Queue-Affectance-based Scheduling in Multi-hop Wireless Networks under SINR Interference Constraints

Changhee Joo
ECE, UNIST
Ulsan, Korea
Email: cjoo@unist.ac.kr

Myeongseon Shin
LG Electronics
Seoul, Korea
Email: myungseon.shin@lge.com

***Abstract*—Most distributed wireless scheduling schemes that are provably efficient have been developed under the protocol model, which describes interference constraints in a binary form. However, the oversimplified interference model imposes fundamental limitations on the performance in practice. The signal-to-interference-plus-noise-ratio (SINR) based interference model is more accurate and realistic accounting for the cumulative nature of the interference signals, but its complex structure makes the design of scheduling schemes much more challenging. In this paper, we focus on the scheduling performance under the SINR model and develop random access scheduling schemes that are amenable to implement in a distributed fashion with only local information. We analytically show that they are provably efficient under the SINR model, and through simulations demonstrate that they empirically perform better than the theoretical performance bound.**

I. Introduction

In multi-hop wireless networks, link scheduling is an essential element in effectively using the limited bandwidth resources by increasing the frequency spectrum efficiency and avoiding collisions from simultaneous transmissions. Due to the shared nature of wireless media, successful transmission is determined by the ratio of the received signal strength to the interference level from simultaneous transmissions. This interference relationship has been modeled using the *protocol model* or the *physical model* [1].

In the protocol model, a node can successfully transmit or receive packets if there are no simultaneous transmissions within a certain distance or hops. This model has been widely adopted since the interference relationship of two links can be easily represented by a conflict graph and their binary relation is independent of the others. The symmetry and independence attributes of the protocol model greatly facilitate the development of efficient scheduling schemes in multi-hop wireless networks. However, while the protocol model captures the essential feature of wireless interference, it also substantially distorts the interference constraints by ignoring the important property of *interference accumulation from all concurrent transmissions in the network*. For example, in a three-link network, where any two links can transmit at the same time but all three links cannot be active simultaneously due to accumulated interference, the protocol model cannot be used to describe the interference constraints.

In contrast, the physical model can successfully take into account the cumulative nature of interference by considering the *Signal-to-Interference-plus-Noise-Ratio* (SINR), i.e., a transmission is successful if the SINR at the receiver exceeds a certain threshold. The authors in [2] show through experimental evaluation that the physical model (or the *SINR model*) accurately predicts the impact of interference.

Given an interference model, let *feasible schedule* denote as a set of links that conform to the interference constraints. The *capacity region* of the network, defined as the set of achievable arrival rates, can be described using the set of feasible schedules [3]. It has been known that Maximum Weight Scheduling (MWS) that maximizes the queue weighted sum at each time can achieve the capacity region without a priori knowledge of arrival rate. However, the throughput-optimal scheme is hard to implement in practice, since it requires centralized information and has to solve an NP-Hard problem even under the protocol model [4].

For practical uses, low-complexity distributed algorithms that can schedule links with provable efficiency have attracted much attention in the literature. They can be classified into a few groups: S_γ*-type* scheduling schemes that achieve a queue weighted sum no smaller than the γ fraction of the maximum [5] (e.g., Greedy Maximal Scheduling (GMS)), *Pick-and-Compare* scheduling schemes that gradually improve the scheduling performance by a random pick [6], [7], *Constant-time Random Access* scheduling schemes that approximate GMS by giving an access priority to the link with relatively large weight [8], [9], and recently developed *throughput-optimal CSMA* scheduling schemes that exploit the carrier-sensing functionality and the Glauber dynamics [10], [11].

However, the above low-complexity scheduling schemes have been developed under the protocol model, and their extensions to the SINR model are not straightforward. Most low-complexity scheduling schemes assume that the *feasibility check* of a schedule can be done in a prompt and reliable manner, e.g., through the carrier-sensing/signal-overhearing functionality or through explicit control message exchanges between neighboring links. Under the SINR model, these approaches are not scalable and hard to implement in a fully distributed manner. Note that a link's transmission may fail by adding a small amount of additional interference from a far-away link. This implies that the feasibility check of a schedule under the SINR model would require global information of all links' activities [12] or global feedback through the network [13]. However, collecting global information incurs a significant amount of control overhead which substantially degrades the scheduling performance in practice. Hence, the complex interference structure of the SINR model makes the design of low-complexity distributed scheduling schemes much more challenging and difficult.

There have been studies that address wireless scheduling problems under the SINR model from different angles. The complex interference relations among the links are extended to a precomputed conflict graph of virtual nodes with binary interference constraints, and the conflict-graph-based previous scheduling schemes can operate over the graph [14], [15]. In [12], the authors localize the interference on each link, and develop a distributed greedy scheduling scheme that can provably guarantee a non-vanishing fraction of the capacity region, which, however, can be arbitrarily small. Instead of maximizing the throughput, the minimum-length-scheduling problem that aims to minimize the time to serve a given initial demand vector under the SINR model has been studied. It has been shown that the problem is solvable in polynomial time in the special case that the demand vector is superincreasing [16], and a novel concept of *affectance* (or relative interference) has been introduced to solve the problem in a greedy manner [17]. From a snapshot point of view, there are efforts to maximize the number of simultaneously scheduled links, i.e., maximum independent set problem. In [18], the authors use a similar concept of affectance and solve the maximum independent set problem using a greedy approach. In [19], a power allocation problem has been addressed to maximize the number of scheduled links. In [20], a linear program has been used to achieve a constant-factor approximation. However, these works under the SINR model often take a greedy approach that requires an ordering of the links, and assume that the feasibility check for a set of links can be done, which commonly results in a significant amount of overhead of control message exchanges and medium sensing [21]. The work of [22] avoids such overhead by accessing the medium with a probability based on the maximum average affectance, but it requires global information that can be obtained only through intensive computations over all the possible schedules. Recently, a distributed power allocation scheme has been developed [23], which can fully explore the rate region without any message passing. However, it requires a priori knowledge of the arrival rate and may suffer from poor delay performance due to the large mixing time of Markov Chain under the CSMA algorithm [24].

In this paper, we focus on the scheduling problem of the capacity maximization under the SINR model, and develop a low-complexity distributed scheduling algorithm that does not require the feasibility check of a schedule and yet guarantees a fraction of the capacity region without priori knowledge of the arrival rate. Our contributions can be summarized as follows.

i) We make use of the affectance as the normalized measure of interference, and investigate the relationship between the affectance sum and the capacity region.
ii) We propose queue-affectance-based random access scheduling schemes that are amenable to implementation in a distributed fashion and operate without the feasibility check, while they guarantee a fraction of the capacity region.
iii) We investigate an upper bound on the affectance sum, which provides a lower bound on the performance of the proposed scheduling schemes.
iv) We evaluate the performance of our proposed schemes through simulations and demonstrate that they outperform other distributed scheduling schemes.

The rest of this paper is organized as follows. System model is described in Section II. We propose our random access scheduling schemes in Section III. After the relationship of the affectance with the capacity region is investigated, the performance of our scheduling schemes through the affectance sum is analyzed in Section IV. An upper bound on the affectance sum is provided in Section V. Finally, we evaluate the performance of the proposed schemes through simulations in Section VI, and conclude our paper in Section VII.

II. System Model

We consider a wireless network $G = (V, E)$ with the set V of nodes and the set E of unidirectional links. Two nodes are connected by a link if one can directly transmit a packet to the other. Time is slotted, and a single frequency channel is shared by all the links. Due to wireless signal propagation, simultaneous transmissions of two links may interfere with each other. Link l consists of the sender node and the receiver node. Let $\hat{d}_{ll}$ denote the distance between the receiver of link l and the sender of link l, and let $\hat{d}_{lk}$ denote the distance between the receiver of link l and the sender of link k. Let $d_{ll} := \max\{1, \hat{d}_{ll}\}$ and $d_{lk} := \max\{1, \hat{d}_{lk}\}$, where a unit distance is defined as the distance from which the transmitted signal power starts attenuating. Further we assume that $d_{ll} \leq d_{max}$ for some d_{max}.

The sender of link l can successfully transmit a packet to the receiver if the received SINR is greater than a certain threshold β. That is, the transmission of link l is successful, if

$$\frac{P_l d_{ll}^{-\alpha} C_l(t)}{\sum_{k \in E, k \neq l} P_k d_{lk}^{-\alpha} C_k(t) + N} > \beta, \tag{1}$$

where P_l denotes the transmission power of link l, α denotes the path loss exponent, N denotes the noise power, and $C_l(t) \in \{0, 1\}$ denotes the activity of link l at time t. We use the vector of the link activities, denoted by $\vec{C}(t)$, to represent a schedule. A schedule $\vec{C}(t)$ is said to be *feasible*, if all active links l (i.e., $C_l(t) = 1$) are successful satisfying (1). Let $\mathcal{C}$ denote the set of all feasible schedules. In this work, we assume that the channel gain depends only on the distance between the sender and the receiver, and that all the links use the same transmission power, i.e., $P_l = P$ for all $l \in E$.

Let $R_l(t)$ denote the number of packets that arrive at link l at the beginning of time slot t. Let λ_l denote the mean arrival rate, i.e., $\lambda_l = E[R_l(t)]$, and let $\vec{\lambda}$ denote its vector, $\vec{\lambda} = [\lambda_1, ..., \lambda_{|E|}]$. We consider single-hop traffic in multi-hop networks, i.e., packets leave the system immediately after successful transmission. Let $D_l(t)$ denote the number of packets that link l can serve in time slot t. For ease of exposition, we assume that all the links have unit link capacity, i.e., a link can transmit one packet per time slot, if the transmission is successful. However, the result can extend to any constant link rates. Let $Q_l(t)$ denote the number of packets in the queue of link l at the beginning of time slot t. The queue length will evolve as

$$Q_l(t+1) = [Q_l(t) - D_l(t) + R_l(t)]^+, \tag{2}$$

where $[\cdot]^+$ denotes $\max\{0, \cdot\}$. A network is said to be *stable* if the queue lengths of all links remain finite.

The *capacity region* Λ is the set of rate vectors $\vec{\lambda}$ for which there is a scheduling policy that keeps the network stable. A *throughput-optimal* scheduling policy is defined as the scheduling policy that achieves the capacity region. We define the *efficiency ratio* of a scheduling policy as the largest number $0 \leq \gamma \leq 1$ such that, the policy can support $\gamma\vec{\lambda}$ for all $\vec{\lambda} \in \Lambda$. Throughout the paper, we use the efficiency ratio as the performance metric of a scheduling policy.

Note that if two links are very close to each other, they cannot simultaneously transmit at an acceptable rate due to mutual interference, and if two links are distant far away, their transmissions are unlikely to interfere with each other. To this end, we consider two ranges $\underline{R}_l$ and $\overline{R}_l$ to denote an exclusive transmission area and a substantial interference area, respectively. Let $\underline{R}_l$ denote the distance d_{lk} such that $\frac{P_l d_{ll}^{-\alpha}}{P_k d_{lk}^{-\alpha} + N} = \beta$. Let $\overline{R}_l$ denote the distance outside which a link's transmission can have insignificant interference to link l, i.e., $P_k d_{lk}^{-\alpha} \ll N$, if $d_{lk} > \overline{R}_l$. Link l will consider the interference from any link within $\overline{R}_l$ distance. We will characterize $\overline{R}_l$ later.

Since we can treat the links outside $\overline{R}_l$ as non-interfering to link l, we confine the neighborhood of link l to links in $\overline{R}_l$. Let N_l be the set of links whose distance from link l is less than $\overline{R}_l$. N_l does not include link l. Note that the neighborhood relation can be asymmetric, i.e., link l can be considered as a neighbor of link k, while link k does not consider link l as a neighbor. Let N_l' denote the set of links whose neighbor set contains l, i.e., $k \in N_l'$ if $l \in N_k$. Further, let $N_l^+ := N_l \cup N_l' \cup \{l\}$.

To measure the interference experienced at a link from its neighborhood, we will use the notion of *affectance* (or relative interference) as in [17], [22].

Definition 1: The *affectance* a_{lk} from link k on link l is defined as

$$a_{lk} := \min\left\{1, \beta \frac{P d_{lk}^{-\alpha}}{P d_{ll}^{-\alpha} - \beta N}\right\}. \tag{3}$$

By combining (1) and (3), in a feasible schedule, the sum of affectance on any active link l from its neighborhood is bounded by 1, i.e., $\sum_{k \in N_l} a_{lk} C_k < 1$ for any active link l in feasible $\vec{C}$. Let A_l denote the maximum affectance sum of link l from its active neighbors (over all the feasible schedules), i.e.,

$$A_l := \max_{\vec{C} \in \mathcal{C}} \left\{1, \sum_{k \in N_l} a_{lk} C_k\right\}. \tag{4}$$

We make use of the affectance in the scheduling decision to take into account the strength of the interference.

We assume that the control message can be transmitted to the neighborhood without loss, which can be justified by the fact that the control message is much smaller than a data packet and often transmitted at the lowest transmission rate. We can also consider a separate channel for the control message using a lower frequency band link Super Wi-Fi. In the sequel, we omit subscript t if there is no confusion.

III. Random Access Scheduling Algorithm

We consider a random access scheme like slotted ALOHA. At the beginning each time slot, each wireless link l transmits its data with probability p_l. Provided that link l attempts at time t, the transmission is successful if the SINR constraint (1) is satisfied.

We design a random access scheduling scheme, named *Queue-length and Affectance based Random Access Scheduling* (QARAS), that sets $p_l = \frac{x_l}{2}$, where

$$x_l := \frac{Q_l}{\max_{k \in N_l^+} \left(\sum_{j \in N_k} a_{kj} Q_j + Q_k\right)}. \tag{5}$$

Through this, we assign higher attempt probability to the link with higher (relative) queue length in its neighborhood, since the high queue length implies less service for the link. Different interference levels of its neighbors are taken into consideration through the affectance-weighted queue length sum in the denominator: For a fully-interfering neighbor, i.e., link k that cannot be ever active with link j, its full queue length is considered in the denominator since $a_{kj} = 1$, and for a partially-interfering neighbor, its queue length is summed with weight $a_{kj} < 1$. Hence, the neighbor with less interference level makes less contribution to the denominator and less decreases the attempt probability.

We highlight that the set N_l^+ and the affectance a_{kj} are fixed and can be known in advance. Each link can calculate (5) by timely exchanging the queue length information with its neighboring links or by piggybacking it on the data packet as in [9], [25]. Practicability of the latter approach has been already tested through experiments in [24].

Remarks: Eq. (5) looks similar to the attempt probabilities of constant-time scheduling policies [9], [25], [26] that have been developed under the binary symmetric interference model. However, there are significant differences. First, it takes into account the cumulative nature of interference signals by using the affectance. Second, there is no contention period, which is because validating the feasibility of a schedule in a distributed manner (i.e., ensuring that all scheduled links satisfy the SINR constraint) will be much more complicated than the case with the binary symmetric interference model. We also note that the LQF scheme in [12] has a linear complexity with respect to the number of links to determine a schedule at each time slot. In contrast, our random access scheme has no such overhead.

In the following, we investigate the relationship of the affectance and the capacity region, and analytically evaluate the performance of QARAS.

IV. Performance Analysis

The affectance can serve as a normalized indicator of the interference from one link to another. In this section, we show how one can relate the affectance to the scheduling performance and analyze the performance of QARAS.

A. Affectance and the Capacity Region

The maximum affectance sum A_l describes the (normalized) maximum interference that link l will receive from its neighboring links In an extreme case, if the interference relationship of any two links is binary, such as the protocol model, then the affectance sum is equivalent to the *interference degree* that is defined as the largest number of active links in the set of a link's interfering neighbors. Note that random access scheduling schemes can achieve a factor of the interference degree of the capacity region in the protocol model [8], [9].

We will show that the maximum affectance sum is the key factor in characterizing the scheduling performance of QARAS under the SINR interference model. To elaborate, we will present an upper bound of the capacity region under the SINR model in terms of the maximum affectance sum and define a norm-like measure, which will be used to demonstrate the efficiency ratio of QARAS.

Recall that Λ denotes the capacity region.

Lemma 1: Any feasible arrival rate $\vec{\lambda} \in \Lambda$ (strictly inside) satisfies that

$$\sum_{k \in N_l} a_{lk}\lambda_k + \lambda_l \leq A_l + 1, \text{ for all } l. \tag{6}$$

Proof: Under a feasible schedule $\vec{C}$, if link l is active, we have $\sum_{k \in N_l} a_{lk} C_k + C_l \leq 2$, because the maximum affectance sum from all active links in N_l on link l should be less than 1 to satisfy (1). Similarly, if link l is inactive, we have $\sum_{k \in N_l} a_{lk} C_k \leq A_l$ from the definition of A_l.

There exists an optimal stationary scheduling policy that schedules links independent of the system state and yet stabilizes the network [27]. Consider a sequence of schedules $\vec{C}^*(t)$ computed by such an optimal scheduling policy, satisfying $\lambda_l < E[C_l^*(t)]$ for all l. Let λ_{N_l} denote the time fraction that one of link l's neighbors transmits data under this optimal scheduling. Since $\vec{\lambda} \in \Lambda$ strictly inside, there is $\delta > 0$ and λ_{N_l} such that link l is scheduled for $\lambda_l + \frac{\delta}{2}$, at least one of link l's neighbors is scheduled for $\lambda_{N_l} + \frac{\delta}{2}$, and $\lambda_l + \lambda_{N_l} + \delta = 1$. We have that

$$\begin{aligned}\sum_{k \in N_l} a_{lk}\lambda_k + \lambda_l &\leq E\left[\sum_{k \in N_l} a_{lk} C_k^*(t) + C_l^*(t)\right] \\ &\leq 2 \cdot P\{l \text{ is active}\} + A_l \cdot P\{l \text{ is inactive}\} \\ &\leq 2 \cdot (\lambda_l + \frac{\delta}{2}) + A_l \cdot (\lambda_{N_l} + \frac{\delta}{2}) \\ &\leq A_l + 1.\end{aligned} \tag{7}$$

where the last inequality comes from the facts that $\lambda_l + \lambda_{N_l} + \delta = 1$, $A_l \geq 1$, and $\lambda_l + \frac{\delta}{2} \leq 1$. ■

Since any feasible arrival $\vec{\lambda} \in \Lambda$ must satisfy (6) for all l, we have $\Lambda \subseteq \Omega$, where

$$\Omega := \left\{ \vec{\lambda} \mid \sum_{k \in N_l} \frac{a_{lk}}{(A_l + 1)}\lambda_k + \frac{\lambda_l}{(A_l + 1)} \leq 1, \text{ for all } l \right\}. \tag{8}$$

We now define the *length* of vector $\vec{x}$ with respect to Ω, following the line of the analysis of [25], as

$$\|\vec{x}\|_\Omega := \left(\sup\{\gamma \mid \gamma \cdot \vec{x} \in \Omega\}\right)^{-1}. \tag{9}$$

We also define the *normalized* vector $\tilde{\mathbf{x}}$ of $\vec{x}$ with respect to Ω as

$$\tilde{\mathbf{x}} = \frac{\vec{x}}{\|\vec{x}\|_\Omega}, \tag{10}$$

i.e., $\tilde{\mathbf{x}}$ is the longest vector in Ω that is in the same direction as $\vec{x}$. The following lemma shows that the length has many properties of a norm [25].

Lemma 2:

1) $\|\vec{x}\|_\Omega \geq 0$. Further, $\|\vec{x}\|_\Omega > 0$ if $\vec{x} \neq 0$.
2) $\|\tilde{\mathbf{x}}\|_\Omega = 1$.
3) If $\vec{y} = \alpha\vec{x}$, where $\alpha > 0$, then $\|\vec{y}\|_\Omega = \alpha\|\vec{x}\|_\Omega$.
4) If $y_l \geq x_l$ for all l, then $\|\vec{y}\|_\Omega \geq \|\vec{x}\|_\Omega$.
5) $\|\vec{x} + \vec{y}\|_\Omega \leq \|\vec{x}\|_\Omega + \|\vec{y}\|_\Omega$.

We omit the proof and refer interested readers to [25].

Let $\tilde{\mathbf{Q}}(t)$ denote the normalized vector of the queue length $\vec{Q}(t)$. From (10) and Part 2 of Lemma 2, it is clear that $\|\tilde{\mathbf{Q}}(t)\|_\Omega = 1$ if $\vec{Q}(t) \neq 0$. Then, since $\tilde{\mathbf{Q}}(t)$ is on the boundary of Ω, it should satisfy that, for all l,

$$\sum_{k \in N_l} \frac{a_{lk} Q_k(t)}{(A_l + 1)\,\|\vec{Q}(t)\|_\Omega} + \frac{Q_l(t)}{(A_l + 1)\,\|\vec{Q}(t)\|_\Omega} \leq 1.$$

This implies that

$$\begin{aligned}\frac{1}{\|\vec{Q}(t)\|_\Omega} &\leq \min_{l \in E} \frac{1}{\sum_{k \in N_l} \frac{a_{lk} Q_k(t)}{(A_l+1)} + \frac{Q_l(t)}{(A_l+1)}} \\ &\leq \frac{1}{\max_{l \in E}\left(\sum_{k \in N_l} \frac{a_{lk} Q_k(t)}{(A_l+1)} + \frac{Q_l(t)}{(A_l+1)}\right)}.\end{aligned} \tag{11}$$

B. Scheduling Performance of QARAS

Under the algorithm described in Section III, we calculate the probability of successful transmission of each link, and obtain the performance bound using the Lyapunov technique.

Let P_l^s denote the probability of successful transmission of link l.

Lemma 3: Setting the transmission probability $p_l = \frac{x_l}{2}$, QARAS achieves

$$P_l^S \geq \frac{x_l}{4} \text{ for all } l. \tag{12}$$

Proof: The successful transmission probability P_l^S is the product of link l's transmissions probability and the probability that (1) holds at link l, i.e.,

$$\begin{aligned} P_l^S &= p_l \cdot \Big(1 - Prob\Big\{ \sum_{k \in N_l} a_{lk} C_k \geq 1 \Big\}\Big) \\ &\geq p_l \cdot \Big(1 - E\Big[\sum_{k \in N_l} a_{lk} C_k \Big]\Big) \\ &= p_l \cdot \Big(1 - \sum_{k \in N_l} a_{lk} p_k \Big), \end{aligned} \tag{13}$$

where the inequality comes from the Markov inequality. Note that for any $k \in N_l$, we have

$$\max_{j \in N_k^+} \Big(\sum_{i \in N_j} a_{ji} Q_i + Q_j \Big) \geq \Big(\sum_{i \in N_l} a_{li} Q_i + Q_l \Big),$$

since $l \in N_k' \subseteq N_k^+$. Then, we have

$$x_k \leq \frac{Q_k}{\sum_{i \in N_l} a_{li} Q_i + Q_l} \quad \text{for all } k \in N_l. \tag{14}$$

Combining (13) and (14), we can obtain that

$$\begin{aligned} P_l^S &\geq p_l \cdot \Big(1 - \frac{1}{2} \sum_{k \in N_l} a_{lk} x_k \Big) \\ &\geq p_l \cdot \Big(1 - \frac{1}{2} \frac{\sum_{k \in N_l} a_{lk} Q_k}{\sum_{k \in N_l} a_{lk} Q_k + Q_l} \Big) \\ &\geq \frac{x_l}{4}. \end{aligned}$$

■

Let $\vec{d}$ denote the service rate vector. Then $d_l = P_l^S \geq \frac{x_l}{4}$ for all l by Lemma 3. Also let $m := \operatorname{argmax}_{j \in N_l^+} (\sum_{i \in N_j} a_{ji} Q_i + Q_j)$. From (11), we can obtain that

$$\begin{aligned} d_l &\geq \frac{1}{4} \cdot \frac{Q_l}{\sum_{i \in N_m} a_{mi} Q_i + Q_m} \\ &\geq \frac{1}{4(A_m + 1)} \cdot \frac{Q_l}{\|\vec{Q}\|_\Omega}. \end{aligned} \tag{15}$$

The following lemma can be proven using the fluid limits and the Lyapunov techniques.

Lemma 4: For $\eta > 0$, if a scheduling scheme satisfies $d_l \geq \eta \cdot \frac{Q_l}{\|\vec{Q}\|_\Omega}$ for all l, then the network system remains stable under the scheduling policy for any arrival rate $\vec{\lambda} \in \eta \Lambda$ (strictly inside).

The proof follows the same line of analysis of [9] and can be found in Appendix.

Let $\bar{A}$ denote an upper bound on A_l for all l. From (15) and Lemma 4, we can obtain the following proposition.

Proposition 1: The efficiency ratio of QARAS is no smaller than $\frac{1}{4(\bar{A}+1)}$.

V. Bound on the Maximum Affectance Sum

In this section, we calculate a bound on the maximum affectance sum A_l. We first obtain the bound on $\underline{R}_l$ and then calculate $\overline{R}_l$, which defines the neighbor set of link l.

Since $\underline{R}_l$ is, by definition, the largest distance d such that no other active link can be active within the distance, we have that

$$\underline{R}_l = \Big(\frac{1}{\beta} d_{ll}^{-\alpha} - \frac{N}{P} \Big)^{-1/\alpha}.$$

Note that $\underline{R}_l$ is an increasing function of d_{ll} when $\frac{1}{\beta} d_{ll}^{-\alpha} > \frac{N}{P}$, which is required to hold for a successful transmission without any interfering link. From the fact that $d_{ll} \geq 1$ by definition, we can find the lower bound on $\underline{R}_l$ when $d_{ll} = 1$. We denote the lower bound on $\underline{R}_l$ by

$$\underline{R}^{min} := \Big(\frac{1}{\beta} - \frac{N}{P} \Big)^{-1/\alpha}.$$

Let I_l^{max} denote the maximum interference (normalized by P) that can be tolerated by the receiver of link l, i.e., $I_l^{max} := \sup\{ I \mid \frac{P \cdot d_{ll}^{-\alpha}}{P \cdot I + N} > \beta \}$. Equivalently, we have

$$I_l^{max} = \frac{1}{\beta} d_{ll}^{-\alpha} - \frac{N}{P} = (\underline{R}_l)^{-\alpha}. \tag{16}$$

Also, let $I_l^{out}(\vec{C})$ denote the on the total normalized interference sum from all the active links except N_l to the receiver of link l under schedule $\vec{C}$, i.e.,

$$I_l^{out}(\vec{C}) := \sum_{k \in E \setminus N_l, k \neq l} d_{lk}^{-\alpha} \cdot C_k.$$

We define $\vec{C}'(l)$ as the schedule that causes the largest interference to l, i.e.,

$$\vec{C}'(l) := \operatorname*{argmax}_{\vec{C} \in \mathcal{C}} I_l^{out}(\vec{C}).$$

We note that $I_l^{out}(\vec{C})$ becomes the largest when the active links outside N_l are as close to link l as possible while satisfying their interference constraints. Since the minimum distance between any two active links is not smaller than $\underline{R}^{min}$, the density Δ, defined as the number of active links within a unit area, is upper bounded as

$$\Delta \leq \frac{4}{\pi (\underline{R}^{min})^2}.$$

We consider an imaginary network where the active links (of zero mass) are distributed uniformly with density Δ over space. Then when $\overline{R}_l$ is sufficiently large, for each active link $k \in \vec{C}'(l)$, we can find an exclusive area of size $\pi(\underline{R}^{min}/2)^2$ such that the farthest point of the area is the location of link k,

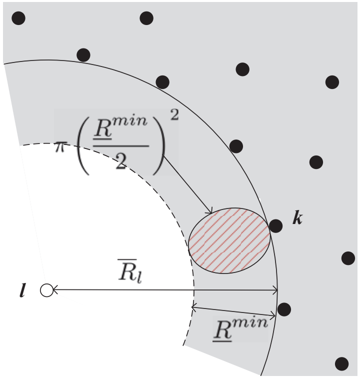


Fig. 1. Neighborhood of link l. Solid dots denote the set of active links in $\vec{C}'(l)$. Since each $k \in \vec{C}'(l)$ occupies an exclusive area of size at least $\pi(\underline{R}^{min}/2)^2$, we slide the area toward link l and find an exclusive area (shaded in the figure) of the same size where link k is located at the farthest point from link l, assuming that $\overline{R}_l$ is sufficiently large. The imaginary links (of zero mass) are uniformly distributed with density Δ in the area. Then the interference from $\vec{C}'(l)$ is no greater than the interference from the imaginary links in the grey area.

as shown in Fig. 1 (shaded area). Then the interference from link k to link l will be bounded by the interference from all the imaginary links in the shaded area, and thus the interference from $\vec{C}'(l)$ will be no greater than the interference from the imaginary links in the grey area in Fig. 1.

Hence, we can bound the interference from the active links outside N_l under $\vec{C}'(l)$, by integrating the interference from all the imaginary links with distance larger than $(\overline{R}_l - \underline{R}^{min})$, i.e.,

$$\begin{aligned} I_l^{out}(\vec{C}'(l)) &\le \int_{\overline{R}_l-\underline{R}^{min}}^{\infty} r^{-\alpha}\cdot\Delta\cdot 2\pi r\cdot dr \\ &= \frac{1}{\left(\underline{R}^{min}\right)^2}\cdot\frac{8}{\alpha-2}\cdot\left(\overline{R}_l-\underline{R}^{min}\right)^{2-\alpha}, \end{aligned}$$

when $\alpha > 2$. By setting $\overline{R}_l \ge \overline{R}_l(\epsilon)$, where

$$\overline{R}_l(\epsilon) := \left(\frac{\alpha-2}{8}\epsilon\left(\underline{R}^{min}\right)^2\underline{R}_l^{-\alpha}\right)^{\frac{1}{2-\alpha}} + \underline{R}^{min}, \tag{17}$$

we obtain that

$$I_l^{out}(\vec{C}) \le \epsilon\cdot(\underline{R}_l)^{-\alpha} = \epsilon I_l^{max},$$

where the last equality comes from (16). This leads to the following lemma.

Lemma 5: For any $\epsilon \in (0,1]$, if $\overline{R}_l \ge \overline{R}_l(\epsilon)$, we can achieve $I_l^{out}(\vec{C}) \le \epsilon I_l^{max}$ for all $\vec{C} \in \mathcal{C}$.

Note that we control the interference margin through ϵ, i.e., the larger ϵ is, the smaller neighboring area a link has, and the more conservatively it behaves to satisfy the interference constraint.

We use (17) to bound the maximum affectance sum A_l over active links in N_l. Note that from $d_{ll} \le d_{max}$, we have $\underline{R}_l \le \left(\frac{d_{max}^{-\alpha}}{\beta} - \frac{N}{P}\right)^{-\frac{1}{\alpha}} =: \underline{R}^{max}$. Also, for a feasible schedule $\vec{C}$, we have

$$A_l = \sum_{k\in\check{N}_l} a_{lk}C_k + \sum_{k\in N_l\setminus\check{N}_l} a_{lk}C_k.$$

where $\check{N}_l := \{k \in N_l \mid d_{lk} \le \underline{R}_l\}$.

Since $a_{lk} = 1$ for $k \in \check{N}_l$, the first term will be bounded by the number active links in $\check{N}_l$ as

$$\sum_{k\in\check{N}_l} a_{lk}C_k \le \left(\frac{\underline{R}_l}{\underline{R}^{min}/2}\right)^2.$$

For the second term, we again consider an imaginary network, where the links are uniformly distributed. Similarly, we can bound the interference by integrating the interference from the imaginary links of distance larger than $\underline{R}_l - \underline{R}^{min}$. But when $\underline{R}_l \le \underline{R}^{min}$, the affectance equals 1 and there can be at most $\left(\frac{\underline{R}_l}{\underline{R}^{min}/2}\right)^2$ such active links. Hence, the second term can be bounded as

$$\sum_{k\in N_l\setminus\check{N}_l} a_{lk}C_k \le \left(\frac{\underline{R}_l}{\underline{R}^{min}/2}\right)^2 + \int_{\underline{R}_l}^{\overline{R}_l(\epsilon)} \Delta\cdot 2\pi r\cdot a_{lk}(r)\cdot dr,$$

where we emphasize that the affectance $a_{lk}(r)$ is a function of distance r. This can be derived as

$$\begin{aligned} &\int_{\underline{R}_l}^{\overline{R}_l(\epsilon)} \Delta\cdot 2\pi r\cdot a_{lk}(r)\cdot dr \\ &\le \frac{8}{(\underline{R}^{min})^2}\int_{\underline{R}_l}^{\overline{R}_l(\epsilon)} \frac{r^{1-\alpha}}{\underline{R}_l^{-\alpha}}dr \\ &= \frac{8}{\alpha-2}\cdot\frac{\underline{R}_l^{\alpha}}{(\underline{R}^{min})^2}\cdot\left(\frac{1}{(\underline{R}_l)^{\alpha-2}} - \frac{1}{(\overline{R}_l(\epsilon))^{\alpha-2}}\right). \end{aligned} \tag{18}$$

By construction, the maximum of (18) will be achieved when $\underline{R}_l = \underline{R}^{max}$. Letting $\theta := \frac{\underline{R}^{min}}{\underline{R}^{max}}$ and combining the two results, we have

$$\begin{aligned} A_l &\le 2\left(\frac{\underline{R}^{max}}{\underline{R}^{min}/2}\right)^2 + \frac{8}{\alpha-2}\cdot\frac{(\underline{R}^{max})^{\alpha}}{(\underline{R}^{min})^2}\cdot\frac{1}{(\underline{R}^{max})^{\alpha-2}} \\ &\quad - \epsilon/\left(1+\left(\epsilon\frac{\alpha-2}{8}\left(\frac{\underline{R}^{min}}{\underline{R}^{max}}\right)^{\alpha}\right)^{\frac{1}{\alpha-2}}\right)^{\alpha-2} \\ &= \frac{8}{\theta^2} + \frac{8}{\alpha-2}\cdot\frac{1}{\theta^2} - \frac{\epsilon}{(1+(\epsilon\frac{\alpha-2}{8}\theta^{\alpha})^{\frac{1}{\alpha-2}})^{\alpha-2}} \\ &\le \frac{8(\alpha-1)}{\alpha-2}\cdot\frac{1}{\theta^2} - \frac{\epsilon}{2^{\alpha-2}}, \end{aligned} \tag{19}$$

where the last inequality comes from the fact that $\epsilon < 1$, $\theta \le 1$, and α is commonly less than 8. This results in the following proposition.

Proposition 2: For all link l, the maximum affectance sum is bounded by

$$A_l \le \bar{A}(\epsilon) \le \bar{A}, \tag{20}$$

where $\bar{A} := \frac{8(\alpha-1)}{\alpha-2}\cdot(\frac{\underline{R}^{max}}{\underline{R}^{min}})^2$ and $\bar{A}(\epsilon) := \bar{A} - \frac{\epsilon}{2^{\alpha-2}}$.

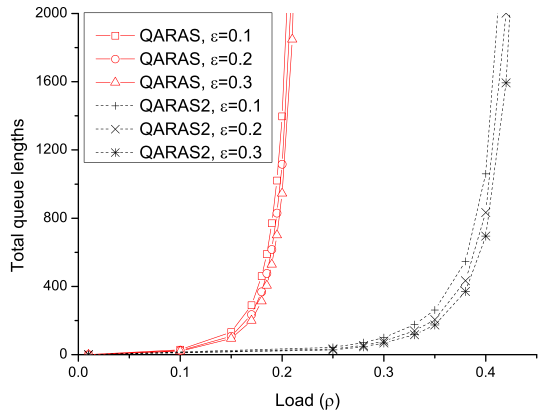


Fig. 2. Performance of QARAS and QARAS2 in a grid network.

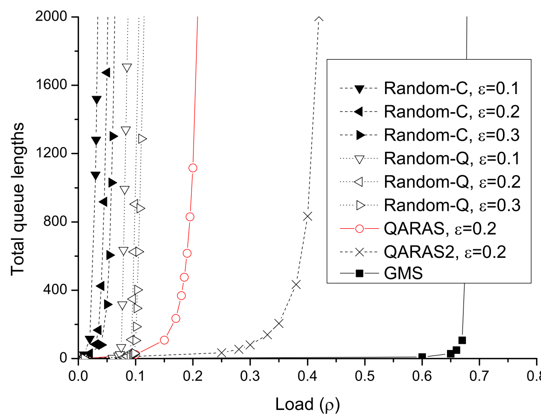


Fig. 3. Performance comparison in a grid network.

Note that $\bar{A}(\epsilon)$ is a decreasing function of ϵ. As ϵ increases, the neighborhood area featured by $\overline{R}_l(\epsilon)$ will shrink, and the expected interference sum from the active neighboring links will decrease.

Remarks: It has been shown in [12] that a distributed version of Longest-Queue-First (LQF) achieves a non-vanishing fraction of the optimal throughput. Our result along with Proposition 1 provides better bound than [12] by taking into account the cumulative interference through affectance.

VI. Simulation Results

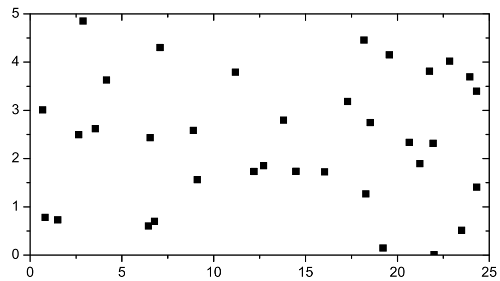
We evaluate empirical performance of QARAS through simulations. We first evaluate its performance with different ϵ values in a grid topology and compare it with other scheduling schemes, including the Greedy Maximal Scheduling (GMS) scheme and conventional random access scheduling schemes under the conflict interference model. We also evaluate the scheduling performance in a random network, where the topology and the traffic are generated at random.

We consider a grid topology network of 49 links. For each link l, we place a vertex with parameter $\hat{d}_{ll}$, and the distance between two vertices l and k indicates $\hat{d}_{lk}$. In the grid topology, we set $\hat{d}_{ll}$ to unit distance (m) for all l, and the minimum distance from a vertex to another is set to 2.1 m. We set the transmission power P for each link to 10 dBm and the noise power N to -96 dBm. The path loss exponent α is set to 4, and the SINR threshold β is set to 6 dB.

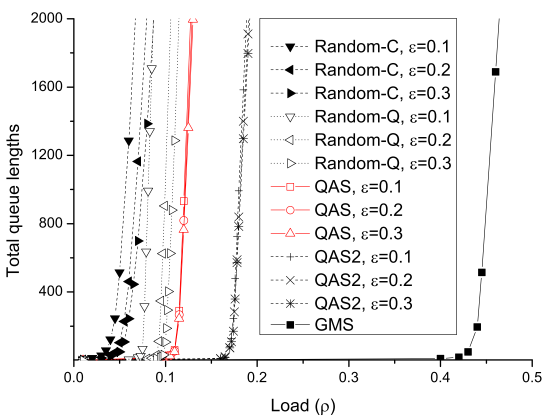
Packet arrivals at each link follow an i.i.d. Bernoulli process with an identical mean rate ρ. At each time slot, a scheduled link can transmit one packet if its SINR constraint is satisfied, and cannot make a successful transmission otherwise. We simulate $10,000$ time slots and measure the total average queue lengths when the simulation terminates. We present an average of 10 simulation runs for each result. We evaluate the performance of the scheduling schemes by increasing ρ. When the traffic load gets closer to the achievable region of the scheduling schemes, we will observe a surge of the queue lengths.

Fig. 2 demonstrates the performance of QARAS and QARAS2 with different ϵ. QARAS2 works as QARAS except that it sets $p_l = x_l$. The motivation of QARAS2 is that a link with no neighboring link can refrain from accessing the medium due to the factor $\frac{1}{2}$. The results show that the performance of QARAS and QARAS2 is insensitive to ϵ. Note that ϵ changes the neighbor set N_l; the smaller ϵ is, the larger N_l link l has. While we have increased ϵ from 0.1 to 0.3, the bound of their achievable region does not change, around 0.19 for QARAS and around 0.38 for QARAS2. The bound remains unchanged even when we increase ϵ to 1. Also, the results show that QARAS2 outperforms QARAS and achieves as almost twice throughput as QARAS. This may imply that the parameters of QARAS are conservative for the practical use.

In Fig. 3, we compare our proposed schemes with other scheduling schemes. GMS schedules the links in the decreasing order of queue length, conforming to the interference constraints. It assumes that one can check the scheduling feasibility when a link is added to the schedule. Since GMS has been known to empirically achieve close-to optimal throughput [9], we consider the performance of GMS as the reference to the bound of the capacity region. We also consider two conventional random access schemes. Random-C considers the cardinality of the neighboring links as in RMS and set $p_l = \frac{1}{|N_l^+|}$. Random-Q considers the relative queue lengths among its neighboring links as in [9] and sets $p_l = \frac{Q_l}{\sum_{k \in N_l} Q_k}$. The results show that our proposed QARAS outperforms both Random-C and Random-Q. It is because Random-C and Random-Q have excessively conservative attempt probability, or they define the neighbor set excessively large. Indeed, it is very difficult for a link to make a binary decision for its neighbor set under the SINR interference model. Further in the results, when $\epsilon = 0.2$, the largest (measured) affectance sum is 2, and thus the performance bound of QARAS is $\frac{1}{12}$ from Proposition 1. However, the numerical results show that the empirical performance of QARAS is much better than the theoretical bound and its ratio to the performance of GMS is close to 0.35.



(a) Topology of 35 links (a vertex represents a link)



(b) Performance

Fig. 4. Performance in a random topology with different ϵ.

Remarks: There are several scheduling schemes that are throughput-optimal in the binary interference model [3], [10], [11]. They require global information to make a scheduling decision under the SINR interference model, and cannot be directly compared with our solution. For example, throughput-optimal CSMA schemes require for each transmitter to have either global information [23], or to be able to check whether its transmission will not violate the interference constraint of other existing transmissions [10], [11], [13]. Under the binary symmetric interference model, it can be easily checked through the carrier-sensing. However, under the SINR interference model, the local carrier-sensing cannot provide the information and global information should be collected.

We also compare the schemes in a random topology network where 35 links are randomly placed in 25×5 area as shown in Fig. 4(a). In this topology, $\hat{d}_{ll}$ is chosen uniformly at random in range $[0.8, 1.2]$, and the packet arrival rate of each link follows an independent Bernoulli processes with mean $\rho \cdot \lambda_l$, where λ_l is randomly chosen between 0.2 (with probability 0.3), 0.5 (with probability 0.4), 1 (with probability 0.3). The other parameters are the same as in the grid network.

Fig. 4(b) presents the total queue lengths of the scheduling schemes with different traffic loads ρ. The results are similar to those in the grid network. Our proposed QARAS and QARAS2 achieve larger throughput than Random-C and Random-Q that do not consider the interference level from the neighbors. Note that when $\epsilon = 0.2$, the largest measured affectance sum is 3.1769. The theoretical performance bound of QARAS is around $\frac{1}{16}$, but the empirical performance is much better as in the grid network, around $\frac{1}{4}$.

VII. Conclusion

In this paper, we are interested in developing low-complexity scheduling schemes under the SINR interference model, such that they are provably efficient and amenable to distributed implementation without requiring global information. Due to the complex structure of the SINR interference constraints, most previous low-complexity scheduling schemes have assumed that a means to check the feasibility of a schedule is available, which, however, comes with a significant amount of overhead and often requires network-wide information exchanges. We develop low-complexity random access scheduling schemes that use only the local information of queue lengths and the interference level from neighboring links, and operate without the feasibility check. By investigating the properties of the cumulative interference through the notion of affectance and its relationship with the capacity region, we analytically show that the proposed scheduling schemes achieve provable efficiency. The numerical results show that our proposed scheme outperforms other random access scheduling schemes without considering the interference levels of the neighboring links, and empirically achieve higher performance than the theoretical result.

Acknowledgements

C. Joo is the corresponding author. M. Shin was with UNIST. This work is in part supported by Basic Science Research Program through the NRF of Korea funded by the MSIP (No. NRF-2014R1A1A1A05002557), and in part by IITP grant funded by the Korea government (MSIP) (No. B0126-15-1064, Research on Near-Zero Latency Network for 5G Immersive Service).

References

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[2] W. L. Tan, P. Hu, and M. Portmann, "Experimental Evaluation of Measurement-based SINR Interference Models," in *WoWMoM*, June 2012.

[3] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximal Throughput in Multihop Radio Networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, December 1992.

[4] C. Joo, G. Sharma, N. B. Shroff, and R. R. Mazumdar, "On the Complexity of Scheduling in Wireless Networks," *EURASIP Journal of Wireless Communications and Networking*, October 2010.

[5] X. Lin and N. B. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Congestion Control in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 302–315, April 2006.

[6] L. Tassiulas, "Linear Complexity Algorithms for Maximum Throughput in Radio Networks and Input Queued Switches," in *IEEE INFOCOM*, April 1998, pp. 533–539.

[7] E. Modiano, D. Shah, and G. Zussman, “Maximizing Throughput in Wireless Networks via Gossiping,” *Sigmetrics Performance Evaluation Review*, vol. 34, no. 1, pp. 27–38, 2006.
[8] X. Lin and S. Rasool, “Constant-Time Distributed Scheduling Policies for Ad Hoc Wireless Networks,” *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 231–242, Feburary 2009.
[9] C. Joo and N. B. Shroff, “Performance of Random Access Scheduling Schemes in Multi-hop Wireless Networks,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, October 2009.
[10] L. Jiang and J. Walrand, “A Distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks,” *IEEE/ACM Trans. Netw.*, vol. 18, no. 13, pp. 960–972, June 2010.
[11] J. Ni, B. Tan, and R. Srikant, “Q-CSMA: Queue-Length Based CSMA/CA Algorithms for Achieving Maximum Throughput and Low Delay in Wireless Networks,” *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, June 2012.
[12] L. Le, E. Modiano, C. Joo, and N. B. Shroff, “Longest-Queue-First Scheduling Under SINR Interface Model,” in *ACM MobiHoc*, September 2010.
[13] J.-G. Choi, C. Joo, J. Zhang, and N. B. Shroff, “Distributed Link Scheduling Under SINR Model in Multihop Wireless Networks,” *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1204–1217, Aug 2014.
[14] S. Lv, X. Wang, and X. Zhou, “Scheduling under SINR Model in Ad Hoc Networks with Successive Interference Cancellation,” in *IEEE GLOBECOM*, December 2010.
[15] D. Qian, D. Zheng, J. Zhang, N. Shroff, and C. Joo, “Distributed CSMA Algorithms for Link Scheduling in Multihop MIMO Networks Under SINR Model,” *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 746–759, June 2013.
[16] S. Borbash and A. Ephremides, “Wireless Link Scheduling With Power Control and SINR Constraints,” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5106–5111, Nov 2006.
[17] M. M. Halldórsson and R. Wattenhofer, “Wireless Communication is in APX,” in *ICALP*, 2009, pp. 525–536.
[18] P.-J. Wan, X. Jia, and F. Yao, “Maximum Independent Set of Links under Physical Interference Model,” in *WASA*, 2009.
[19] T. Kesselheim, “A Constant-factor Approximation for Wireless Capacity Maximization with Power Control in the SINR Model,” in *ACM-SIAM SODA*, 2011, pp. 1549–1559.
[20] M. Halldorsson and P. Mitra, “Wireless Capacity and Admission Control in Cognitive Radio,” in *IEEE INFOCOM*, March 2012, pp. 855–863.
[21] G. Pei and A. K. S. Vullikanti, “Brief Announcement: Distributed Algorithms for Maximum Link Scheduling in the Physical Interference Model,” in *DISC*, 2012, pp. 407–408.
[22] T. Kesselheim and B. Vöcking, “Distributed Contention Resolution in Wireless Networks,” in *DISC*, 2010.
[23] P. Chaporkar and A. Proutiere, “Optimal Distributed Scheduling in Wireless Networks under SINR Interference Model,” in *the 51st Annual Allerton Conference on Communications, Control, and Computing*, 2013.
[24] C. Joo, X. Lin, J. Ryu, and N. B. Shroff, “Distributed Greedy Approximation to Maximum Weighted Independent Set for Scheduling with Fading Channels,” in *ACM MobiHoc*, 2013.
[25] X. Lin and S. Rasool, “Constant-Time Distributed Scheduling Policies for Ad Hoc Wireless Networks,” in *IEEE CDC*, December 2006.
[26] A. Gupta, X. Lin, and R. Srikant, “Low-Complexity Distributed Scheduling Algorithms for Wireless Networks,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1846–1859, December 2009.
[27] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, “Scheduling in a Queueing System with Asynchronously Varying Service Rates,” *Probab. Eng. Inf. Sci.*, vol. 18, no. 2, pp. 191–217, Apr. 2004.
[28] J. G. Dai, “On Positive Harris Recurrence of Multiclass Queueing Networks: A Unified Approach via Fluid Limit Models,” *Annals of Applied Probability*, vol. 5, no. 1, pp. 49–77, 1995.

APPENDIX
PROOF OF LEMMA 4

To prove the system stability, we use the fluid limits and the Lyapunov techniques as in [9]. We extend the discrete arrival and departure processes to a continuous process by defining $X_l(t) = X_l(\lfloor t \rfloor)$, where $\lfloor t \rfloor$ is the largest integer no greater than t, and also interpolate the queue length to a continuous process. Then there exist limits $q_l(t)$ and $d_l(t)$ such that $q_l(t)$ is absolutely continuous on t, and as $n \to \infty$,

$$\begin{aligned}
&\frac{1}{n}\int_0^{nt} R_l(s)ds \to \lambda_l t,\\
&\frac{1}{n}\int_0^{nt} D_l(s)ds \to \int_0^t d_l(s)ds,\\
&\qquad \frac{1}{n} Q_l(nt) \to q_l(t),
\end{aligned}$$

and they satisfy

$$\begin{aligned}
&q_l(t) \geq 0,\\
&\frac{d}{dt} q_l(t) = \begin{cases} \lambda_l - d_l(t), & \text{if } \lambda_l - d_l(t) \geq 0 \text{ or } q_l(t) > 0,\\ 0, & \text{otherwise,}\end{cases}\\
&d_l(t) \geq \eta \cdot \frac{q_l(t)}{\|\vec{q}(t)\|_\Omega}.
\end{aligned} \tag{21}$$

We consider the following Lyapunov function:

$$V(t) := \|\vec{q}(t)\|_\Omega.$$

It has been well-known that if the Lyapunov function has a negative drift, then the system is stable. Since $\vec{q}(t)$ is continuous, for any $\vec{q}(t) \neq 0$, there exist a small $\delta t > 0$ such that for all l, $q_l(t) - d_l(t)\delta t \geq 0$ for $[t, t+\delta t]$. For such $\vec{q}(t) \neq 0$ and δt, we have that

$$\begin{aligned}
V(t+\delta t) &\leq \|\vec{q}(t) + \vec{\lambda}\delta t - \vec{d}(t)\delta t\|_\Omega + o(\delta t)\\
&\leq \|\vec{q}(t) - \vec{d}(t)\delta t\|_\Omega + \|\vec{\lambda}\|_\Omega \delta t + o(\delta t)\\
&\quad \text{(by Part 5 of Lemma 2)}\\
&\leq \left\| \vec{q}(t) \cdot \left(1 - \frac{\eta}{\|\vec{q}(t)\|_\Omega}\delta t\right)\right\|_\Omega + \|\vec{\lambda}\|_\Omega \delta t + o(\delta t)\\
&\quad \text{(by (21))}\\
&= \|\vec{q}(t)\|_\Omega - \eta \cdot \delta t + \|\vec{\lambda}\|_\Omega \delta t + o(\delta t)\\
&\quad \text{(by Part 3 of Lemma 2)}\\
&= V(\vec{q}(t)) + (\|\vec{\lambda}\|_\Omega - \eta)\delta t + o(\delta t).
\end{aligned}$$

Hence, if $\|\vec{\lambda}\|_\Omega - \eta < 0$, then $\limsup_{\delta t \downarrow 0} \frac{V(t+\delta t) - V(t)}{\delta t} < 0$. This implies that for any arrival rate $\vec{\lambda} \in \eta\Lambda$ (strictly inside), the Lyapunov function has a negative drift, and thus the fluid limits system is stable. Then from [28], the original system is also stable.