Optimal Delay Bound for Maximum Weight Scheduling Policy in Wireless Networks

Cem Boyacı and Ye Xia
Computer and Information Science and Engineering Department
University of Florida, Gainesville, Florida, USA
Emails: cem.boyaci@gmail.com, yx1@cise.ufl.edu

***Abstract*—We provide an improved bound for the expectation of the stationary delay under the maximum weight scheduling (MWS) policy in one-hop wireless networks. In the model, the interference of the links is characterized by an interference graph $G = (V, E)$. For a vector $\mu \in \mathbb{R}_+^{|V|}$, let $\chi_f(G, \mu)$ be the weighted fractional coloring number for the graph G under the weight vector μ. For an arrival rate vector λ in the capacity region, we define a quantity $\epsilon^{(\lambda)}$ to be the unique value satisfying the condition $\chi_f(G, \lambda+\epsilon^{(\lambda)}e) = 1$, where $e = (1, 1, \ldots, 1)^T$. We show that the stationary delay is upper-bounded by $B/(2(\sum_{i=1}^{|V|} \lambda_i)\epsilon^{(\lambda)})$, where B is a constant depending on the arrival process. We show that the new bound is the tightest single-parameter bound obtainable with an often-used analytical framework. Generalizing the above, we also provide the tightest bounds for all MWS-w policies.**

***Index Terms*—wireless link scheduling; delay; maximum weight schedule; stability; queue size**

I. Introduction

In many wireless communication systems, wireless links share a common communication medium and simultaneous link transmissions may cause interference. Researchers address the interference problem by devising link scheduling policies (algorithms) that optimize performance objectives of choice. Throughput has been considered as one of the most important performance metrics and the throughput performance of various scheduling algorithms have been intensively studied. Recently, researchers are increasingly interested in the delay performance of link scheduling algorithms [1] [2]. This paper studies the expected delay and the sum of the expected queue sizes[1] under the well-known maximum weight scheduling (MWS) policy and one of its generalizations, which we call MWS-w.

Link interference may have different consequences in different families of wireless communication technologies. The paper focuses on one of the important cases, known as the *protocol interference model*, in which the transmission of a link (i.e., a transmitter-receiver pair) can be successful only when there is no interference from other link transmissions.

In [3], Tassiulas et al. proposed the MWS policy and showed that it is throughput optimal when applied to the protocol model. Under the MWS policy, the set of links that are scheduled for transmission on each time slot corresponds to a weighted maximum independent set (WMIS) of the interference graph, where the weight of a node is the queue length of the corresponding wireless link. Later, researchers considered a generalization, the MWS-w policy, which is also throughput optimal [2]. The difference is that, in MWS-w where $w = (w_i)_i$ is a positive vector, the weight of each node i is equal to its queue length multiplied by w_i. It has been observed experimentally that the queue sizes and delays are small under the MWS policy for the 2-hop interference model, which is a special case of the protocol interference model [2]. It is worth pointing out that the WMIS problem is NP-hard in general. Hence, the good throughput and delay performance does not come for free.

The most relevant prior work to this paper includes [1] [2] [4]. In [4], we presented an upper bound for the expected delay under the MWS policy, which is an improvement over the bound given in [1]. The bounds in [1] and [4] are what we call *single-parameter bounds*. Given an arrival rate vector λ in the interior of the capacity region, a single parameter is created to characterize the traffic intensity to the entire network system and the bounds directly depend on such a parameter. The relevant bound for an MWS-w policy in [2], with an appropriately chosen w, is a multi-parameter bound.

This paper not only presents improved single-parameter bounds, but asks how much more improvement can be made. Applying the analytical framework used in [2] [4], we first generate a family of bounds. Then, we find the tightest bound by minimizing over the entire family. As a result, we show that the new bound is the smallest single-parameter bound for the MWS policy that are obtainable through the aforementioned framework. Generalizing that result, we have also derived the tightest bound under each MWS-w policy, one for each fixed w. Furthermore, we show that an important bound found in [2] is the best multi-parameter bound for all MWS-w policies. In developing the new bounds, we are able to relate the new bounds with those in [2] [4].

On the theoretical side, in both [4] and the current paper, we have made extensive use of graph-theoretical quantities, such as the fractional coloring number. The connection with graph theory was instrumental in revealing the hardness of the wireless scheduling problem in prior work [5] [6]. Further exploration of the connection can be fruitful in the future. In addition, the graph-theoretical quantities used in this paper can be defined by linear optimization problems. The paper emphasizes the geometric aspects of these problems and provides geometric interpretation on where the tightest bounds are achieved.

We next provide some additional discussion about the

[1]The two are related through Little's Law, and having one gives the other.

978-1-4799-3360-0/14/$31.00 © 2014 IEEE

related literature. In [2], the aforementioned multi-parameter bound is used to show that the expected delay under that MWS-w policy is at least as good as any randomized policy that assigns constant activation probabilities to the schedules. In addition, [2] contains results on the lower bound for the expected delay under any scheduling policy. In recent years, there has been an increasing number of studies on the delay performance under various other scheduling algorithms. For example, in [7], the authors propose a backpressure algorithm for multi-hop networks using first-in-first-out (FIFO) queueing. The algorithm is proven to be utility-optimal and a delay bound is given. Our work in [4] also contains delay bounds under the longest-queue-first policy and approximate MWS policies. Sample studies about the delay performance under CSMA-like randomized algorithms include [8] and [9]. Previous research on network-switch scheduling supplied many results that are applicable to wireless link scheduling with appropriate modifications [10] [11]. In [11], the authors investigate bounds for the expected queue size in virtual-output-queued switches. Their object of study corresponds to a special case of the wireless link scheduling problem, which is the node exclusive interference model on a complete bipartite network graph. The authors of [1] provide a generalization for arbitrary networks and interference relationship under the MWS algorithm.

There is a stream of literature that studies the expected queue sum under the MWS-w policy in the heavy traffic regime [12] [13]. The relevant result in [13] is roughly as follows. Let ν be a point in the relative interior of a facet of the capacity region whose outer normal is c (also with unit length). Let $\lambda^{\epsilon} = \nu - \epsilon c$ for $\epsilon > 0$. Under any scheduling policy, the expectation of the weighted queue sum, $E[c^T Q]$, grows at least as fast as Z/ϵ as $\epsilon \to 0$, where Z is a constant depending on c; the MWS policy achieves ζ. In our case, the arrival rate vector λ is fixed. The expected queue sum is upper-bounded by $\frac{1}{2}B$ divided by a "distance" of λ to the boundary of the capacity region, where B is constant. We ask: What is the best measure of distance that gives the tightest upper bound under the MWS policy?

This paper is structured as follows. In Section II, we provide the system model. Useful graph theoretical quantities and supporting lemmas are given in Section III. In Section IV, we derive single-parameter bounds for the expected queue-size sum and the expected delay under the MWS and MWS-w policies. In Section V, we show the new bounds are the best of a family. We also show that the multi-parameter bound in [2] is also the tightest and how it is related to our bounds. Additional theoretical results are given in Section VI. The conclusions are in Section VII.

II. System Models

We assume the following widely used models.

A. Network and Interference Model

We consider a wireless network with one-hop traffic. That is, any data is transmitted only once, and after that, it leaves the network. Let L be the set of the wireless links. The system is time-slotted, the packet sizes are identical, and the capacity of every link is one packet per time slot.

We model the interference relations of the wireless links with an interference graph, $G = (V, E)$. Each wireless link in the physical network is represented by a node $v \in V$. Two nodes $v_1, v_2 \in V$ are connected in G if and only if the corresponding links in the network interfere with each other. We assume symmetric interference relation; thus G is undirected. This interference model is also known as the *protocol interference model*. Let the number of nodes in G be denoted by N, i.e., $N = |V|$. The links in L and the nodes in V are both indexed as $1, 2, \ldots, N$ and they have one-to-one correspondence in that order.

A *feasible schedule* is defined to be a set of non-interfering nodes in G. A *maximal schedule* is a feasible schedule that cannot include any more nodes without causing interference. A feasible schedule corresponds to an independent set in G. We denote the set of all maximal schedules by M_L. When applicable, M_L is also regarded as a $N \times |M_L|$ 0-1 matrix. Each column of the matrix is a 0-1 vector representation of a maximal independent set of G, with 1 indicating that the corresponding node (in G) is selected (and the corresponding wireless link is activated in the schedule) and 0 otherwise. Throughout the paper, we use the term *schedule* to refer to a maximal schedule, unless mentioned otherwise.

B. Traffic Model

The packets arriving for a wireless link are queued at the transmitter (sender) of the link. There is one queue for each link, or equivalently, there is one queue associated with each node in the interference graph. Each arrival process is i.i.d. in time; and the arrival processes for all the links are mutually independent. Let $A_i(t)$ denote the number of packets arriving at link i at time t. Since the distribution of $A_i(t)$ is time invariant, we use $E[A_i^k]$ to denote its k-th moment, if it exists. We assume the second moment $\mathrm{E}[A_i^2]$ exists for every i. Let $A(t)$ denote the vector $(A_i(t))_{i \in V}$. Let $Q(t)$ be the queue length vector, i.e., $Q(t) = (Q_i(t))_{i \in V}$, where $Q_i(t)$ is the queue length of link i (or equivalently, node i).

A link can transmit a packet on a time slot only if it is selected in the current schedule and its queue is non-empty. At each queue, at most one packet can be served on any time slot. For each link i, let $D_i(t)$ indicate whether or not link i transmits a packet on time slot t. Note that

$$D_i(t) = \begin{cases} 1, & \text{if } Q_i(t) \geq 1 \text{ and } i \text{ is scheduled.} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Since the capacity of each link is one packet per time slot, $D_i(t)$ is also the number of departures from queue i at time t. Let $D(t) = (D_i(t))_{i \in V}$. The evolution of the queues is as follows:

$$Q(t+1) = Q(t) + A(t) - D(t). \tag{2}$$

C. System Stability and Scheduling Policies

We adopt the conventional definition of the capacity region Λ (see [3]).

$$\Lambda = \{\lambda \mid 0 \leq \ \lambda \leq \mu \text{ for some } \mu \in Co(M_L)\},$$

where $Co(M_L)$ stands for the convex hull of the column vectors in M_L. The interior of the capacity region, denoted by Λ^o, is defined as:

$$\Lambda^o = \{\lambda \mid 0 \leq \lambda < \mu \text{ for some } \mu \in Co(M_L)\}.$$

$\Lambda \backslash \Lambda^o$ is called the boundary of Λ.

Stability of the queueing system is defined as: The irreducible discrete-time Markov chain (DTMC) that represents the system is positive recurrent. It has been shown in [3] that the stability region is Λ^o. A scheduling algorithm is said to be *throughput-optimal* if it keeps the queues stable under any arrival rate vector in Λ^o.

In general, for any vector x, let x_i denote its i-th component. For a vector or matrix x, let x^T denote its transpose. Let $e = (1, 1, \ldots, 1)^T$; its dimension depends on the context.

The MWS policy is defined as follows. The selected schedule on each time slot t, denoted by $m(t)$, has the maximum total weight; that is, $m(t)^T Q(t) \geq m^T Q(t)$ for all $m \in M_L$. The weight of node i is its queue size at time t, $Q_i(t)$. The MWS-w policy will be defined when needed.

III. Preliminary: Graph Theoretical Quantities

This section presents key concepts, notations and supporting lemmas related to certain graph-theoretical quantities that are central to the new bounds.

Under the protocol interference model, the capacity region can be characterized using the concept of fractional coloring. Let $\lambda \in \mathbb{R}^N_+$ be a (component-wise) non-negative vector. Consider the following optimization problem, where $\alpha = (\alpha_m)_{m \in M_L}$ is the decision variable.

$$\chi_f(G, \lambda) \triangleq \min \; e^T\alpha, \text{ subject to } M_L\alpha \geq \lambda, \, \alpha \geq 0. \quad (3)$$

The optimal value $\chi_f(G, \lambda)$ is called the *weighted fractional coloring/chromatic number*, where λ is the weight vector.

We denote the *weighted independence number* of graph G, i.e., the total weight of a WMIS, by $\pi(G, z)$, where $z \in \mathbb{R}^N_+$ is the weight vector. We denote $\chi_f(G, e)$ and $\pi(G, e)$ by $\chi_f(G)$ and $\pi(G)$, respectively. Then, $\chi_f(G)$ is the fractional coloring number and $\pi(G)$ is the size of a maximum independent set.

When λ is the arrival rate vector to the queues in the wireless network, $\chi_f(G, \lambda)$ can be understood as the traffic load or intensity to the queueing system. The following is an easy extension of a result in [5].

Lemma 1. *(i) For each $k > 0$, $\lambda \in k\Lambda^o$ if and only if $\chi_f(G, \lambda) < k$; (ii) $\lambda \in \Lambda \backslash \Lambda^o$ if and only if $\chi_f(G, \lambda) = 1$.*

The following simple properties hold. Recall that $\chi_f(G, \lambda)$ is only defined on $\mathbb{R}^N_+$.

Lemma 2. *(i) For any scalar $b \geq 0$ and $\lambda \in \mathbb{R}^N_+$, $\chi_f(G, b\lambda) = b\chi_f(G, \lambda)$; (ii) $\chi_f(G, \lambda) = 0$ if and only if $\lambda = 0$.*

It can be checked that $\chi_f(G, \lambda)$ satisfies the triangle inequality.

Lemma 3. *For any $\lambda, \mu \in \mathbb{R}^N_+$,*

$$\chi_f(G, \lambda) + \chi_f(G, \mu) \geq \chi_f(G, \lambda + \mu), \quad (4)$$

where equality is achieved if $\lambda = b\mu$ for some $b \geq 0$.

Thus, $\chi_f(G, \lambda)$ possesses norm-like properties. However, its domain of definition is $\mathbb{R}^N_+$, which is not a vector space.

Lemma 4. *For $\lambda \in \mathbb{R}^N_+$, $\max_i \lambda_i \leq \chi_f(G, \lambda) \leq e^T\lambda$.*

Proof: Consider an arbitrary $\alpha \geq 0$ such that $M_L\alpha \geq \lambda$. Since the entries of the matrix M_L are 0 or 1, we have $e^T\alpha \geq (M_L\alpha)_i$ for all i. Thus, we have $e^T\alpha \geq \lambda_i$ for all i. Hence, $\chi_f(G, \lambda) \geq \lambda_i$ for all i.

For the upper bound, consider the dual problem to the problem in (3).

$$w_f(G, \lambda) \triangleq \max \lambda^T\beta, \text{ subject to } M_L^T\beta \leq e, \, \beta \geq 0. \quad (5)$$

Note that the matrix M_L cannot have a row with all zeros. Let the column vectors of M_L^T be denoted by $c_1, c_2, \ldots, c_N$. Hence, $c_i \neq 0$ for all i. Then, the inequality $M_L^T\beta \leq e$ can be written as $\sum_{i=1}^N \beta_i c_i \leq e$. For each $i = 1, 2, \ldots, N$, we must have $\beta_i c_i \leq e$. Since c_i is a 0-1 valued vector with at least one non-zero entry, we see that $\beta_i \leq 1$ for every i. Thus, $\beta \leq e$ and the objective value $\lambda^T\beta \leq \lambda^T e$. ■

The optimal value of the problem in (5), $w_f(G, \lambda)$, is known as the *fractional weighted clique number*.

IV. Improved Single-Parameter Bounds for Delay and Queue Sum

In this section, we define a new parameter $\epsilon^{(\lambda)}$ for $\lambda \in \Lambda^o$ and use it to derive new bounds for the expected delay and the sum of the expected queue sizes under the MWS policy. We show that the new bounds are improvement over earlier single-parameter bounds.

A. Improved Bounds through $\epsilon^{(\lambda)}$

In our earlier work [4], we provided the following bound for the stationary queue sum under MWS. For $\lambda \in \Lambda^o$,

$$\mathrm{E}[\sum_{i=1}^N Q_i(t)] \leq \frac{B}{2\zeta^{(\lambda)}}, \quad (6)$$

where $B = \sum_{i=1}^N (\lambda_i + \mathrm{E}[A_i^2] - 2\lambda_i^2)$ and $\zeta^{(\lambda)} = \frac{1 - \chi_f(G,\lambda)}{\chi_f(G)}$.

Next, we will define another scalar $\epsilon^{(\lambda)}$ and show that it can replace $\zeta^{(\lambda)}$ and provide a better bound.

Definition 1. *For any $\lambda \in \Lambda$, define $\epsilon^{(\lambda)}$ to be the unique value satisfying $\chi_f(G, \lambda + e\epsilon^{(\lambda)}) = 1$.*

We will check that $\epsilon^{(\lambda)}$ is well defined. It is convenient to define $\epsilon^{(\lambda)}$ for all $\lambda \in \hat{\Lambda}$, where $\hat{\Lambda} \triangleq \Lambda + \{be : b \in \mathbb{R}_+\}$.

Lemma 5. *For any $\lambda \in \hat{\Lambda}$, the solution ϵ that satisfies $\chi_f(G, \lambda + \epsilon e) = 1$ exists and is unique.*

Proof: The existence of such ϵ is trivial since $\lambda + \epsilon e$ hits $\Lambda \backslash \Lambda^o$ for the right value of ϵ. Next, suppose there exists $\epsilon' \neq \epsilon$ such that $\chi_f(G, \lambda + \epsilon' e) = 1$. Without loss of generality, suppose $\epsilon' > \epsilon$. We then have a vector $\beta \geq 0$ such that $e^T\beta = 1$ and $M_L\beta \geq \lambda + \epsilon' e > \lambda + \epsilon e$. Then, there exist a constant b, where $0 < b < 1$, and a vector $\beta' = b\beta$ such that $M_L\beta' \geq \lambda + \epsilon e$. Since $e^T\beta' = b < 1$, we get $\chi_f(G, \lambda + \epsilon e) < 1$, which is a contradiction. ■

The following can be checked.

Lemma 6. *If $\lambda \in \Lambda^o$, then $\epsilon^{(\lambda)} > 0$. If $\lambda \in \Lambda \backslash \Lambda^o$, then $\epsilon^{(\lambda)} = 0$. If $\lambda \in \hat{\Lambda} \backslash \Lambda$, then $\epsilon^{(\lambda)} < 0$.*

Lemma 7. *For $\lambda \in \hat{\Lambda}$ and $b \in \mathbb{R}$ such that $\lambda + be \in \hat{\Lambda}$,*

$$\epsilon^{(\lambda)} = b + \epsilon^{(\lambda+be)}. \tag{7}$$

Proof: $\epsilon^{(\lambda+be)}$ is equal to the value ϵ such that $\min_{\alpha \geq 0, M_L\alpha \geq \lambda+be+\epsilon e} e^T\alpha = 1$. For such ϵ, $b + \epsilon$ is equal to $\epsilon^{(\lambda)}$ by definition. ∎

Next, we show $\epsilon^{(\lambda)}$ is no less than $\zeta^{(\lambda)}$.

Lemma 8. *For $\lambda \in \Lambda$, $\epsilon^{(\lambda)} \geq \zeta^{(\lambda)}$. Equality is achieved if $\lambda = be \in \Lambda$, where $b \in \mathbb{R}_+$.*

Proof: By the triangle inequality (see Lemma 3) and the definition of $\epsilon^{(\lambda)}$, we get

$$\chi_f(G, \lambda) + \chi_f(G, \epsilon^{(\lambda)}e) \geq \chi_f(G, \lambda + e\epsilon^{(\lambda)}) = 1, \tag{8}$$

where equality is achieved if $\lambda = be$ for some $b \geq 0$. For $\lambda \in \Lambda$, by Lemma 6, $\epsilon^{(\lambda)} \geq 0$. By Lemma 2, we have $\chi_f(G, \epsilon^{(\lambda)}e) = \epsilon^{(\lambda)}\chi_f(G, e) = \epsilon^{(\lambda)}\chi_f(G)$. By substituting this result back to (8), we get

$$\epsilon^{(\lambda)} \geq \frac{1 - \chi_f(G, \lambda)}{\chi_f(G)} = \zeta^{(\lambda)},$$

where equality is achieved if $\lambda = be$ for some $b \geq 0$. ∎

Corollary 1. *If $\lambda = be \in \Lambda$, where $b \geq 0$, then*

$$\epsilon^{(\lambda)} = \frac{1 - b\chi_f(G)}{\chi_f(G)} = \zeta^{(\lambda)}.$$

The next lemma shows an equivalent definition of $\epsilon^{(\lambda)}$ by a linear optimization problem.

Lemma 9. *For a vector $\lambda \in \Lambda$, $\epsilon^{(\lambda)}$ is the optimal value of the following optimization problem.*

$$\epsilon^{(\lambda)} \triangleq \max \epsilon, \text{ subject to } M_L\alpha \geq \lambda + \epsilon e,\ e^T\alpha = 1, \alpha \geq 0. \tag{9}$$

Proof: Let ϵ^* and α^* be an optimal solution to the problem in (9). Clearly, $\chi_f(G, \lambda + \epsilon^* e) \leq 1$. Suppose $\chi_f(G, \lambda + \epsilon^* e) < 1$. Then, by Lemma 1, $\lambda + \epsilon^* e \in \Lambda^o$. By the definition of Λ^o, there exists $\mu \in Co(M_L)$ such that $\mu > \lambda + \epsilon^* e$. That is, there exists $\alpha \geq 0$ such that $e^T\alpha = 1$ and $\mu = M_L\alpha > \lambda + \epsilon^* e$. Then, there exists ϵ_o such that $\epsilon_o > \epsilon^*$ and $M_L\alpha \geq \lambda + \epsilon_o e$. This violates the optimality of ϵ^* for the problem in (9). We conclude that $\chi_f(G, \lambda + \epsilon^* e) = 1$ and hence $\epsilon^* = \epsilon^{(\lambda)}$. ∎

The next theorem provides an improved bound for the expected queue-length sum under the stationary distribution if the system is under the MWS policy.

Theorem 1. *Under the MWS policy, for any arrival rate vector $\lambda \in \Lambda^o$, the following holds under the stationary distribution.*

$$\mathrm{E}[\sum_{i=1}^{N} Q_i(t)] \leq \frac{B}{2\epsilon^{(\lambda)}}, \tag{10}$$

where $B = \sum_{i=1}^{N}(\lambda_i + \mathrm{E}[A_i^2] - 2\lambda_i^2)$.

Proof: Consider the quadratic Lyapunov function $\Phi : \mathbb{R}_+^N \to \mathbb{R}_+$ defined by $\Phi(Q) = \sum_{i=1}^{N} Q_i^2$. The drift of the Lyapunov function satisfies the following:

$$\begin{aligned}
& \Phi(Q(t+1)) - \Phi(Q(t)) \\
= {} & (Q(t+1) - Q(t))^T(Q(t+1) + Q(t)) \\
= {} & (A(t) - D(t))^T(2Q(t) + A(t) - D(t)) \\
= {} & 2(A(t) - D(t))^T Q(t) + (A(t) - D(t))^T(A(t) - D(t)) \\
= {} & 2(A(t) - D(t))^T Q(t) \\
& + A(t)^T A(t) + D(t)^T D(t) - 2A(t)^T D(t).
\end{aligned} \tag{11}$$

If the network is stable under a given policy, $(Q(t), A(t), D(t))$ forms a positive recurrent DTMC [3] and a stationary distribution exists. Under the stationary distribution, for a lower-bounded k-th-degree polynomial function $h : \mathbb{R}^N \to \mathbb{R}$, we have $\mathrm{E}[h(Q(t+1)) - h(Q(t))] = 0$, if the k-th moments exists [14] [15]. Since Φ satisfies the condition, under the stationary distribution,

$$\begin{aligned}
0 &= \mathrm{E}[\Phi(Q(t+1)) - \Phi(Q(t))] \\
&= 2\,\mathrm{E}[(A(t) - D(t))^T Q(t)] + \mathrm{E}[A(t)^T A(t)] \\
&\quad + \mathrm{E}[D(t)^T D(t)] - 2\,\mathrm{E}[A(t)^T D(t)].
\end{aligned} \tag{12}$$

Next, we bound the terms in (12). First,

$$\mathrm{E}[A(t)^T Q(t)] = \lambda^T \mathrm{E}[Q(t)].$$

Under the MWS policy, for any $\mu \in \Lambda$, we have[2]

$$D(t)^T Q(t) \geq \mu^T Q(t). \tag{13}$$

Suppose $\lambda \in \Lambda$. Let $\mu = \lambda + \epsilon^{(\lambda)}e$. We know that $\mu \in \Lambda$. Thus, we have

$$\begin{aligned}
2\,\mathrm{E}[(A(t) - D(t))^T Q(t)] &\leq 2(\lambda - \mu)^T \mathrm{E}[Q(t)] \\
&= -2\epsilon^{(\lambda)}e^T \mathrm{E}[Q(t)].
\end{aligned}$$

Since $D_i(t) \in \{0, 1\}$ for each i, we have $D_i^2(t) = D_i(t)$. At stationarity, we also have $\mathrm{E}[A(t)] = \mathrm{E}[D(t)]$. Thus,

$$\mathrm{E}[D(t)^T D(t)] = \mathrm{E}[\sum_{i=1}^{N} D_i(t)] = \sum_{i=1}^{N} \lambda_i. \tag{14}$$

Since $A_i(t)$ is independent of $D_i(t)$, we have

$$\mathrm{E}[A(t)^T D(t)] = \mathrm{E}[A(t)]^T \mathrm{E}[D(t)] = \sum_{i=1}^{N} \lambda_i^2. \tag{15}$$

$$0 \leq -2\epsilon^{(\lambda)}e^T \mathrm{E}[Q(t)] + \sum_{i=1}^{N} \lambda_i + \sum_{i=1}^{N} \mathrm{E}[A_i^2] - 2\sum_{i=1}^{N} \lambda_i^2.$$

After rearrangement, we arrive at (10). ∎

By applying Little's law, which says the product of the average arrival rate and the average delay experienced by an

[2] By the definition of the MWS policy, the schedule selected at time t, denoted by $m(t)$, satisfies $m(t)^T Q(t) = \pi(G, Q(t))$. If $Q_i(t) = 0$, then $D_i(t)Q_i(t) = m_i(t)Q_i(t) = 0$. By (1), if $Q_i(t) \geq 1$ and $m_i(t) = 1$, then $D_i(t) = 1$; if $Q_i(t) \geq 1$ and $m_i(t) = 0$, then $D_i(t) = 0$. In all these cases, we have $D_i(t)Q_i(t) = m_i(t)Q_i(t)$. Hence, (13) holds.

arrival is equal to the average queue size, the following bound on the expected delay, $\bar{d}$, is obtained.

Corollary 2. *Under MWS, for $\lambda \in \Lambda^o$,*

$$\bar{d} \leq \frac{B}{2(\sum_{i=1}^N \lambda_i)\epsilon^{(\lambda)}}. \tag{16}$$

B. Challenges in Computing $\epsilon^{(\lambda)}$

Computing $\epsilon^{(\lambda)}$ is NP-hard, which can be shown by a reduction of the fractional coloring problem. Let the scalar $b \geq 0$ be such that $be \in \Lambda^o$. By Corollary 1, $\epsilon^{(be)} = (1 - b\chi_f(G))/\chi_f(G)$. Hence, one can solve for $\epsilon^{(be)}$ if and only if one can solve for $\chi_f(G)$, the fractional coloring number. For an interference graph $G = (V, E)$, the vector $e/|V| \in \Lambda$, and hence, $e/(2|V|) \in \Lambda^o$. One can take $b = 1/(2|V|)$.

The fractional coloring problem (more precisely, its decision version) is NP-complete. Grötschel et al. [16] [17] established an equivalence between the fractional coloring problem and the maximum weight independent set (MWIS) problem. Furthermore, an approximate solver for the MWIS problem can be used to derive an approximation algorithm for the fractional coloring problem. In [18] [19], the approximation ratio of the MWIS solver is directly carried to the latter problem's solution. However, it is also known that the MWIS problem does not admit a polynomial time approximation algorithm with an approximation ratio $|V|^\delta$ for some $\delta > 0$ [20].

It is NP-hard to compute $\chi_f(G, \lambda)$ in general. Also the existence of a polynomial-time algorithm that provides approximation within $\frac{N}{2^{\sqrt{\log N}}}$ implies P= NP [21].

C. Performance Comparison between $\epsilon^{(\lambda)}$ and $\zeta^{(\lambda)}$

Next, we will investigate the gap between $\epsilon^{(\lambda)}$ and $\zeta^{(\lambda)}$. We will show that $\epsilon^{(\lambda)}/\zeta^{(\lambda)}$ may scale linearly in the number of network links.

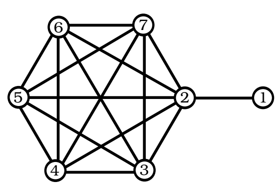

Fig. 1. Illustration of K-clique with an extra node attached; $K = 6$.

Consider an interference graph G that consists of a K-clique, where $K \geq 3$, and an extra node attached to the clique by a single edge. The extra node is indexed as node 1. Fig. 1 illustrates one of such graphs where $K = 6$. Consider a vector $\lambda = (1 - 2/K, 0, \ldots, 0)^T$. Note that only the extra node attached to the clique has a non-zero value in λ. By Lemma 4, $\max_j \lambda_j \leq \chi_f(G, \lambda) \leq e^T\lambda$. Hence, $\chi_f(G, \lambda) = 1 - 2/K$. Also, it is easy to check that $\chi_f(G) = K$. Thus, $\zeta^{(\lambda)} = \frac{1-(1-2/K)}{K} = 2/K^2$.

Let $\mu = \lambda + \frac{1}{K}e = (1 - 1/K, 1/K, \ldots, 1/K)^T$. Observe that the maximal schedules are: $(1, 0, 1, 0, \ldots, 0)^T$, $(1, 0, 0, 1, 0 \ldots, 0)^T$, ..., $(1, 0, 0, \ldots, 0, 1)^T$ together with $(0, 1, 0, 0, \ldots, 0)^T$. It is easy to verify that the vector $\alpha = (1/K, \ldots, 1/K)^T$ is the optimal solution for computing $\chi_f(G, \mu)$, which implies $\chi_f(G, \mu) = 1$, and hence, $\epsilon^{(\lambda)} = 1/K$. Thus, we have $\epsilon^{(\lambda)} = \frac{K}{2}\zeta^{(\lambda)}$, indicating that $\epsilon^{(\lambda)}/\zeta^{(\lambda)}$ may scale linearly with the node count in the interference graph G.

V. Optimality of the New Bounds

In this section, we will show that the new bound for the sum of the expected queue sizes is not only an improvement over the earlier bound, it is also the best bound that is obtainable using the often-used analytical framework of this paper. To reach that conclusion, we will derive some general bounds under the MWS-w policy, and then, specialize the results to the case of the MWS policy.

Throughout this section, we assume that the given arrival rate vector λ satisfies $\lambda \in \Lambda^o$ unless mentioned otherwise.

A. Tightest Single-Parameter Bound under MWS-w Policy

The MWS-w policy is a generalization of the MWS policy in that the weight assignment at each node $i \in V$ is its queue length scaled by a positive constant w_i. More precisely, let $w = (w_i)_i$ be a positive scaling vector. Let the weight of node i at time t be $z_i(t) = w_i Q_i(t)$, and let $z(t) = (z_i(t))_i$ be the weight vector at time t. Under the MWS-w policy, the chosen schedule $m(t)$ at each time slot t satisfies $m(t)^T z(t) \geq m^T z(t)$ for all $m \in M_L$.

Consider a fixed scaling vector $w > 0$. We will first derive a whole family of upper bounds for the sum of the expected queue sizes. They are all derivable by analyzing the drift of a family of Lyapunov functions.

Lemma 10. *Under the MWS-w policy where $w > 0$, for any μ such that $\mu \in \Lambda$ and $\mu > \lambda$, we have*

$$\sum_{i=1}^N E[Q_i(t)] \leq \frac{\sum_{i=1}^N w_i B_i}{\min_i \ w_i(\mu_i - \lambda_i)}, \tag{17}$$

where $B_i = \frac{1}{2}(\lambda_i + \mathrm{E}[A_i^2(t)] - 2\lambda_i^2)$ for each i.

Proof: Given $w > 0$, define a Lyapunov function Φ^w by

$$\Phi^w(Q) = \frac{1}{2}\sum_{i=1}^N w_i Q_i^2. \tag{18}$$

Let the one-step drift at time t be $\Delta(t; w) = \Phi^w(Q(t+1)) - \Phi^w(Q(t))$. We have the following.

$$\begin{aligned}\Delta(t; w) =& \frac{1}{2}\sum_{i=1}^N w_i(Q_i(t+1) - Q_i(t))(Q_i(t+1) + Q_i(t)) \\ =& \sum_{i=1}^N w_i(A_i(t) - D_i(t))Q_i(t) \\ &+ \frac{1}{2}\sum_{i=1}^N w_i(A_i(t) - D_i(t))^2.\end{aligned}$$

By taking the conditional expectation and using the fact that, under the assumed scheduling policy,

$$\sum_{i=1}^N w_i D_i(t) Q_i(t) \geq \sum_{i=1}^N w_i \mu_i Q_i(t), \text{ for all } \mu \in \Lambda, \tag{19}$$

we get

$$E[\Delta(t;w)|Q(t)] \le \sum_{i=1}^{N} w_i(\lambda_i - \mu_i)Q_i(t) + \frac{1}{2}\sum_{i=1}^{N} w_i E[(A_i(t) - D_i(t))^2|Q(t)]. \quad (20)$$

It is easy to check

$$\frac{1}{2}\sum_{i=1}^{N} w_i \,\mathrm{E}[(A_i(t) - D_i(t))^2] = \sum_{i=1}^{N} w_i B_i.$$

Continue from (20). After taking another expectation and by setting $E[\Delta(t;w)] = 0$ under stationarity, we get

$$\sum_{i=1}^{N} w_i(\mu_i - \lambda_i)E[Q_i(t)] \le \sum_{i=1}^{N} w_i B_i. \quad (21)$$

Observe that, for any fixed vectors w and μ, where $w \ge 0$ and $\mu \ge \lambda$, and for any non-negative vector $(E[Q_i(t)])_i$, the following holds.

$$(\min_i w_i(\mu_i - \lambda_i)) \sum_{i=1}^{N} E[Q_i(t)] \le \sum_{i=1}^{N} w_i(\mu_i - \lambda_i)E[Q_i(t)]. \quad (22)$$

Equality is achieved in the case where $E[Q_i(t)] > 0$ only for $i \in \arg\min_i w_i(\mu_i - \lambda_i)$.

Then, under the given $w > 0$, for any μ such that $\mu \in \Lambda$ and $\mu > \lambda$, there is an upper bound for $\sum_{i=1}^{N} E[Q_i(t)]$, as given in (17). ∎

Next, we wish to make the upper bound as tight as possible over $\mu \in \Lambda$ and $\mu > \lambda$. We will argue that it is enough to consider the following optimization problem.

$$\min_{\mu} \frac{\sum_{i=1}^{N} w_i B_i}{\min_i \; w_i(\mu_i - \lambda_i)} \quad (23)$$

$$\text{s.t. } \mu \ge \lambda, \mu \in Co(M_L). \quad (24)$$

In other words, it is enough to consider only those $\mu \ge \lambda$ that satisfies $\mu \in Co(M_L)$. First, suppose the tightest bound is achieved at some $\mu \in \Lambda$ where $\mu > \lambda$. By the definition of Λ, there always exists $\nu \in Co(M_L)$ such that $\nu \ge \mu$. Clearly, the bound achieved under ν is at least as tight as that under μ. Hence, it is enough to consider the constraints defined by $\mu > \lambda$ and $\mu \in Co(M_L)$. Second, recall that $\lambda \in \Lambda^o$, which means that there exists $\mu \in Co(M_L)$ such that $\lambda < \mu$. Hence, the problem in (23)-(24) is feasible. Furthermore, any optimal solution to the problem must occur at some $\mu > \lambda$. Otherwise, the objective value will be infinitely large, violating the feasibility of the problem. Therefore, it makes no difference whether to write the constraints as $\mu \ge \lambda$ or $\mu > \lambda$.

Definition 2. *For any $\lambda \in \Lambda$, $w \in \mathbb{R}_+^N$ and $w > 0$, define $\epsilon^{(\lambda,w)}$ to be the unique value satisfying $\chi_f(G, \lambda + \epsilon^{(\lambda,w)}w) =$ 1.*

Note that $\epsilon^{(\lambda)}$ is a special case: $\epsilon^{(\lambda)} = \epsilon^{(\lambda,e)}$. As in Lemma 9, $\epsilon^{(\lambda,w)}$ has an alternative definition through a linear program.

$$\epsilon^{(\lambda,w)} \triangleq \max_{\alpha,\epsilon} \epsilon \quad \text{s.t. } M_L\alpha \ge \lambda + \epsilon w, \quad e^T\alpha = 1, \alpha \ge 0.$$

We can now generalize Theorem 1.

Theorem 2. *Under the MWS-w policy, where $w > 0$, for any arrival rate vector $\lambda \in \Lambda^o$, the following holds under the stationary distribution.*

$$\mathrm{E}[\sum_{i=1}^{N} Q_i(t)] \le \frac{\sum_{i=1}^{N} w_i B_i}{\epsilon^{(\lambda,w^{-1})}}, \quad (25)$$

where w^{-1} denotes the vector $(1/w_1, \ldots, 1/w_N)^T$.

Proof: Start with the problem in (23)-(24). Since the constant factor $\sum_{i=1}^{N} w_i B_i$ does not affect the optimal solution, we will omit it for notational simplicity and re-write the problem as

$$\max_{\mu} \; \min_i w_i(\mu_i - \lambda_i) \quad \text{s.t. } \mu \ge \lambda, \mu \in Co(M_L).$$

Since the optimization problem above is satisfied by a vector $\mu \in Co(M_L)$, we can re-write the constraints accordingly and get the following equivalent problem.

$$\max_{\mu,\alpha} \; \min_i w_i(\mu_i - \lambda_i) \quad \text{s.t. } \mu \ge \lambda, M_L\alpha = \mu, \quad e^T\alpha = 1, \alpha \ge 0.$$

Let $\epsilon = \min_i w_i(\mu_i - \lambda_i)$, which implies $\mu_j - \lambda_j \ge \epsilon/w_j$ for all j. Then, the optimization problem can be rewritten as

$$\max_{\alpha,\epsilon} \epsilon \quad \text{s.t. } M_L\alpha \ge \lambda + \epsilon w^{-1}, \quad e^T\alpha = 1, \alpha \ge 0.$$

From the linear programming definition of $\epsilon^{(\lambda,w)}$, we see that the optimal objective value the above problem is equal to $\epsilon^{(\lambda,w^{-1})}$. Bringing back the constant factor $\sum_{i=1}^{N} w_i B_i$, we arrive at the conclusion of the theorem. ∎

Corollary 3. *For any $w \in \mathbb{R}_+^N$ with $w > 0$ and for any arrival rate vector $\lambda \in \Lambda^o$, the following holds.*

$$\frac{\sum_{i=1}^{N} w_i B_i}{\epsilon^{(\lambda,w^{-1})}} \le \frac{\sum_{i=1}^{N} w_i B_i}{\min_i \; w_i(\mu_i - \lambda_i)}, \textit{ for all } \mu \in \Lambda, \mu > \lambda.$$

Proof: The proof for Theorem 2 shows that, for all $\mu \in Co(M_L)$ and $\mu \ge \lambda$,

$$\frac{\sum_{i=1}^{N} w_i B_i}{\epsilon^{(\lambda,w^{-1})}} \le \frac{\sum_{i=1}^{N} w_i B_i}{\min_i \; w_i(\mu_i - \lambda_i)}.$$

The comments after (23)-(24) show that the above inequality holds for all $\mu \in \Lambda$ and $\mu > \lambda$. ∎

Remark. Lemma 10, Theorem 2 and Corollary 3 together demonstrate that, under the MWS-w policy, if we wish to find the tightest bound for the queue sum, we only need to look for a vector $\tilde{\mu} \in \Lambda$ and $\tilde{\mu} > \lambda$ in the direction w^{-1}, starting from the vector $\lambda \in \Lambda^o$. When the ray $\{\lambda + \epsilon w^{-1} : \epsilon \geq 0\}$ hits the boundary of Λ, we have $\epsilon = \epsilon^{(\lambda, w^{-1})}$ and the resulting vector $\tilde{\mu} = \lambda + \epsilon^{(\lambda, w^{-1})} w^{-1}$ provides the tightest bound.

Note that $||\tilde{\mu} - \lambda|| = \epsilon^{(\lambda, w^{-1})} ||w^{-1}||$, where the norm is the Euclidean norm. Hence, the bound in (25) is inversely proportional to the distance from λ to the boundary of Λ along the direction w^{-1}. Also note that $w_i = \epsilon^{(\lambda, w^{-1})} / (\tilde{\mu}_i - \lambda_i)$ for each i.

By letting $w = e$, we arrive at the conclusion that under MWS, the bound in Theorem 1, $\frac{B}{2\epsilon^{(\lambda)}}$, is not only an improvement, but also the tightest with our analytical framework. Since this is a key result of the paper, we state it formally.

Corollary 4. *Under the MWS policy, for any arrival rate vector $\lambda \in \Lambda^o$, $\frac{B}{2\epsilon^{(\lambda)}}$ is the tightest bound for the sum of the expected queue sizes under the stationary distribution (obtainable through the analytical framework considered here).*

The analysis conducted above establishes the significance of $\epsilon^{(\lambda)}$. We must emphasize that $\frac{B}{2\epsilon^{(\lambda)}}$ is the tightest bound obtainable through the analytical framework used here, which can be summarized as follows.

- A family of bounds are obtained by analyzing the drift of the Lyapunov function Φ^w given in (18).
- The best bound is found by minimization over the family.

B. Tightest Multi-Parameter Bound over All MWS-w Policies

In this section, we will show that the multi-parameter bound developed in [2] is the tightest within our analytical framework.

The authors of [2] provided the following important bound. Define B_o by

$$B_o \triangleq \min_{\mu} \sum_{i=1}^{N} \frac{B_i}{\mu_i - \lambda_i}, \text{subject to } \mu > \lambda, \mu \in \Lambda, \qquad (26)$$

where each B_i is the same as given in Lemma 10. Let μ^* be an optimal solution to the problem in (26) and let the vector w^* be such that $w_i^* = 1/(\mu_i^* - \lambda_i)$ for each i. Then, under the particular MWS-w^* policy, $\mathrm{E}[\sum_{i=1}^{N} Q_i(t)] \leq B_o$. The bound B_o requires multiple parameters, μ_i^*, for $i = 1, \ldots, N$. Hence, we call it a *multi-parameter bound*. The bound B_o is closely connected to the results and method of this paper, and we now explore the connection.

In the following, we will address why w^* is chosen to be $w_i^* = 1/(\mu_i^* - \lambda_i)$ for each i, or whether a better policy with a better bound exists. Specifically, one may wonder whether there exists some $\tilde{w} > 0$ such that the MWS-$\tilde{w}$ policy has a tighter bound. We now provide the answers.

Consider a fixed $\lambda \in \Lambda^o$. Recall that by Theorem 2, for any fixed $w > 0$, the tightest bound for the queue sum under the MWS-w policy is $\sum_{i=1}^{N} w_i B_i / \epsilon^{(\lambda, w^{-1})}$. The following theorem says that B_o is the tightest bound among all the bounds corresponding to different MWS-w policies.

Theorem 3. *For any arrival rate vector $\lambda \in \Lambda^o$,*

$$B_o \leq \sum_{i=1}^{N} w_i B_i / \epsilon^{(\lambda, w^{-1})}, \quad \textit{for all } w > 0.$$

Proof: We start by trying to make the upper bound in Lemma 10 as tight as possible over all $w > 0$, $\mu \in \Lambda$ and $\mu > \lambda$. Using a similar argument given after (23)-(24), it is enough to consider the following optimization problem.

$$\min_{\mu, w} \frac{\sum_{i=1}^{N} w_i B_i}{\min_i \ w_i(\mu_i - \lambda_i)} \qquad (27)$$

$$\text{s.t. } \mu > \lambda, \mu \in Co(M_L), w > 0. \qquad (28)$$

The problem in (27)-(28) can be solved by first minimizing over w and then minimizing over μ. For the minimization over w under a fixed μ, the objective function value, $\frac{\sum_{i=1}^{N} w_i B_i}{\min_i \ w_i(\mu_i - \lambda_i)}$, is invariant with respect to linear scaling of w. Hence, we can set $\min_i \ w_i(\mu_i - \lambda_i) = 1$. Thus, for a fixed $\mu > \lambda$, we need to solve the following.

$$\min_{w} \sum_{i=1}^{N} w_i B_i$$
$$\text{s.t. } w > 0, \ \min_i w_i(\mu_i - \lambda_i) = 1.$$

Since each B_i is non-negative, an optimal solution is $w_i = 1/(\mu_i - \lambda_i)$ for every i. Then, the minimization over μ becomes[3]

$$B_o = \min_{\mu} \sum_{i=1}^{N} \frac{B_i}{\mu_i - \lambda_i}, \text{subject to } \mu > \lambda, \mu \in Co(M_L).$$

■

To summarize, what we have shown here is that B_o is the tightest multi-parameter bound over all MWS-w policy. That is, no MWS-w policy can do better than MWS-w^* in terms of the obtainable bound. The decision of setting $w_i^* = 1/(\mu_i^* - \lambda_i)$ for each i is optimal.

C. Single vs. Multi-Parameter Bounds

The multi-parameter bound $B_o = \sum_{i=1}^{N} \frac{B_i}{\mu_i^* - \lambda_i}$ is very interesting. As pointed out in [2], B_o is a sum of terms corresponding to individual queues. Also, since $\mu^* \in \Lambda$, it can be viewed as a service rate vector. Consider a system of N independent single-server queues with an arrival rate λ_i and service μ_i^* for each queue i. Under fairly general conditions, the expected size of each queue i is proportional to $\frac{1}{\mu_i^* - \lambda_i}$. For our system, one may conjecture whether each queue i is individually bounded by $\frac{B_i}{\mu_i^* - \lambda_i}$, whether the bound B_o is indeed equal to the expected queue sum under a wide class of arrival processes, or more strongly, whether each term $\frac{B_i}{\mu_i^* - \lambda_i}$ is the expected queue size of queue i.

Note that the only relaxation in the steps of deriving the bound are inequalities (19) and (22). In the case of $w_i^* = 1/(\mu_i^* - \lambda_i)$ for each i, (22) is in fact equality. A key question

[3] As before, the constraint $\mu > \lambda$ can be replaced by $\mu \geq \lambda$ and $\mu \in Co(M_L)$ can be replaced by $\mu \in \Lambda$.

is therefore whether or when the following holds under MWS-w^*.

$$E[\sum_{i=1}^{N} \frac{D_i(t)Q_i(t)}{\mu_i^* - \lambda_i}] = E[\sum_{i=1}^{N} \frac{\mu_i^* Q_i(t)}{\mu_i^* - \lambda_i}]. \tag{29}$$

The question points to an interesting line of future inquiries.

On the other hand, consider the single-parameter bound in (25) for a fixed $w > 0$. The value of the bound can also be derived by solving the problem in (23)–(24). Let $\tilde{\mu}$ be an optimal solution, which depends on the fixed w. Then, the bound is $\frac{\sum_{i=1}^{N} w_i B_i}{\min_i w_i(\tilde{\mu}_i - \lambda_i)}$. Although the bound can also be interpreted as a sum of individual terms, $\frac{B_j}{\min_i w_i(\tilde{\mu}_i - \lambda_i)/w_j}$, one for each queue j, it is unlikely that every term is a good bound for the corresponding queue. The reason is that, in general, the denominator $\min_i w_i(\tilde{\mu}_i - \lambda_i)/w_j$ is not equal to a service rate minus the arrival rate for queue j, unless $j \in \arg\min_i w_i(\tilde{\mu}_i - \lambda_i)$, or $w_j = 1/(\tilde{\mu}_j - \lambda_j)$ for all j.

The distinction between the single or multi-parameter bounds has to do with whether or not some measure of the traffic intensity to the queueing system is specified by a single parameter or multiple parameters. In the single-parameter case, $1 - \epsilon^{(\lambda)}$ can be viewed as taking over the role of the traffic intensity (i.e., load). In the multi-parameter case, the traffic intensity is not described by a single value but by a set of values, λ_i/μ_i^*, for $i = 1, \ldots, N$. Finally, the multi-parameter bound applies only to the MWS-w^* policy whereas there is a single-parameter bound for every MWS-w policy with a fixed w.

VI. Additional Results

In this section, we investigate the class of graphs for which the new and old single-parameter bounds are always identical. We denote this class by Ω. One of the goals is to further explore the geometric and graph-theoretical aspects of the wireless scheduling problem.

Definition 3. *Ω is the set of all graphs G such that $\epsilon^{(\lambda)} = \zeta^{(\lambda)}$ for all $\lambda \in \Lambda(G)$*[4].

Lemma 11. *The following are equivalent:*
(i) $G \in \Omega$;
(ii) $\chi_f(G, \lambda) + \chi_f(G, \epsilon^{(\lambda)} e) = 1$ for all $\lambda \in \Lambda(G)$[5] *;*
(iii) $\chi_f(G, \lambda + be) = \chi_f(G, \lambda) + \chi_f(G, be)$ for all $\lambda \in \Lambda(G)$ and all $b \in \mathbb{R}_+$ such that $\lambda + be \in \Lambda(G)$.

Proof: (i) $\Leftrightarrow$ (ii): By the definition of $\zeta^{(\lambda)}$, $\epsilon^{(\lambda)} = \zeta^{(\lambda)}$ means that $1 - \chi_f(G, \lambda) = \epsilon^{(\lambda)} \chi_f(G) = \chi_f(G, \epsilon^{(\lambda)} e)$. That is, $\chi_f(G, \lambda) + \chi_f(G, \epsilon^{(\lambda)} e) = 1$.

(ii) $\Rightarrow$ (iii): Given an arbitrary $\lambda \in \Lambda$, consider an arbitrary $b \geq 0$ such that $\lambda + be \in \Lambda$. By Lemma 7, $\epsilon^{(\lambda)} = b + \epsilon^{(\lambda+be)}$. Then,

$$\begin{aligned} \chi_f(G, \epsilon^{(\lambda)} e) &= \chi_f(G, (b + \epsilon^{(\lambda+be)})e) \\ &= \chi_f(G, be) + \chi_f(G, \epsilon^{(\lambda+be)} e). \end{aligned}$$

Hence, $\chi_f(G, \lambda) + \chi_f(G, \epsilon^{(\lambda)} e) = 1$ can be written as

$$\chi_f(G, \lambda) + \chi_f(G, be) + \chi_f(G, \epsilon^{(\lambda+be)} e) = 1.$$

Also, since $\lambda + be \in \Lambda$,

$$\chi_f(G, \lambda + be) + \chi_f(G, \epsilon^{(\lambda+be)} e) = 1.$$

From the last two equalities, we get

$$\chi_f(G, \lambda + be) = \chi_f(G, \lambda) + \chi_f(G, be).$$

That is, the triangle inequality holds with equality for any $\lambda \in \Lambda$ and any $b \geq 0$ such that $\lambda + be \in \Lambda$.

(iii) $\Rightarrow$ (ii): Take $b = \epsilon^{(\lambda)}$. ■

Recall that $\chi_f(G)$ is defined by the following optimization problem.

$$\chi_f(G) \triangleq \min \ e^T \alpha, \text{ subject to } M_L \alpha \geq e, \ \alpha \geq 0. \tag{30}$$

Lemma 12. *A graph G is in Ω only if every optimal solution α to the problem in (30) satisfies $M_L \alpha = e$.*

Proof: Suppose a graph G is in Ω. Then, for all $\lambda \in \Lambda(G)$, $\epsilon^{(\lambda)} = \zeta^{(\lambda)}$. Let α be an optimal solution to the optimization problem in (30). Then, $\sum_j \alpha_j = \chi_f(G)$ and $(M_L \alpha)_j \geq 1$ for all j. Assume there exists an index i such that $(M_L \alpha)_i > 1$. Let $\delta_i = ((M_L \alpha)_i - 1)/\chi_f(G)$. Now define a vector $\lambda = (\lambda_i)$ by $\lambda_i = \delta_i$ and $\lambda_j = 0$ for $j \neq i$ and consider $\chi_f(G, \lambda + \frac{1}{\chi_f(G)} e)$. Note that $(M_L \alpha)_j / \chi_f(G) \geq \lambda_j + 1/\chi_f(G)$ for all j. Thus, $\chi_f(G, \lambda + \frac{1}{\chi_f(G)} e) \leq \sum_j \alpha_j / \chi_f(G) = 1$, which implies $\lambda + \frac{1}{\chi_f(G)} e \in \Lambda$ by Lemma 1. By Lemma 6, $\epsilon^{(\lambda + \frac{1}{\chi_f(G)} e)} \geq 0$.

Since $\chi_f(G) = e^T \alpha \geq \max_i (M_L \alpha)_i$, we have $\delta_i < 1$. Thus, $\lambda \leq (1, 0, \ldots, 0)^T$. Since the latter vector is in Λ, we must have $\lambda \in \Lambda$.

By Lemma 7,

$$\begin{aligned} \epsilon^{(\lambda + \frac{1}{\chi_f(G)} e)} &= \epsilon^{(\lambda + \frac{1}{\chi_f(G)} e - \frac{1}{\chi_f(G)} e)} - \frac{1}{\chi_f(G)} \\ &= \epsilon^{(\lambda)} - \frac{1}{\chi_f(G)} \geq 0. \end{aligned}$$

Thus, $\epsilon^{(\lambda)} \geq 1/\chi_f(G)$.

By Lemma 4, $\delta_i = \max_j \lambda_j \leq \chi_f(G, \lambda) \leq e^T \lambda = \delta_i$. That is, $\chi_f(G, \lambda) = \delta_i$. Therefore,

$$\zeta^{(\lambda)} = \frac{1 - \chi_f(G, \lambda)}{\chi_f(G)} = \frac{1 - \delta_i}{\chi_f(G)} < \frac{1}{\chi_f(G)} \leq \epsilon^{(\lambda)}, \tag{31}$$

which is a contradiction to the assumption that $\zeta^{(\lambda)} = \epsilon^{(\lambda)}$ for all λ.[6] ■

Remark. It is unknown whether Lemma 12 can be made into an "if and only if" statement.

We next show that vertex-transitive graphs are potential members of Ω. First, we need to introduce some definitions (see [22] for a more details). In the following, we consider undirected graphs with no loops and no more than one edge between any two different nodes, i.e., the *simple graphs*.

[4] $\Lambda(G)$ and $\Lambda^o(G)$ denote the capacity region and its interior, respectively, under the interference graph G.

[5] Since $\chi_f(G, \lambda + \epsilon^{(\lambda)} e) = 1$, we see that $\epsilon^{(\lambda)} = \zeta^{(\lambda)}$ is equivalent to $\chi_f(G, \lambda + \epsilon^{(\lambda)} e) = \chi_f(G, \lambda) + \chi_f(G, \epsilon^{(\lambda)} e)$. That is, the triangle inequality holds as an equality.

[6] Alternatively, we can apply Lemma 11 (iii). $\chi_f(G, \lambda + e/\chi_f(G)) = \chi_f(G, \lambda) + \chi_f(G, e/\chi_f(G)) = \delta_i + 1$, which is a contradiction to an earlier statement that $\chi_f(G, \lambda + e/\chi_f(G)) \leq 1$.

Definition 4. *An isomorphism from a graph $G = (V_G, E_G)$ to a graph $H = (V_H, E_H)$ is a bijection $f : V_G \to V_H$ such that $(u, v) \in E_G$ if and only if $(f(u), f(v)) \in E_H$.*

Definition 5. *An automorphism of a graph G is an isomorphism from G to G.*

Definition 6. *A graph $G = (V, E)$ is vertex-transitive if for every pair $u, v \in V$ there is an automorphism that maps u to v.*

Cyclic graphs are vertex-transitive, e.g., a cycle with 5 nodes, which is denoted as C_5.

Lemma 13. *For any vertex-transitive graph G, every optimal solution α to the problem in (30) satisfies $M_L\alpha = e$.*

Proof: Consider the dual problem of (30):

$$w_f(G, e) = \max e^T\beta, \text{ subject to } M_L^T\beta \leq e, \beta \geq 0. \quad (32)$$

For a vertex-transitive graph G, it is known that the problem in (32) admits an optimal solution where $\beta_i = 1/\pi(G)$ for $i = 1, \ldots, N$ [23]. Here, $\pi(G)$ is the size of a maximum independent set of the graph. Since $\beta_i > 0$ for all i, by complementary slackness, we have at least one optimal solution α^* to the primal problem in (30) where $M_L\alpha^* = e$.

Suppose there exists another optimal solution to (30), α^i, for which $(M_L\alpha^i)_i > 1$. Now consider another node $j \in V$, where $j \neq i$. By the definition of vertex-transitivity, there is an automorphism f that maps node i to node j and, for that mapping, there is a new optimal solution vector α^{ij} for which $(M_L\alpha^{ij})_j > 1$. Since j is arbitrary except that $j \neq i$, we can get a collection of optimal solutions, $\{\alpha^{ik}\}_{k\neq i}$, such that for each $k \neq i$, $(M_L\alpha^{ik})_k > 1$. For convenience, define $\alpha^{ik} = \alpha^i$.

Since each α^{ik} is optimal, we have $\sum_{i=1}^N e^T\alpha^{ik}/N = \chi_f(G)$. Note that $M_L\sum_{k=1}^N \alpha^{ik}/N > e$. Thus, the vector $\sum_{i=1}^N \alpha^{ik}/N$ can be scaled by some scalar b, $0 < b < 1$, while still satisfying the constraints in (30). Now, $b\sum_{i=1}^N e^T\alpha^{ik}/N < \chi_f(G)$, which contradicts the assumption that $\chi_f(G)$ is the optimal value of the problem in (30). ■

VII. Conclusions

In this paper, our primary objective is to provide the tightest, single-parameter upper bound for the expected delay (or the sum of the expected queue sizes) under the MWS policy. In process, we have done the same for the entire family of the MWS-w policies for all $w > 0$. Although there are earlier attempts to bound the expected delay (or queue sum) such as in [1], [2], [4], our emphasis is to collect the common analytical framework in those earlier studies and find the best bounds that the framework can provide. Using the weighted fractional coloring number $\chi_f(G, \lambda)$, we define the parameter $\epsilon^{(\lambda)}$ (or, more generally, $\epsilon^{(\lambda, w^{-1})}$) and show how the best delay or queue bound depends on it. We also describe its geometric meaning and various properties. We evaluate the improvement of the new bound over the old one in a class of interference graphs and show that the new bound can be superior by a factor that increases linearly in the number of nodes.

The method and results of this paper have close connections with an important multi-parameter bound in [2], which corresponds to the MWS-w policy for a specifically chosen w. We describe those connections and show that the chosen w is optimal in the sense that no other MWS-w policy has a smaller bound.

References

[1] K. Kar, X. Luo, and S. Sarkar, "Delay guarantees for throughput-optimal wireless link scheduling," in *Proc. of IEEE INFOCOM*, 2009, pp. 2331–2339.

[2] G. R. Gupta and N. B. Shroff, "Delay analysis for wireless networks with single hop traffic and general interference constraints," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 393–405, April 2010.

[3] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[4] C. Boyaci and Y. Xia, "Delay analysis of the approximate maximum weight scheduling in wireless networks," in *Proceedings of the 9th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC)*, July 2013.

[5] C. Boyaci, B. Li, and Y. Xia, "An investigation on the nature of wireless scheduling," in *Proc. of IEEE INFOCOM*, March 2010.

[6] D. Shah, D. N. C. Tse, and J. N. Tsitsiklis, "Hardness of low delay network scheduling," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7810–7817, December 2011.

[7] L. Bui, R. Srikant, and A. Stolyar, "Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing," *ArXiv e-prints, 0901.1312v3*, May 2010. [Online]. Available: http://arxiv.org/abs/0901.1312v3

[8] M. Lotfinezhad and P. Marbach, "Throughput-optimal random access with order-optimal delay," *ArXiv e-prints, 1009.5944*, September 2010. [Online]. Available: http://arxiv.org/abs/1009.5944

[9] D. Shah and J. Shin, "Delay optimal queue-based CSMA," in *Proceedings of the ACM SIGMETRICS*, New York, NY, USA, 2010.

[10] I. Keslassy and N. McKeown, "Analysis of scheduling algorithms that provide 100% throughput in input-queued switches," in *Proc. of Allerton Conference on Communication, Control and Computing*, 2001.

[11] D. Shah and M. Kopikare, "Delay bounds for approximate maximum weight matching algorithms for input queued switches," in *Proc. of IEEE INFOCOM*, 2002, pp. 1024–1031.

[12] A. L. Stolyar, "Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 1–516, 2004.

[13] A. Eryilmaz and R. Srikant, "Asymptotically tight steady-state queue length bounds implied by drift conditions," *Queueing Systems: Theory and Applications*, vol. 72, no. 3-4, pp. 311–359, Dec. 2012.

[14] P. Kumar and S. Meyn, "Stability of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. 40, no. 2, pp. 251 –260, February 1995.

[15] E. Leonardi, M. Mellia, F. Neri, and M. Ajmone Marsan, "Bounds on average delays and queue size averages and variances in input-queued cell-based switches," in *Proc. of IEEE INFOCOM 2001*, vol. 2, 2001, pp. 1095–1103.

[16] M. Grötschel, L. Lovász, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica*, vol. 1, pp. 169–197, 1981.

[17] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, 2nd ed., ser. Algorithms and Combinatorics. Springer, 1993, vol. 2.

[18] K. Jansen, "Approximate strong separation with application in fractional graph coloring and preemptive scheduling," *Theor. Comput. Sci.*, vol. 302, no. 1-3, pp. 239–256, 2003.

[19] K. Jansen and L. Porkolab, "On preemptive resource constrained scheduling: Polynomial-time approximation schemes," *SIAM J. Discret. Math.*, vol. 20, no. 3, pp. 545–563, 2006.

[20] C. Lund and M. Yannakakis, "On the hardness of approximating minimization problems," *J. ACM*, vol. 41, no. 5, pp. 960–981, 1994.

[21] L. Trevisan, "Non-approximability results for optimization problems on bounded degree instances," in *Proceedings of the thirty-third annual ACM symposium on Theory of computing (STOC)*, 2001, pp. 453–461.

[22] D. B. West, *Introduction to Graph Theory*, 2nd ed. Prentice Hall, August 2000.

[23] C. Godsil and G. Royle, *Algebraic Graph Theory*. Springer, April 2001.