forall t \geq 1$.

Arrival generation: For each link l,

(1) If $t = 1$, then $X_l^{(\epsilon)}[t] = B$;
(2) else if $\sum_{i=1}^{t-1} X_l^{(\epsilon)}[i] < \sum_{i=1}^{t-1} x_l^{(\epsilon)}[i]$, then $X_l^{(\epsilon)}[t] = B$; $X_l^{(\epsilon)}[t] = 0$, otherwise.

Then, inject $X_l^{(\epsilon)}[t]$ packets into each data queue l and increase virtual queue length $Y_l^{(\epsilon)}[t]$ by $x_l^{(\epsilon)}[t]$.

Scheduling: Perform Maximum Weight Scheduling (MWS) algorithm among virtual queues, that is,

$$\mathbf{S}^{(\epsilon)}[t] \in \underset{\boldsymbol{\eta}=(\eta_l)_{l=1}^L \in \mathcal{S}^{J[t]}}{\text{argmax}} \sum_{l=1}^{L} Y_l^{(\epsilon)}[t]\eta_l, \tag{13}$$

where $J[t] \in \mathcal{J}$ denotes the channel state at time t. Use $\mathbf{S}^{(\epsilon)}[t]$ to serve data queues.

Queue evolution: Let $Q_l^{(\epsilon)}[1] = Y_l^{(\epsilon)}[1], \forall l$, and for $t \geq 2$, update the data queue length and virtual queue length as follows:

$$Q_l^{(\epsilon)}[t+1] = \left(Q_l^{(\epsilon)}[t] + X_l^{(\epsilon)}[t] - S_l^{(\epsilon)}[t]\right)^+, \forall l, \tag{14}$$

$$Y_l^{(\epsilon)}[t+1] = \left(Y_l^{(\epsilon)}[t] + x_l^{(\epsilon)}[t] - S_l^{(\epsilon)}[t]\right)^+, \forall l. \tag{15}$$

Remarks: (1) Recent advances in the design of distributed Newton's method (e.g, [12], [13]) show the promise in generating sequence $\{\mathbf{x}^{(\epsilon)}[t]\}_{t\geq 1}$ in quick and distributed way. In addition, we can also use Gradient Projection method to solve ϵ-NUM.

(2) The purpose of maintaining the virtual queue is to help show the stability of data queues. In fact, directly performing MWS among data queues does not hurt the convergence speed, as we will see in the simulations. However, the Lyapunov drift argument to show the stability of the proposed algorithm does not work in such a case, since the deterministic arrivals are alternating between 0 and B, which leads to the potential positiveness of the one-step Lyapunov drift given the current queue length state.

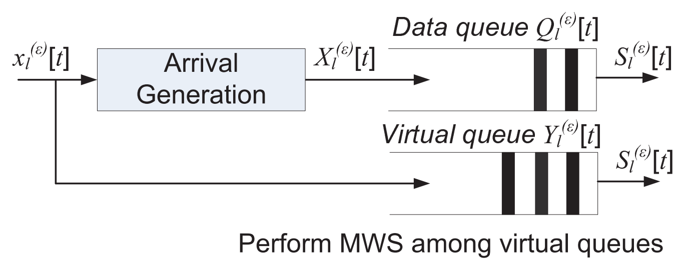


Fig. 3: The operation of the RDO Algorithm at link l.

Figure 3 shows the operation of the RDO Algorithm at link l. Next, we show that the RDO Algorithm can achieve rate deviation optimality if the generated sequence $\{\mathbf{x}^{(\epsilon)}[t]\}_{t\geq 1}$ converges fast enough. To that end, we need the following lemma exhibiting that the generated arrivals closely track the generated sequence $\{\mathbf{x}^{(\epsilon)}[t]\}_{t\geq 1}$.

Lemma 2: For each link l, we have

$$\left|\sum_{t=1}^{T}(X_l^{(\epsilon)}[t] - x_l^{(\epsilon)}[t])\right| \leq B, \qquad \forall T \geq 1. \tag{16}$$

Proof: The proof follows by induction. Please see our technical report [14] for details. ∎

Based on Lemma 2, we can show that for each link, the data queue length is upper-bounded by the sum of some constant and the virtual queue length for all sample paths, which is useful in establishing the rate deviation optimality of the RDO Algorithm.

Lemma 3: For each link l, the data queue length is upper-bounded by the sum of the virtual queue length and $2B$ for all sample paths, i.e.,

$$Q_l^{(\epsilon)}[T] \le Y_l^{(\epsilon)}[T] + 2B, \qquad \forall T \ge 1, \tag{17}$$

holds for all sample paths.

Proof: The proof follows from Lemma 2 and the Lindley's equation

$$Q_l^{(\epsilon)}[T] = \max_{1\le k\le T-1}\left\{\sum_{t=k}^{T-1} X_l^{(\epsilon)}[t] - \sum_{t=k}^{T-1} S_l^{(\epsilon)}[t], 0\right\}. \tag{18}$$

Please see our technical report [14] for details. ■

We are now ready to establish the rate deviation optimality of the RDO Algorithm.

Proposition 2: For the RDO Algorithm with parameter $\epsilon > 0$, as long as the flow controller satisfies $\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{x}^{(\epsilon)}[t] - \overline{\mathbf{x}}^{(\epsilon)}\| \le \frac{K_1}{T}$, for all $T \ge 1$, the generated link departure sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\ge1}$ satisfies

$$\mathbb{E}[\phi(\overline{\mathbf{d}}^{(\epsilon)}[T], \overline{\mathbf{x}}^{(\epsilon)})] \le \frac{K_1^{(\epsilon)}}{T}, \qquad \forall T \ge 1, \tag{19}$$

where K_1 and $K_1^{(\epsilon)}$ are some positive constants.

Remark: $\{\mathbf{x}^{(\epsilon)}[t]\}_{t\ge1}$ generated by the distributed Newton's method (e.g., [11], [12], [13]) satisfies the condition for Proposition 2.

Proof: We first give an upper bound on the expected rate deviation by using triangle inequality. Then, we show the boundedness of the virtual queue length by using Lyapunov Drift argument, which implies the boundedness of the data queue length by using Lemma 3. Finally, the result follows by using Lemma 2 and the condition that the flow rate vector converges fast enough. The detailed proof can be found in our technical report [14]. ■

Proposition 3: For the RDO Algorithm with parameter $\epsilon > 0$, if at least one coordinate of $\overline{\mathbf{x}}^{(\epsilon)}$ is a non-integer and the same condition in Proposition 2 holds, then the RDO Algorithm is rate deviation optimal (cf. Definition 3).

Proof: The result directly follows from Propositions 1, 2 and the definition of rate deviation optimality. ■

So far, we have observed that the discrete choice of transmission rates significantly limits the convergence speed to $\Omega(\frac{1}{T})$ and provided an algorithm that can achieve the optimal convergence speed in terms of rate deviation. In [15], the authors showed that for dual algorithm, the convergence speed of the running average of primal variables over T slots can be as fast as $\Omega\left(\frac{1}{T}\right)$ in terms of utility benefit. To the best of our knowledge, there does not exist a convergence speed analysis of dual methods in terms of rate deviation metric due to the non-smoothness of the dual function (see [16]). This motivates us to investigate the optimality of dual algorithm in terms of its convergence speed of the utility benefit metric under additional assumptions of non-randomness. These assumptions are necessary to establish the fundamental upper bound on the utility benefit under random environment, since the aggregation over links and the randomness (such as random arrivals, randomized scheduling or channel fading) distort the discrete structure. Thus, we focus on the deterministic system in next section, where there is no randomness in the system. It is still quite difficult and non-trivial to establish the convergence speed optimality in terms of utility benefit in such a system.

IV. Convergence speed in Utility benefit

In this section, we first establish the fundamental upper bound on the utility benefit for any algorithm in the deterministic system. Then, we show that both deterministic version of the RDO Algorithm and the well-known dual algorithm can achieve this upper bound and establish their optimality under utility benefit metric.

A. An upper bound on the utility benefit

In this subsection, we establish an upper bound on the utility benefit $U(\overline{\mathbf{d}}^{(\epsilon)}[T])$ for any algorithm in class $\mathcal{P}$ with parameter ϵ. We do not require the integer Assumption 1 for the deterministic system. Without out loss of generality, we assume that each link has a finite set of transmission rates $\mathcal{F} \triangleq \{a_1, a_2, ..., a_B\}$, where $0 \le a_1 < a_2 < ... < a_B$. To establish the fundamental upper bound on the utility benefit, we need the following assumption on the scheduling:

Assumption 2: Each link with queue length less than a_B is not to be scheduled. ◇

Remark: This scheduling assumption helps establish the fundamental bound on the utility benefit. Removing this assumption does not speedup the convergence, which is validated through simulations in our technical report [14].

We also need the following assumptions on the utility function:

Assumption 3: (1) The utility function $U(\mathbf{r})$ is additive, that is, $U(\mathbf{r}) = \sum_{l=1}^{L} U_l(r_l)$, where $U_l(y)$ is a concave and non-decreasing function of y;
(2) $h_{\min} \le U_l'(y) \le h_{\max}$, $\forall y$, where $0 < h_{\min} < h_{\max} < \infty$;
(3) $-\beta_{\max} \le U_l''(y) \le -\beta_{\min}$, $\forall y$, where $0 < \beta_{\min} < \beta_{\max}$. ◇

Examples of such utility functions include $U_l(y) = \log(y+\gamma)$ and $U_l(y) = \frac{(y+\gamma)^{1-m}}{1-m}$, where m and γ are positive constants. Now, we are ready to establish the fundamental upper bound on the utility benefit for any algorithm in class $\mathcal{P}$.

Proposition 4: (1) Under Assumption 2 and 3, for any $\delta \in (0, \max_{\mathbf{r}\in\mathcal{R}}\|\mathbf{r}-\mathbf{r}^*\|)$ and any policy in $\mathcal{P}$ with parameter ϵ, there exists a constant $c^{(\delta)} > 0$ and a positive integer-valued sequence $\{T_k\}_{k=1}^{\infty}$ such that

$$U(\overline{\mathbf{d}}^{(\epsilon)}[T_k]) \le U(\mathbf{r}^*) - \frac{1}{2}\beta_{\min}\sqrt{L}\delta^2 - \frac{c^{(\delta)}}{T_k}, \forall T_k \le \frac{c^{(\delta)}}{H\delta}, \tag{20}$$

where $H \triangleq \sqrt{L}(2a_B\beta_{\max} + h_{\max})$.

(2) If we further have $\sum_{l=1}^{L} U_l'(r_l^*)r_l^* \notin \mathcal{H}$, then there exists a sequence $\{c^{(\delta)}\}_{\delta>0}$ such that $c^{(0)} \triangleq \lim_{\delta\downarrow0} c^{(\delta)} > 0$ and $\lim_{\delta\downarrow0}\frac{c^{(\delta)}}{H\delta} = \infty$, and (20) becomes

$$U(\overline{\mathbf{d}}^{(\epsilon)}[T_k]) \le U(\mathbf{r}^*) - \frac{c^{(0)}}{T_k}, \qquad \forall k \ge 1, \tag{21}$$

where $\mathcal{H} \triangleq \left\{\sum_{l=1}^{L} U_l'(r_l^*)I_l : \mathbf{I} = (I_l)_{l=1}^{L} \in \mathcal{R} \text{ and } I_l \in \mathcal{F}, \forall l\right\}$.

Remark: The finiteness of the set $\mathcal{F}$ implies that the set $\mathcal{H}$ also has a finite number of elements. Thus, it is unlikely that $\sum_{l=1}^{L} U_l'(r_l^*)r_l^*$ is in the set $\mathcal{H}$ in practice.

Proof: The proof starts by finding a proper rate vector $\mathbf{r}^{(\delta)} \in \mathcal{R}$ depending on parameter δ. Then, we establish the relationship between the utility benefit and $U(\mathbf{r}^{(\delta)})$ by using the first-order optimization condition. Finally, by properly finding the intrinsic discrete structure, we have the desired result. Please see our technical report [14] for more details. ■

The first part of Proposition 4 establishes a fundamental bound on how close the utility benefit can be to the optimal utility level within a finite range of time. The second part, then, shows that, under an additional mild assumption on $\mathbf{r}^*$, the range over which the bound holds can be made to extend to infinity by letting the error go to zero. To illustrate the nature of this result, Figure 4 shows the utility benefit of an algorithm in class $\mathcal{P}$ over time. It shows that the utility benefit repeatedly falls below the fundamental bound until time $\frac{c^{(\delta)}}{H\delta}$, which, from the second part of the proposition, goes to infinity as δ vanishes.

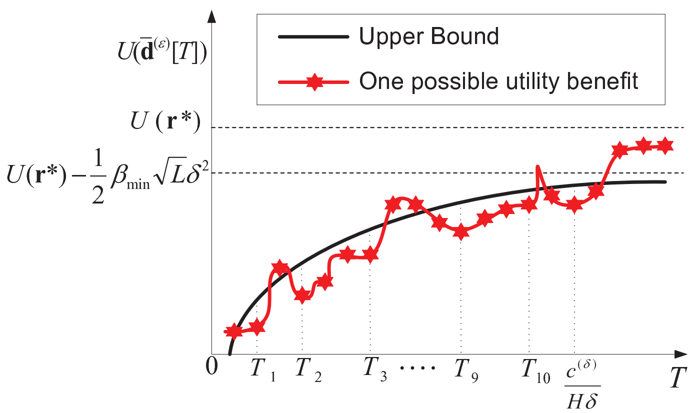


Fig. 4: The utility benefit of an algorithm in class $\mathcal{P}$.

Note that it is impossible for any policy in class $\mathcal{P}$ with parameter $\epsilon > 0$ that (20) holds for all $T \geq 1$, since the "good" policy (e.g., where ϵ is sufficiently small) can achieve the optimal value at arbitrary accuracy and thus $U(\overline{\mathbf{d}}^{(\epsilon)}[T])$ will exceed $U(\mathbf{r}^*) - \frac{1}{2}\beta_{\min}\sqrt{L}\delta^2$ eventually. In addition, inequality (21) implies that the utility benefit of any algorithm cannot be beyond the optimal value. Thus, these fundamental upper bounds on the utility benefit motivate the definition of utility benefit optimality of an algorithm given next.

Definition 5: (**Utility benefit optimality**) For any $\delta > 0$, an algorithm in class $\mathcal{P}$ with parameter $\epsilon > 0$ is called *utility benefit optimal*, if its generated departure rate vector sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\geq 1}$ satisfies

$$U(\overline{\mathbf{d}}^{(\epsilon)}[T]) \geq U(\mathbf{r}^*) - \frac{1}{2}\beta_{\min}\sqrt{L}\delta^2 - \frac{F_3^{(\epsilon)}}{T}, \forall T \leq F_4^{(\delta)}, \tag{22}$$

where $F_3^{(\epsilon)} > 0$ and $F_4^{(\delta)} > 0$ with $\lim_{\delta \to 0} F_4^{(\delta)} = \infty$. ◇

Next, we first investigate the utility benefit optimality of the deterministic version of the RDO Algorithm.

B. The utility benefit optimality of the RDO Algorithm

In this subsection, we show that the deterministic version of the RDO Algorithm is utility benefit optimal under Assumptions 2 and 3.

The deterministic version of the RDO Algorithm works as follows:

Algorithm 2: (**Deterministic RDO (DRDO) Algorithm with parameter** $\epsilon > 0$): At each time slot t,

Flow control: $\{\mathbf{x}^{(\epsilon)}[t] = (x_l^{(\epsilon)}[t])_l\}_{t\geq 1}$ is a sequence generated by a numerical optimization algorithm solving ϵ-NUM.

Arrival: Inject $x_l^{(\epsilon)}[t]$ amount of data into each queue l;

Scheduling: Perform the MWS algorithm among links whose queue length are no less than a_B, that is,

$$\mathbf{S}^{(\epsilon)}[t] \in \underset{\boldsymbol{\eta}=(\eta_l)_{l=1}^{L} \in \mathcal{R}}{\operatorname{argmax}} \sum_{l=1}^{L} Q_l^{(\epsilon)}[t] \mathbb{1}_{\{Q_l^{(\epsilon)}[t] \geq a_B\}} \eta_l; \tag{23}$$

Queue evolution: Update the queue length as follows:

$$Q_l^{(\epsilon)}[t+1] = (Q_l^{(\epsilon)}[t] + x_l^{(\epsilon)}[t] - S_l^{(\epsilon)}[t])^+, \forall l. \tag{24}$$

Next, we give a lower bound of the DRDO Algorithm under utility benefit metric.

Proposition 5: Under Assumption 3 on the utility function U, for the DRDO Algorithm with parameter $\epsilon > 0$, if $\frac{1}{T}\sum_{t=1}^{T} \|\mathbf{x}^{(\epsilon)}[t] - \overline{\mathbf{x}}^{(\epsilon)}\| \leq \frac{K_2}{T}$, for all $T \geq 1$, then its departure sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\geq 1}$ satisfies

$$U(\overline{\mathbf{d}}^{(\epsilon)}[T]) \geq U(\mathbf{r}^*) - h_{\max}\sqrt{L}\rho^{(\epsilon)} - \frac{\sqrt{L}h_{\max}K_2^{(\epsilon)}}{T}, \forall T \geq 1,$$

where K_2 and $K_2^{(\epsilon)}$ are some positive constants.

Proof: The proof first establishes the connection between the utility benefit of the DRDO Algorithm and its rate deviation by using the concavity of the utility function. Then, we establish the upper bound on the rate deviation by using similar technique in Proposition 2. The details can be found in our technical report [14]. ■

Proposition 6: Under Assumptions 2 and 3, the DRDO Algorithm is utility benefit optimal (c.f. Definition 5), i.e., for any $\delta > 0$, by choosing $\epsilon > 0$ such that $\rho^{(\epsilon)} \leq \frac{\beta_{\min}\delta^2}{2h_{\max}}$, the DRDO Algorithm can achieve the upper bound in (20).

Proof: The proof immediately follows from Propositions 4, 5 and the definition of utility benefit optimality. ■

Next, we study the utility benefit optimality of the well-known dual algorithm.

C. Utility benefit optimality of dual algorithm

In this subsection, we establish the utility benefit optimality of the well-known dual algorithm (e.g., [1], [2], [4], [5]). The dual algorithm can be obtained by Lagrangian relaxation and naturally decomposes the network function into the two main components: the congestion control and the scheduling. Next, we give the definition of the dual algorithm for completeness.

Definition 6: (**Dual Algorithm with parameter** $\epsilon > 0$):
Flow control: Given $\mathbf{Q}^{(\epsilon)}[t] = (Q_l^{(\epsilon)}[t])_{l=1}^L$, solve the following optimization problem:

$$\mathbf{x}^{(\epsilon)}[t] \in \underset{0 \le \boldsymbol{w} \le M}{\operatorname{argmax}} \ \frac{1}{\epsilon} U(\boldsymbol{w}) - \sum_{l=1}^{L} Q_l^{(\epsilon)}[t] w_l; \qquad (25)$$

Scheduling: Perform MWS algorithm among links whose queue length are no less than a_B, that is,

$$\mathbf{S}^{(\epsilon)}[t] \in \underset{\boldsymbol{\eta}=(\eta_l)_{l=1}^L \in \mathcal{R}}{\operatorname{argmax}} \sum_{l=1}^{L} Q_l^{(\epsilon)}[t] \mathbb{1}_{\{Q_l^{(\epsilon)}[t] \ge a_B\}} \eta_l; \qquad (26)$$

Queue evolution:

$$Q_l^{(\epsilon)}[t+1] = (Q_l^{(\epsilon)}[t] + x_l^{(\epsilon)}[t] - S_l^{(\epsilon)}[t])^+, \forall l, \qquad (27)$$

where M is the maximum allowable input rate. ◇

The Dual Algorithm also uses the scheduling assumption as the DRDO Algorithm that does not schedule links with queue length less than a_B, which helps establish its utility benefit optimality. However, removing this scheduling constraint does not improve the convergence speed, which is validated through simulations in our technical report [14].

We are now ready to give the convergence speed of the Dual Algorithm in terms of utility benefit.

Proposition 7: For the Dual Algorithm with parameter $\epsilon > 0$, the generated departure sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t \ge 1}$ satisfies

$$U(\overline{\mathbf{d}}^{(\epsilon)}[T]) \ge U(\mathbf{r}^*) - \frac{\epsilon}{2T} \|\mathbf{Q}^{(\epsilon)}[1]\|^2 - \frac{\epsilon L}{2}(M^2 + 3a_B^2) - \frac{h_{\max}\sqrt{L}}{T}\left(\|\mathbf{Q}^{(\epsilon)}[1]\| + G^{(\epsilon)}\sqrt{L}\right), \qquad \forall T \ge 1, \qquad (28)$$

where $G^{(\epsilon)} \triangleq \sqrt{W} + \frac{h_{\max}}{\epsilon}$ and $W \triangleq \left(\frac{\beta_{\max}}{\epsilon} + 2\right) LM^2 + \left(\frac{3\beta_{\max}}{\epsilon} + 2\right) La_B^2$.

Proof: The proof first shows the boundedness of queue length at all times for each link. Then, we establish the relationship between the utility benefit and the utility of the running average of flow rate vector sequence. By using similar technique in [15], we can give the lower bound on the utility of the running average of flow rate vector sequence. Note that the scheduling component in our setup makes our analysis more challenging than that in [15]. Please see our technical report [14] for details. ■

From (28), we can see that the utility benefit $U(\overline{\mathbf{d}}^{(\epsilon)}[T])$ converges to the optimal value $U(\mathbf{r}^*)$ within error level $\frac{\epsilon L}{2}(M^2 + 3a_B^2)$ with the speed of $\Omega\left(\frac{1}{T}\right)$. When the parameter ϵ decreases, the error level will decrease in the price of the slower convergence speed. Next, we establish the utility benefit optimality of the Dual Algorithm.

Proposition 8: The Dual Algorithm is utility benefit optimal (c.f. Definition 5), i.e., for any $\delta > 0$, by choosing $\epsilon \le \frac{\beta_{\min}\delta^2}{\sqrt{L}(M^2+3a_B^2)}$, the Dual Algorithm can achieve the upper bound in (20).

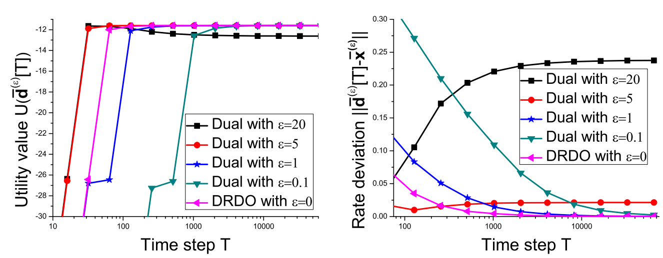
Proof: The proof directly follows from propositions 4, 7 and the definition of utility benefit optimality. ■

V. Simulation Results

In this section, we consider a single-hop network topology with $L = 5$ links in both non-fading and fading channels. In each time slot, at most one link can be active. We take the additive utility function with $U_l(y) = \log(y + \gamma), \forall l$, where $\gamma = 10^{-8}$, for both non-fading and fading scenarios. Recall that this function satisfies Assumption 3 on the utility function to establish the utility benefit optimality of both DRDO Algorithm and the Dual Algorithm. For the non-fading scenario, each link has a fixed rate and the link rate vector is $\mathbf{p} = [0.8, 0.4, 0.6, 0.5, 0.3]$. For the fading scenario, each link suffers from ON-OFF channel fading independently and the link ON probability vector is also $\mathbf{p}$. For DRDO and RDO algorithms with parameter ϵ, we use Newton's method (see [8]) to generate sequence $\{\mathbf{x}^{(\epsilon)}[t]\}_{t \ge 1}$ that satisfies the condition for Proposition 2.

A. Non-fading scenario

In this subsection, we mainly investigate the impact of parameter ϵ on the performance of the Dual Algorithm and compare it with the DRDO Algorithm with $\epsilon = 0$. In our technical report [14], we also study the impact of ϵ on the performance of the DRDO Algorithm and observe that ϵ does not have significant influence on the convergence speed under both utility benefit and rate deviation metrics, but has significant impact on the closeness to the optimal value.



(a) Utility benefit convergence (b) Rate deviation convergence

Fig. 5: Dual Algorithm Performance with varying ϵ.

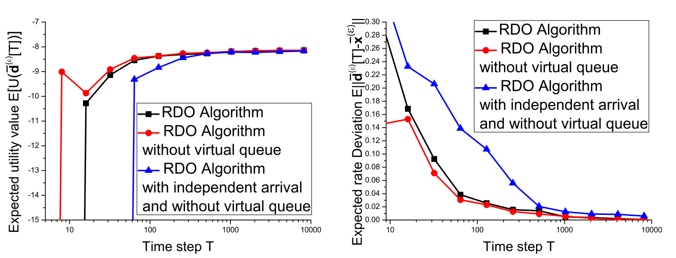
Figure 5a and 5b show the impact of parameter ϵ on the convergence speed of the Dual Algorithm. From Figure 5a, we can observe that when ϵ is too large (e.g., $\epsilon = 20$), the utility benefit cannot converge to the optimal value. From Figure 5b, we can see that the convergence under rate deviation metric requires much smaller ϵ than that under utility benefit metric. Here, it is worth mentioning that the large ϵ (e.g., $\epsilon = 1$) can still lead to the convergence to the optimal value under both metrics of interest. This is a little contradictory with the traditional dual algorithm in wireless networks, where the convergence property requires much smaller ϵ (e.g., $\epsilon = 0.01$). The reason is that we are interested in the time average metric rather than the instantaneous value. If ϵ is relatively large, the instantaneous value oscillates around the optimal value. However, the time average value may still be arbitrary close to the optimal value. In addition, among the set of parameters ϵ guaranteeing the convergence to the optimal value, the smaller ϵ leads to the slower convergence speed under both interest metrics. Thus, for the Dual Algorithm, we need to choose ϵ as large as possible among the set of parameters ϵ guaranteeing convergence to the optimal value.

In Figure 5a and 5b, we also compare the performance between the Dual Algorithm and the DRDO Algorithm with $\epsilon = 0$. We can observe that the Dual Algorithm with proper parameter ϵ (e.g., $\epsilon = 5$) converges slightly faster than the DRDO Algorithm under the utility benefit metric. This does not contradict our result in Section IV that both Dual and DRDO algorithms are utility benefit optimal and thus their convergence speed may differ at a constant factor. However, the DRDO Algorithm converges faster than the Dual Algorithm under rate deviation metric, which matches our theoretical result that the DRDO Algorithm is still optimal and the optimality of the Dual Algorithm is unknown under such metric.

B. Fading scenario

In this subsection, we mainly consider the performance of variants of the RDO Algorithm over wireless fading channels. The convergence speed comparison between the RDO Algorithm and a version of the Dual Algorithm (see [3]) under fading can be found in our technical report [14], where the observations are quite similar to that under non-fading scenario.

We consider a variant of the RDO Algorithm that does not require maintaining a virtual queue and performs MWS directly among data queues, and another variant of the RDO Algorithm that has independent random arrivals and performs MWS directly among data queues. Recall that the purpose of introducing virtual queues in the RDO Algorithm is to show the boundedness of the average data queue length by avoiding the difficulty in using Lyapunov Drift argument. From Figure 6a and 6b, we can observe that for both rate deviation and utility benefit metrics, the original RDO Algorithm converges faster than a variant of the RDO Algorithm with independent random arrivals, but slower than a variant of the RDO Algorithm without virtual queues. Thus, we suggest to use the arrival generation component in the RDO Algorithm and perform MWS among data queues directly in practice.



(a) Utility benefit convergence (b) Rate deviation convergence

Fig. 6: Performance of RDO Algorithm Variants.

VI. Conclusion

In this paper, we considered the convergence speed of joint flow control and scheduling algorithms in Network Utility Maximization (NUM) problem in multi-hop wireless networks. We realized that the discreteness of scheduling constraints and transmission rates are two of the most important features in wireless networks. We incorporated these important characteristics into the analysis and design of algorithms in terms of their convergence speed by defining two metrics of interest: rate deviation and utility benefit.

We showed that the convergence speed of any algorithm cannot be faster than $\Omega\left(\frac{1}{T}\right)$ for both rate deviation and utility benefit metrics due to the discrete choices of transmission rates at each link. This interesting and fundamental finding reveals that designing faster (e.g., Interior-Point or Newton based) algorithms for the flow rate control cannot break the barrier of $\Omega\left(\frac{1}{T}\right)$ in wireless networks caused by the scheduling component. Then, we provided an algorithm that can achieve optimal convergence speed under both rate deviation and utility benefit metrics. Moreover, we showed that the well-known dual algorithm also has optimal convergence speed in terms of utility benefit, which is a somewhat surprising outcome in view of the first-order nature of its iteration.

References

[1] F. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: Shadow prices, proportionalfairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[2] S. Low and D. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7(6):861–874, December 1999.

[3] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length based scheduling and congestion control. In *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, Miami, Florida, March 2005.

[4] A. Eryilmaz and R. Srikant. Joint congestion control, routing and mac for stability and fairness in wireless networks. *IEEE Journal on Selected Areas in Communications, special issue on Nonlinear Optimization of Communication Systems*, 14(8):1514–1524, August 2006.

[5] X. Lin and N. Shroff. The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks. In *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, Miami, Florida, March 2005.

[6] M. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. In *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, Miami, Florida, March 2005.

[7] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*. Kluwer, Boston, 2004.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, 2004.

[9] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1995.

[10] H. Feng, C. Xia, Z. Liu, and L. Zhang. Linear-speed interior-path algorithms for distributed control of information networks. *Performance Evaluation*, 67(11):1107–1122, 2010.

[11] A. Jadbabaie, A. Ozdaglar, and M. Zargham. A distributed newton method for network optimization. In *Proc. IEEE Conference on Decision and Control (CDC)*, Shanghai, China, December 2009.

[12] E. Wei, A. Ozdaglar, and A. Jadbabaie. Distributed newton method for network utility maximization. In *Proc. IEEE Conference on Decision and Control (CDC)*, Atlanta, Georgia, December 2010.

[13] J. Liu and H. Sherali. A distributed newton's method for joint multi-hop routing and flow control: Theory and algorithm. In *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, Florida, March 2012.

[14] B. Li, A. Eryilmaz, and R. Li. Wireless scheduling for network utility maximization with optimal convergence speed. Technical Report, 2013. Available online at http://www.ece.osu.edu/~eryilmaz/ConvergenceSpeed_report.pdf.

[15] A. Nedic and A. Ozdaglar. Subgradient methods in network resource allocation: Rate analysis. In *42nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, New Jersey, March 2008.

[16] D. Bertsekas, A. Nedic, and A. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, Mass., 2003.

APPENDIX A
PROOF OF LEMMA 1

Note that $\left|\frac{1}{T}\sum_{t=1}^{T} I[t]-b\right| \le \frac{c_b}{T}$ is equivalent to

$$-c_b \le \sum_{t=1}^{T} I[t]-Tb \le c_b. \quad (29)$$

(i) If $I[t+1]\le a_i$, then we have

$$\begin{aligned}\sum_{t=1}^{T+1} I[t]-(T+1)b &\le \sum_{t=1}^{T} I[t]-Tb+a_i-b\\ &\le c_b+a_i-b \quad (30)\\ &< -c_b, \quad (31)\end{aligned}$$

where (30) follows from (29), and (31) follows from $c_b < \frac{b-a_i}{2}$.

(ii) If $I_l[t+1]\ge a_{i+1}$, then we have

$$\begin{aligned}\sum_{t=1}^{T+1} I[t]-(T+1)b &\ge \sum_{t=1}^{T} I[t]-Tb+a_{i+1}-b\\ &\ge -c_b+a_{i+1}-b \quad (32)\\ &> c_b, \quad (33)\end{aligned}$$

where (32) follows from (29), and (33) follows from $c_b < \frac{a_{i+1}-b}{2}$.

Thus, by combining (31) and (33), we have the desired result.

APPENDIX B
PROOF OF PROPOSITION 1

We first show the following claim:

Claim 1: If $\overline{x}_l^{(\epsilon)} \in (i,i+1)$, for some $i=0,...,B-1$, then, there exists a $c_{\overline{\mathbf{x}}^{(\epsilon)}} \in (0,\min\{\frac{\overline{x}_l^{(\epsilon)}-i}{2},\frac{i+1-\overline{x}_l^{(\epsilon)}}{2}\})$ and a positive integer-valued sequence $\{T_k\}_{k\ge1}$ such that

$$\|\overline{\mathbf{d}}^{(\epsilon)}[T_k]-\overline{\mathbf{x}}^{(\epsilon)}\| \ge \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{T_k}, \forall k, \quad (34)$$

holds for any sample path of departure rate vector sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\ge1}$. ◇

Since all elements in the closure of $\overline{\mathcal{X}}^{(\epsilon)}$ have non-integer-valued coordinate, by Claim 1, if $\overline{x}_l^{(\epsilon)} \in (i_{\overline{\mathbf{x}}^{(\epsilon)}}, i_{\overline{\mathbf{x}}^{(\epsilon)}}+1)$ for some non-negative integer $i_{\overline{\mathbf{x}}^{(\epsilon)}}$, then we can take $c'_{\overline{\mathbf{x}}^{(\epsilon)}} = \frac{1}{4}\min\{\overline{x}_l^{(\epsilon)}-i_{\overline{\mathbf{x}}^{(\epsilon)}}, i_{\overline{\mathbf{x}}^{(\epsilon)}}+1-\overline{x}_l^{(\epsilon)}\}, \forall \overline{\mathbf{x}}^{(\epsilon)} \in \overline{\mathcal{X}}^{(\epsilon)}$, and there exists a positive integer-valued sequence $\{T_k\}_{k\ge1}$ such that

$$\|\overline{\mathbf{d}}^{(\epsilon)}[T_k]-\overline{\mathbf{x}}^{(\epsilon)}\| \ge \frac{c'_{\overline{\mathbf{x}}^{(\epsilon)}}}{T_k}, \forall k, \quad (35)$$

holds for any sample path of departure rate vector sequence.

If $c \triangleq \inf_{\overline{\mathbf{x}}^{(\epsilon)}\in\overline{\mathcal{X}}^{(\epsilon)}} c'_{\overline{\mathbf{x}}^{(\epsilon)}} = 0$, then, one of the limiting points of the set $\overline{\mathcal{X}}^{(\epsilon)}$ has all integer-valued coordinates, which contradicts our assumption that the closure of set $\overline{\mathcal{X}}^{(\epsilon)}$ does not contain a vector with all integer-valued coordinates. Thus, we have $c>0$ and obtain

$$\phi(\overline{\mathbf{d}}^{(\epsilon)}[T_k],\overline{\mathcal{X}}^{(\epsilon)}) = \inf_{\overline{\mathbf{x}}^{(\epsilon)}\in\overline{\mathcal{X}}^{(\epsilon)}} \|\overline{\mathbf{d}}^{(\epsilon)}[T_k]-\overline{\mathbf{x}}^{(\epsilon)}\| \ge \frac{c}{T_k}, \forall k, \quad (36)$$

holds for any sample path of departure rate vector sequence $\{\mathbf{D}^{(\epsilon)}[t]\}_{t\ge1}$, which implies (9).

Next, we prove Claim 1 to complete the proof. Since

$$\|\overline{\mathbf{d}}^{(\epsilon)}[T_k]-\overline{\mathbf{x}}^{(\epsilon)}\| \ge |\overline{d}_l^{(\epsilon)}[T_k]-\overline{x}_l^{(\epsilon)}|, \forall k, \quad (37)$$

we only need to show

$$|\overline{d}_l^{(\epsilon)}[T_k]-\overline{x}_l^{(\epsilon)}| \ge \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{T_k}, \forall k. \quad (38)$$

Indeed, since $c_{\overline{\mathbf{x}}^{(\epsilon)}} < \frac{1}{2}$ and $\overline{x}_l^{(\epsilon)} > 0$, we have

$$\frac{m+1}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}} - \left(\frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}\right) = \frac{1-2c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}} > 0, \quad (39)$$

which implies

$$(\frac{m}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}) \cap (\frac{m+1}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m+1}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}) = \emptyset,$$

for any non-negative integer m. Since

$$\frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}} - \left(\frac{m}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}\right) = \frac{2c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}} < 1, \quad (40)$$

each interval $(\frac{m}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}})$ can at most contain one non-negative integer.

If the interval $(\frac{m}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}})$ contains some non-negative integer T for some non-negative integer m, i.e., $|\frac{m}{T} - \overline{x}_l^{(\epsilon)}| \le \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{T}$, where $\overline{x}_l^{(\epsilon)} \in (i,i+1)$ for some $i=0,...,B-1$ and $c_{\overline{\mathbf{x}}^{(\epsilon)}} \in (0,\min\{\frac{\overline{x}_l^{(\epsilon)}-i}{2},\frac{i+1-\overline{x}_l^{(\epsilon)}}{2}\})$, then, by taking the set $\mathcal{I}=\mathbb{N}^0 \triangleq \{0,1,2,...\}$ in Lemma 1, we have $|\frac{m+l}{T+1} - \overline{x}_l^{(\epsilon)}| \ge \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{T+1}$ for any positive integer l, which implies

$$T+1 \notin (\frac{m+l}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m+l}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}) \text{ for any positive integer } l.$$

Thus, $\cup_{m=0}^{\infty}(\frac{m}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{m}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}})$ does not cover all positive integers and thus there exists a sequence of positive integers $\{T_k\}_{k=1}^{\infty}$ such that

$$T_k \notin (\frac{j}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{j}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}) \text{ for any non-negative integer } j.$$

For any sample path of departure rate vector sequence $\{D_l^{(\epsilon)}[t], t\ge1\}$, $\sum_{t=1}^{T_k} D_l^{(\epsilon)}[t]$ is an integer and thus we have

$$T_k \notin (\frac{\sum_{t=1}^{T_k} D_l^{(\epsilon)}[t]}{\overline{x}_l^{(\epsilon)}} - \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}, \frac{\sum_{t=1}^{T_k} D_l^{(\epsilon)}[t]}{\overline{x}_l^{(\epsilon)}} + \frac{c_{\overline{\mathbf{x}}^{(\epsilon)}}}{\overline{x}_l^{(\epsilon)}}), \quad (41)$$

which is equivalent to (38).