On the Performance of Largest-Deficit-First for Scheduling Real-Time Traffic in Wireless Networks

Xiaohan Kang*, Weina Wang*, Juan José Jaramillo† and Lei Ying*

*School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85287, USA
{xkang6, wwang136, lying6}@asu.edu

†Department of Applied Math and Engineering
Universidad EAFIT
Medellín, Colombia
jjaram93@eafit.edu.co

ABSTRACT

This paper considers the problem of scheduling real-time traffic in wireless networks. We consider an ad hoc wireless network with general interference and general one-hop traffic. Each packet is associated with a deadline and will be dropped if it is not transmitted before the deadline expires. The number of packet arrivals in each time slot and the length of a deadline are both stochastic and follow certain distributions. We only allow a fraction of packets to be dropped. At each link, we assume the link keeps track of the difference between the minimum number of packets that need to be delivered and the number of packets that are actually delivered, which we call deficit. The largest-deficit-first (LDF) policy schedules links in descending order according to their deficit values, which is a variation of the largest-queue-first (LQF) policy for non-real-time traffic. We prove that the efficiency ratio of LDF can be lower bounded by a quantity that we call the real-time local-pooling factor (R-LPF). We further prove that given a network with interference degree β, the R-LPF is at least $1/(\beta + 1)$, which in the case of the one-hop interference model translates into an R-LPF of at least 1/3.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*wireless communication*

General Terms

Theory, performance

Keywords

Stability; real-time scheduling; largest-deficit-first; local-pooling factor; fluid limit

MobiHoc'13, July 29–August 1, 2013, Bangalore, India.

ACM 978-1-4503-2193-8/13/07 ...$15.00.

1. INTRODUCTION

With the increasing number of real-time applications in wireless networks, scheduling traffic of packets with hard deadlines has become a very important problem. However, the problem is very challenging due to the stochastic nature of the traffic arrivals and deadlines. Hou et al. first proposed a frame-based analytical framework for studying scheduling real-time traffic in wireless networks [6]. In the frame-based framework it is assumed that each frame is a number of consecutive time slots, and all packets arrive at the beginning of a frame and have to be scheduled before the end of the frame. They also characterized the real-time capacity region and developed the optimal scheduling algorithm for collocated networks. Later, the frame-based framework has been generalized to networks with heterogeneous delays, fading, congestion control, etc. [7, 8, 9, 10, 11] In particular, Jaramillo et al. extended the idea to general arrival/deadline patterns within a frame and general network topology, and found the optimal scheduling policy [11], where they assumed that packets can arrive at any time slot during a frame, and the deadline of a packet can be any time after its arrival and before the end of the frame. Their paper assumes that the arrival and deadline information is available at the beginning of the frame, so future knowledge is assumed. Furthermore, the computational complexity of the optimal algorithm is prohibitively high except for some special cases such as collocated networks.

In this paper, we consider the case of general real-time traffic patterns without the assumption of frames and with a general interference model. Under the general settings, the stability region is difficult to characterize, and the optimal policy is unknown. In this paper, we are interested in the performance of a low-complexity greedy policy called the largest-deficit-first (LDF) policy [6], which is the real-time variation of the longest-queue-first (LQF) policy that iteratively selects the link with the largest deficit that does not interfere with those links that are already selected. It has been shown that the largest-deficit-first policy is optimal for scheduling real-time traffic in collocated networks [6, 11] under the frame-based model. The performance of the LDF in general networks has not been studied.

Since LDF can be directly applied to networks with general, non-frame-based real-time traffic, we are interested in characterizing the performance of LDF. Although the capacity region and optimal scheduling algorithm for networks with general real-time traffic remain unknown, we are able to establish the efficiency ratio of LDF by connecting it to the frame-based optimal scheduling algorithm, and obtain a lower bound on the efficiency ratio in terms of a new quantity, called the real-time local-pooling factor (R-LPF). The R-LPF extends the idea of the local-pooling factor for non-real-time traffic [12] and its extension for fading channels [18].

We show using the fluid limit technique [3] that this R-LPF can be successfully used to provide a minimum performance guarantee of LDF under real-time traffic. More interestingly, we are able to connect the R-LPF with the interference degree, and prove that the R-LPF is bounded by $1/(\beta+1)$, where β is the interference degree [2]. Our contributions are therefore twofold:

1. We formulate the construction of the R-LPF and prove that it is a lower bound of the efficiency ratio of LDF in the presence of general deadline constraints.
2. We show that in a network with interference degree β, the R-LPF is at least $1/(\beta+1)$. In particular, the R-LPF is at least 1/3 in a network with one-hop interference model.

We would like to emphasize again that for general (non-frame-based) real-time traffic, to the best of our knowledge, there are no known theoretical results on any scheduling policy, which makes the lower bound obtained in this paper a novel contribution.

2. MODEL

In this paper, we consider a wireless network consisting of K links. The set of links is denoted by $\mathcal{K}$. Assume time is slotted, and at each time slot one packet can be successfully transmitted over a link if no interfering links are transmitting at the same time. We remark that the constant service rate assumption has been widely used in the literature, e.g., [5, 13]. We consider a general interference model. We call a set of links $\mathcal{Z} \subseteq \mathcal{K}$ a maximal schedule if links in $\mathcal{Z}$ can be scheduled at the same time without interfering with each other, but no other link can be further scheduled without interfering with links in $\mathcal{Z}$. We assume that there are R possible maximal schedules and the set of maximal schedules is represented by a maximal schedule matrix M, which is a K-by-R matrix with binary entries such that each column represents a distinct maximal schedule and the set of links that are included in this schedule have value 1 in that column. For example, let M_r be the r^{th} column of matrix M, then the set of links $\{l \in \mathcal{K} \colon M_{r,l} = 1\}$ is a maximal schedule, where $M_{r,l}$ is the (r,l) entry of the matrix. By abuse of notation we also let $M = \{M_1, M_2, \ldots, M_R\}$. It is easy to see that any subset of a maximal schedule is a feasible schedule (i.e., all links in that set can be scheduled at the same time).

We consider single hop traffic with deadline constraints. Let $a_l(t)$ denote the number of packets that arrive at the beginning of time slot t at link l, where we assume that all packets have the same size and can be transmitted in a single time slot. We assume that $a(\cdot)$ is a stochastic process that is temporally independent and identically distributed (i.i.d.) and independent across links, with probability mass function (p.m.f.) $(f_l(i) \colon i = 0, 1, 2, \ldots)$. We also assume that $f_l(i) = 0$ for $i > N$; i.e., the number of packets arriving on a link at each time slot is at most N. Denote by α_l the rate of arrivals on link l; i.e., $\alpha_l = \mathbb{E}[a_l(t)] = \sum_{i=1}^{N} i f_l(i)$.

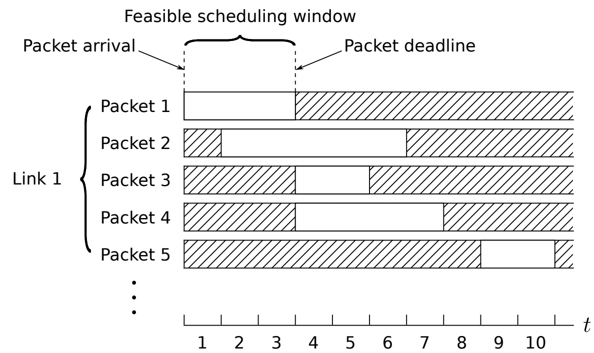


Figure 1: An example of the arrival and maximum delay pattern of packets on a link. For each packet, the beginning of the blank bar is the time slot when that packet arrives, and the end of the blank bar is the deadline associated with that packet. So the feasible scheduling window denoted by the blank bar represents the time slots when the packet is available for transmission, while the shaded part indicates that the packet is not available, either because it has not arrived or because its deadline has passed. Note that here the cumulative numbers of packet arrivals to link 1 are $A_1(\cdot) = (1, 2, 2, 4, 4, 4, 4, 4, 5, 5, \ldots)$ and the maximum delays are $\tau_1(\cdot) = (3, 5, 2, 4, 2, \ldots)$.

Each packet is associated with a maximum delay τ, which is a random variable with integer value between $\tau_{\min}$ and $\tau_{\max}$ and follows a p.m.f. $(\gamma(\tau) \colon \tau_{\min} \le \tau \le \tau_{\max})$. Furthermore, let $A_l(t)$ be the cumulative number of packet arrivals to link l up to time slot t for any $l \in \mathcal{K}$ and any nonnegative integer t; i.e., $A_l(t) = \sum_{t'=1}^{t} a_l(t')$. We order the packets arriving on link l according to the arriving time with arbitrary tie-breakings Then we let $b_l(n)$ be the time slot during which the n^{th} packet arrives on link l; i.e., $b_l(n) = \min\{t \colon A_l(t) \ge n\}$. We also let $e_l(n)$ be the deadline of the n^{th} packet on link l. Note that $e_l(n) = b_l(n) + \tau_l(n)$, where $\tau_l(n)$ is the maximum delay associated with the n^{th} packet on link l. Then the n^{th} arriving packet on link l will be immediately dropped if the deadline is missed. Note that $A(\cdot)$, $\tau(\cdot)$, $b(\cdot)$ and $e(\cdot)$ are all stochastic processes, and $A(\cdot)$ and $\tau(\cdot)$ determine $b(\cdot)$ and $e(\cdot)$. Denote the space of sample paths of the cumulative arrival process $(A_l(\cdot) \colon l \in \mathcal{K})$ and the maximum delay process $(\tau_l(\cdot) \colon l \in \mathcal{K})$ by $\mathcal{A}$. An example of a sample path of the arrival and maximum delay processes on a link during the first 10 time slots is shown in Figure 1.

We assume that each link l is associated with a minimum delivery rate p_l, which is the minimum fraction of packets that should be delivered on link l. The goal of a scheduling policy is to keep the long term delivery rate on link l at least p_l for all $l \in \mathcal{K}$.

Now consider a scheduling policy μ. Denote by $S^{\mu}(t)$ the cumulative service up to time t, in which $S_l^{\mu}(t)$ is the service link l received up to time slot t. For any scheduling policy, it is easy to see that the following three conditions hold:

1. (Initialization) $S_l^{\mu}(0) = 0$ for all $l \in \mathcal{K}$.
2. (Feasibility) The incremental service vector is a feasible schedule; i.e., $0 \preceq S^{\mu}(t) - S^{\mu}(t-1) \preceq M_r$ for some $M_r \in M$, for any positive integer t, where $\preceq$ denotes entrywise less than or equal to.

3. (Deadline constraint) All served packets are served before their deadlines. Formally, let $\zeta_l^\mu(n)$ be the time slot in which the n^{th} packet on link l is scheduled by μ if that packet is ever scheduled, and $\zeta_l^\mu(n) = 0$ if that packet is never scheduled by μ. Then the deadline constraint can be stated as follows: For any n and any l with $\zeta_l^\mu(n) > 0$,

$$b_l(n) \leq \zeta_l^\mu(n) \leq b_l(n) + \tau_l(n).$$

In this paper, we will consider a greedy scheduling policy, called Largest-Deficit-First (LDF) [6] based on the following *deficit process* $D^\mu(t)$ (also known as debts or virtual queues)

1. (Initialization) $D_l^\mu(0) = 0$ for all $l \in \mathcal{K}$.
2. (Dynamics) The evolution of the deficit process for link l is given by

$$D_l^\mu(t) = [D_l^\mu(t-1) + (B_l(t) - B_l(t-1)) - (S_l^\mu(t) - S_l^\mu(t-1))]^+,$$

where $(\cdot)^+ = \max\{0, \cdot\}$ and $B_l(t)$ is the cumulative deficit arrival on link l given by

$$B_l(0) = 0$$

and

$$B_l(t) - B_l(t-1) = \sum_{n=A_l(t-1)+1}^{A_l(t)} c_l(n),$$

where by definition $B_l(t) - B_l(t-1) = 0$ if $A_l(t-1) = A_l(t)$, and $c_l(\cdot)$ is an i.i.d. Bernoulli process with mean p_l. Hence $c_l(n)$ determines whether the n^{th} arriving packet on link l is counted as a deficit arrival or not.

Observe from the definition that the deficit process keeps track on the amount of service we owe to a link in order to fulfill the maximum allowable packet drop probability. To see that, note that the arrival rate of deficit on link l is $\alpha_l p_l$. The deficit of link l reduces by one when a packet is successfully transmitted over link l before its deadline. So if all deficits are bounded, then the requirements on packet drop probabilities are fulfilled.

The LDF scheduling policy is defined as follows. At each time slot, LDF first sorts the links $\mathcal{K}$ according to the current deficits D with arbitrary tie-breaks, and gets the index vector I such that $D_{I_1} \geq D_{I_2} \geq \cdots \geq D_{I_K}$. LDF starts with the *selection* $\mathcal{E} = \{I_1\}$, which only consists of the link with the largest deficit. Then LDF repeatedly considers the link with the next largest deficit I_i for i from 2 to K and adds it into the selection $\mathcal{E}$ if the following two conditions are satisfied:

1. Link I_i does not interfere with any link in $\mathcal{E}$.
2. There is at least one packet available for transmission on link I_i; i.e., $Q_{I_i} > 0$, where Q_l is the number of available packets on link l.

The procedure ends when all links have been considered, and the final selection of links is the desired LDF schedule.

3. PRELIMINARIES

In this section, we introduce basic definitions on stability and efficiency ratio that will be used in the following sections. We first define the stability of the system [17].

DEFINITION 1. *The system is* stable *under a scheduling policy μ if the corresponding deficit process $D^\mu(\cdot)$ satisfies*

$$\limsup_{C\to\infty} \limsup_{t\to\infty} \Pr\left(\sum_{l\in\mathcal{K}} D_l^\mu(t) \geq C\right) = 0.$$

Obviously, the stability of the system depends on the arrival distributions given by $f(\cdot)$, the maximum delay distribution given by $\gamma(\cdot)$, and the required minimum delivery rate $p = (p_l : l \in \mathcal{K})$. Without loss of generality, we fix f and γ and consider the stability of the system in terms of the deficit arrival rate $\lambda = (\lambda_l : l \in \mathcal{K})$ with $\lambda_l = \alpha_l p_l$. We then have the following definition for characterizing such a relation.

DEFINITION 2. *The deficit arrival rate vector λ is* supportable *by a scheduling policy if the system is stable under that policy with deficit rate λ_l for each link l.*

DEFINITION 3. *The* stability region *of a scheduling policy μ is*

$$\Lambda_\mu = \{\lambda \succeq 0 : \lambda \text{ is supportable by } \mu\},$$

where $\succeq$ denotes pairwise greater than or equal to.

Let the set of all causal scheduling policies be $\mathcal{M}$, where a causal scheduling policy, also known as an online policy, is one that makes decision on past and statistical information but not future information. We then have the following characterization.

DEFINITION 4. *The* stability region *of the system is*

$$\Lambda = \bigcup_{\mu\in\mathcal{M}} \Lambda_\mu.$$

That is, the stability region is the set of deficit arrival vectors that can be supported by some causal scheduling policy.

For a given scheduling policy μ, the efficiency ratio of the scheduling policy is defined as follows.

DEFINITION 5. *The* efficiency ratio *of a scheduling policy μ is*

$$\gamma_\mu^* = \sup\{\gamma : \gamma\Lambda \subseteq \Lambda_\mu\}.$$

While refined characterizations of the stability region are possible [14, 15], the efficiency ratio is still a critical metric to evaluate the throughput performance of a scheduling policy.

4. MAIN RESULTS

In this section we present the main results of the LDF policy for scheduling real-time traffic in wireless networks. The first result is Theorem 1, which provides a lower bound on the efficiency ratio of the LDF policy, called the real-time local-pooling factor. The second result is Theorem 2, which states that the efficiency ratio is at least $1/(\beta+1)$ in a network with interference degree β.

We provide a roadmap of the proof of Theorem 1 in Figure 2. The goal of Theorem 1 is to establish the connection between Λ, the stability region of the system, and Λ_{LDF}, the stability region of the LDF policy. However, characterizing Λ turns out to be extremely difficult due to the general arrival and maximum delay distributions. We therefore have to introduce a region called $\Lambda_{\text{NC}}(F)$, which is the stability region by dividing the time into frames with length F and

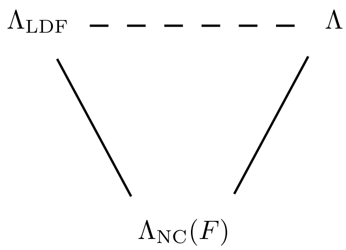


Figure 2: A roadmap of the proof of Theorem 1

assume (i) all information within a frame (arrivals and maximum delays) are known at the beginning of a frame and (ii) at the end of a frame, all packets that have not been transmitted are dropped. The region is denoted by $\Lambda_{\text{NC}}(F)$ since the frame length is F and the system is non-causal because of condition (i).

This novel frame concept was first introduced by Hou et al. [6] for real-time scheduling in wireless networks and provides an analytical framework for understanding real-time communication in wireless networks. The framework has then been extended to general networks and traffic patterns. In particular, the capacity of the non-causal system and heterogeneous deadlines has been characterized by Jaramillo et al. [11]; i.e., $\Lambda_{\text{NC}}(F)$ is known.

We will use $\Lambda_{\text{NC}}(F)$ to bridge Λ and Λ_{LDF}. In the theorem, we will first show that

$$\text{int}(\Lambda) \subseteq \liminf_{F\to\infty} \Lambda_{\text{NC}}(F),$$

where $\text{int}(\Lambda)$ is the interior of the set Λ and $\liminf_{F\to\infty} \Lambda_{\text{NC}}(F)$ is the limit set of $\Lambda_{\text{NC}}(F)$ as F goes to infinity. After that, we will prove that

$$\sigma^* \,\text{int}\,(\Lambda_{\text{NC}}(F)) \subseteq \Lambda_{\text{LDF}},$$

where σ^* is the constant, called the real-time local-pooling factor whose definition is presented in Section 4.1. Combining the two results together, we will be able to prove that σ^* is a lower bound on the efficiency ratio. *We remark that the second step is non-trivial since we will compare the time-slot-based, causal LDF (not frame based LDF) with the frame-based, non-causal system.*

As for the proof of Theorem 2, we first show that the real-time local-pooling factor used in Theorem 1 has a lower bound, which is the ratio of the number of links scheduled under the (causal) LDF to the maximum number of links that can be scheduled within a frame. A similar result has been observed by Reddy et al. (Theorem 3 [18]) for characterizing the local-pooling factor for fading channels. We then make use of the fundamental fact that a one-hop neighborhood of a link under the one-hop interference model contains at most two scheduled links in one time slot, and group the scheduled links by LDF and any arbitrary policy in such a way that each link scheduled by LDF corresponds to at most three links by the other policy, with the correspondence covering all the scheduled links by both policies. At that point, the 1/3 result can be obtained. The proof can be easily generalized to get the $1/(\beta+1)$ bound where the interference degree is β.

4.1 Real-Time Local-Pooling Factor

We will define a quantity analogous to the local-pooling factor [13] and the fading local-pooling factor [18]. Before we do that, we need the following two definitions.

Definition 6. *A non-causal, frame-based scheduling policy called* F-framed *for abbreviation is defined as follows: the packet arrivals and deadlines in the k^{th} frame are known at the beginning of the frame and all packets that arrive during the k^{th} frame are dropped at the end of the frame if not transmitted. Formally, for any $l \in \mathcal{K}$ and positive integer n with $\zeta_l^\mu(n) > 0$, there exists a positive integer k such that*

$$kF + 1 \le b_l(n) \le \zeta_l^\mu(n) \le (k+1)F,$$

where $\zeta_l^\mu(n)$ was defined in Section 2 in the deadline constraint condition.

Let the set of all F-framed policies be $\mathcal{M}_{\text{NC}}(F)$. Note that $\mathcal{M}_{\text{NC}}(F)$ is not a subset of $\mathcal{M}$ since policies in $\mathcal{M}_{\text{NC}}(F)$ can be non-causal. The frame concept (alternatively called intervals or periods) has been used in the literature for tractable analytical analysis of delay constrained traffic [7, 8, 9, 10, 11], where packets that arrive in a frame have deadlines in the same frame. In this paper, we adopt this concept to derive the real-time local-pooling factor for the general traffic model.

Definition 7. *The* stability region of F-framed policies *for a positive integer F is*

$$\Lambda_{\text{NC}}(F) = \bigcup_{\mu \in \mathcal{M}_{\text{NC}}(F)} \Lambda_\mu.$$

We now introduce some notations needed for the main results. Let $\mathcal{J}(F)$ be the set of arrival and maximum delay patterns in a frame of F time slots. We will call an element of $\mathcal{J}(F)$ an F-pattern. An F-pattern is represented by $J = (A, \tau)$ with $A = (A_l(t) \colon l \in \mathcal{K}, 1 \le t \le F)$ and $\tau = (\tau_l(n) \colon l \in \mathcal{K}, 1 \le n \le A_l(F))$, where $A_l(t)$ is the cumulative packet arrival to link l by time slot t in the frame, and $\tau_l(n)$ is the maximum delay associated with the n^{th} packet on link l. Due to the i.i.d. distributions of the arrival and deadline given by f and γ, there is a stationary distribution of the set of F-patterns, denoted by $(\pi(J) \colon J \in \mathcal{J}(F))$.

For a given F-pattern $J = (A, \tau)$, a *schedule* $s = (s_l(n) \colon l \in \mathcal{K}, 1 \le n \le A_l(F))$ specifies the time slot at which each packet is scheduled to be transmitted (if it ever gets scheduled), where $s_l(n)$ is a nonnegative integer that indicates the n^{th} packet on link l is scheduled at time slot $s_l(n)$ if $s_l(n) \in \{1, 2, \ldots, F\}$, and is never scheduled if $s_l(n) = 0$. A schedule s is *feasible* for the F-pattern J if each scheduled packet is scheduled within its feasible scheduling window and no two interfering packets (either two packets on the same link or two packets on two interfering links) are scheduled at the same time slot. We also say that a schedule s is *maximal* for J if no more packets can be further scheduled (i.e., no $s_l(t)$ can be changed from 0 to a positive integer) without breaking feasibility. We denote the maximal feasible schedules for J by $S^*(J)$.

We define the *total service vector of schedule* s to be the column vector $W(s) = (W_i(s) \colon i \in \mathcal{K})$ with $W_i(s) = \sum_{n=1}^{A_i(F)} I(s_i(n) \ne 0)$, where $I(\cdot)$ is the indicator function. Then $W(s)$ is the vector of total number of scheduled packets on each link for the schedule s. Let the *maximal service*

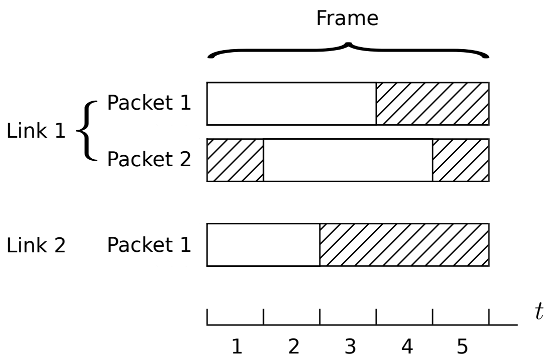


Figure 3: An example of a 5-pattern for two links

matrix for J be

$$M_J = \{W(s)\colon s \in S^*(J)\}.$$

Then the columns of M_J are the total service vectors of the maximal schedules. We note that M_J does not contain all-zero columns if and only if J includes at least one packet arrival on some link, since schedules in $S^*(J)$ are maximal. Similarly, define $M_{J,L}$ to be the maximal service matrix restricted to the set of links L for given pattern J. Then $M_{J,L}$ has no all-zero columns if and only if the pattern J includes at least one packet on some link in L. Also note that $M_{J,L}$ has $|L|$ rows while M_J has $|\mathcal{K}|$ rows.

We use the example in Figure 3 to illustrate the above notations and the concept of the maximal service matrix. As shown in the figure, we consider a frame with size 5 and a 5-pattern J with two packets arriving to link 1 and one packet arriving to link 2, whose arriving times and deadlines are indicated by the blank bars in the figure. The corresponding pattern can be represented by $J = (A, \tau)$, where $A_1(\cdot) = (1, 2, 2, 2, 2)$, $A_2(\cdot) = (1, 1, 1, 1, 1)$, $\tau_1(\cdot) = (3, 3)$, and $\tau_2(\cdot) = (2)$. Assume the two links interfere with each other, so at each time slot only one of them can be scheduled. We can check that there are eight maximal feasible schedules in $S^*(J)$ as follows:

$$s^1 = \begin{pmatrix} 1 & 2 \\ 0 & \end{pmatrix}, s^2 = \begin{pmatrix} 1 & 3 \\ 2 & \end{pmatrix}, s^3 = \begin{pmatrix} 1 & 4 \\ 2 & \end{pmatrix}, s^4 = \begin{pmatrix} 2 & 3 \\ 1 & \end{pmatrix},$$

$$s^5 = \begin{pmatrix} 2 & 4 \\ 1 & \end{pmatrix}, s^6 = \begin{pmatrix} 3 & 2 \\ 1 & \end{pmatrix}, s^7 = \begin{pmatrix} 3 & 4 \\ 1 & \end{pmatrix}, s^8 = \begin{pmatrix} 3 & 4 \\ 2 & \end{pmatrix},$$

where the first row of s^i is the schedule for the two packets on link 1, and the second row is the schedule for the packet on link 2. Then the total service vectors are

$$W(s^1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \text{ and } W(s^i) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ for } 2 \le i \le 8.$$

Hence the maximal service matrix is

$$M_J = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}.$$

We remark from the above example that unlike in the scenario of non-real-time traffic [5], the total service vector of one maximal schedule could be dominated by that of another in the real-time setting. Thus the maximal service matrix M_J can be huge and hard to compute, especially for large frame size F and complex traffic pattern J.

Definition 8. *The* real-time local-pooling factor *(R-LPF) for the F-framed scheduling policies for a set of links $L \subseteq \mathcal{K}$ is*

$$\sigma_L^*(F) = \inf\{\sigma\colon \exists \phi_1, \phi_2 \in \Phi_L(F) \text{ such that } \sigma\phi_1 \succeq \phi_2\},$$

where $\Phi_L(F)$ is the stability region restricted to the set of links $L \subseteq \mathcal{K}$ *defined by*

$$\Phi_L(F) = \{\phi\colon \phi = \sum_{J \in \mathcal{J}(F)} \pi(J)\eta_J, \eta_J \in \mathcal{CH}(M_{J,L})\},$$

and $\mathcal{CH}(M_{J,L})$ defines the convex hull over the columns of the matrix $M_{J,L}$.

Definition 9. *The R-LPF for the F-framed scheduling policies is*

$$\sigma^*(F) = \min_{L \subseteq \mathcal{K}} \sigma_L^*(F).$$

Definition 10. *The R-LPF for the system is*

$$\sigma^* = \liminf_{F \to \infty} \sigma^*(F).$$

We then have the following theorem stating that R-LPF is a lower bound on the efficiency ratio of LDF. The proof is presented in Section 5.1.

Theorem 1. $\gamma_{\text{LDF}}^* \ge \sigma^*$.

By the definition of R-LPF, we can get the R-LPF by solving the following linear program for each $L \subseteq \mathcal{K}$, as suggested in [5]:

$$\begin{aligned} \sigma_L^*(F) = \max_{x, (\rho(J)), (\theta(J))} \quad & \sum_{J \in \mathcal{J}(F)} \pi(J)\rho(J) \\ \text{s.t.} \quad & x' M_{J,L} \succeq \rho(J)\mathbf{1}' \quad \forall J \in \mathcal{J}(F) \\ & x' M_{J,L} \preceq \theta(J)\mathbf{1}' \quad \forall J \in \mathcal{J}(F) \\ & \sum_{J \in \mathcal{J}(F)} \pi(J)\theta(J) = 1 \\ & x \succeq 0, \end{aligned}$$

where x is a column vector of length L, $\rho(J)$ and $\theta(J)$ are scalars for all J, $\mathbf{1}$ is the all-one column vector of length equal to the number of columns of $M_{J,L}$, which we denote by $r_{J,L}$, and x' is the transpose of x. That said, computing the exact R-LPF is usually complex, as it involves roughly

$$\sum_{L \subseteq \mathcal{K}} \left(2 \sum_{J \in \mathcal{J}(F)} r_{J,L} + 1 \right)$$

constraints for each F, which increases exponentially with both the size of the network $\mathcal{K}$ and the frame size F. Thus, we seek lower bounds of the R-LPF in the next subsection.

4.2 A Lower Bound on R-LPF under One-Hop Interference Model

Theorem 2. *Given a network with interference degree β, which is the maximum number of links in the interference neighborhood of some link that can be scheduled without interference, we have*

$$\sigma^* \ge \frac{1}{\beta + 1}. \tag{1}$$

So under the one-hop interference model, $\sigma^* \ge 1/3$.

Remark. This result for one-hop interference model is related to the well-known result that LQF has efficiency ratio at least 1/2 in packet switches [19, 4] and in wireless networks under the one-hop interference [16, 20]. The result makes use of the fact that a one-hop neighborhood contains at most two scheduled links under such an interference model.

Proof outline. We only give the proof for the one-hop interference model case where $\beta = 2$ since the general case follows from nearly identical arguments. The result is proved by showing that $\sigma_L^*(F) \geq 1/3$ for any L and F. The proof consists of two lemmas. The first lemma gives a lower bound on the R-LPF similar to Theorem 3 in [18]. Basically the lower bound on $\sigma_L^*(F)$ can be found by finding the ratio between the "smallest" maximal schedule and the "largest" maximal schedule, both in terms of the number of links scheduled.

The second lemma uses a *multigraph* representation of the network. A multigraph is a graph that allows multiple edges between two nodes. The idea is that we transform the original graph of the network into a multigraph by replacing a link with multiple packets by multiple links with a single packet on each link. Then we prove that the number of links scheduled under LDF over a frame is at least 1/3 of any scheduling policy over the same frame.

The detailed proofs can be found in Section 5.2.

5. PROOFS

5.1 Proof of Theorem 1

Lemma 1. *The stability region of the F-framed policies can be characterized by*

$$\overline{\Lambda_{\mathrm{NC}}(F)} = \left\{ \lambda \succeq 0 \colon \lambda F \preceq \sum_{J \in \mathcal{J}(F)} \pi(J)\eta_J, \eta_J \in \mathcal{CH}(M_J) \right\},$$

where $\overline{A}$ denotes the closure of A and $\mathcal{CH}(M_J)$ is the convex hull over the set of columns of M_J.

Proof of Lemma 1. If λ is strictly outside the stability region, it can be proved that the total amount of deficits increase to infinity with probability one using the strictly separating hyperplane theorem [1] and Lyapunov drift arguments. If λ is strictly inside the stability region, then we can find $\eta = (\eta_J \colon J \in \mathcal{J}(F))$ that dominates λ and make the long-term-average of the scheduling process be at least $\eta \succ \lambda$, where $\succ$ denotes strict pairwise greater than. So the system can be stabilized. □

We then have the next lemma.

Lemma 2. *If $F > \tau_{\max}$, then*

$$\Lambda_{\mathrm{LDF}} \supseteq \sigma^* \cdot \mathrm{int}(\Lambda_{\mathrm{NC}}(F)).$$

Proof of Lemma 2. Let $\lambda' \in \mathrm{int}(\Lambda_{\mathrm{NC}}(F))$ and let $\lambda = \sigma^* \lambda'$. Then by the definition of interior point and the characterization of $\Lambda_{\mathrm{NC}}(F)$ in Lemma 1, there exist $(\xi_J \colon J \in \mathcal{J}(F))$ with $\xi_J \in \mathcal{CH}(M_J)$ for each $J \in \mathcal{J}(F)$ and $\delta > 0$ such that

$$\lambda' + \delta \mathbf{1} \preceq \frac{1}{F} \sum_{J \in \mathcal{J}(F)} \pi(J)\xi_J, \quad (2)$$

where $\mathbf{1}$ is a vector with all 1's. Let $D(t)$ and $S(t)$ be the cumulative deficit and service processes under LDF (without frame). We sample $D(t)$ and $S(t)$ every F time slots, and let the fluid limits of sampled $D(tF)$ and $S(tF)$ be $\bar{D}(t)$ and $\bar{S}(t)$. The construction of fluid limits follows the standard procedure given in [3]. Let $L_0(t)$ be the set of links with the largest deficit fluids, and let $L(t) \subseteq L_0(t)$ be the set of links in $L_0(t)$ with largest derivatives at time t; i.e.,

$$L_0(t) = \left\{ l \in \mathcal{K} \colon \bar{D}_l(t) = \max_{i \in \mathcal{K}} \bar{D}_i(t) \right\}$$

and

$$L(t) = \left\{ l \in L_0(t) \colon \frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_l(t) = \max_{i \in L_0(t)} \frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_i(t) \right\},$$

where we assume t is a regular point; i.e., the derivatives of the fluid limits exist at t. Then we can construct $\eta_J \in \mathcal{CH}(M_J)$ such that for any $l \in L(t)$, the service fluids satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{S}_l(t) \geq \sum_{J \in \mathcal{J}(F)} \pi(J)\eta_{J,l} - \tau_{\max}, \quad (3)$$

where $\eta_{J,l}$ is the l'th entry of the vector $\eta_J \in \mathcal{CH}(M_J)$ for all $J \in \mathcal{J}(F)$. To understand (3), note that $\bar{S}(t)$ is the fluid limit of $S(t)$ sampled every F time slots, so the derivative of $\bar{S}(t)$ is the average service over F time slots under LDF. Now consider a frame of F time slots with arrival and maximum delay pattern J, and denote by s_J^{LDF} the link schedule under LDF during the F time slots. We next construct another link schedule s_J^F, which is a maximal link schedule under the F-framed policy. The construction is to remove those transmissions in s_J^{LDF}, which serve packets that arrived before the frame started, and then add more transmissions to make it a maximal schedule. So for link l,

$$W(s_J^{\mathrm{LDF}})_l - o_{J,l} + n_{J,l} = W(s_J^F)_l,$$

where $o_{J,l}$ is the number of removed transmissions on link l, $n_{J,l}$ is the number of added transmissions on link l, $W(s_J^F) \in \mathcal{M}_J$, and $(\cdot)_l$ denotes the l^{th} component of the vector. Note that those removed transmissions must occur at the first $\tau_{\max}$ time slots because the maximum delay is $\tau_{\max}$ so none of the packets that arrived before the frame can be transmitted after the first $\tau_{\max}$ time slots. This also implies that the added transmissions must be in the first $\tau_{\max}$ time slots as well. Therefore, $n_{J,l} \leq \tau_{\max}$, and

$$W(s_J^{\mathrm{LDF}})_l \geq W(s_J^F)_l - \tau_{\max}$$

holds for any J.

Now assuming $\bar{D}_l(t) > 0$ for $l \in L(t)$, the derivative of $\bar{D}_l(t)$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_l(t) = \lambda_l F - \frac{\mathrm{d}}{\mathrm{d}t}\bar{S}_l(t) \quad (4)$$

$$\leq \lambda_l F - \sum_{J \in \mathcal{J}(F)} \pi(J)\eta_{J,l} + \tau_{\max} \quad (5)$$

$$\leq \sigma^* \left(\sum_{J \in \mathcal{J}(F)} \pi(J)\xi_{J,l} - \delta F \right)$$

$$- \sum_{J \in \mathcal{J}(F)} \pi(J)\eta_{J,l} + \tau_{\max} \quad (6)$$

$$= \left[\sigma^* \left(\sum_{J \in \mathcal{J}(F)} \pi(J)\xi_{J,l} \right) \right. \quad (7)$$

$$\left. - \left(\sum_{J \in \mathcal{J}(F)} \pi(J)\eta_{J,l} \right) \right] \quad (8)$$

$$- \sigma^* \delta F + \tau_{\max}, \quad (9)$$

where (5) comes from (3), and (6) holds because $\lambda = \sigma^*\lambda'$ and inequality (2). By the definition of R-LPF and the fact that $L(t)$ has higher scheduling priority over $\mathcal{K}\backslash L(t)$, there exists $i \in L(t)$ such that

$$\sigma^* \left(\sum_{J \in \mathcal{J}(F)} \pi(J)\xi_{J,i} \right) \leq \left(\sum_{J \in \mathcal{J}(F)} \pi(J)\eta_{J,i} \right).$$

Thus by definition of $L(t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_l(t) = \frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_i(t) \leq \tau_{\max} - \sigma^*\delta F.$$

We note that for any positive integer k,

$$\Lambda_{\mathrm{NC}}(F) \subseteq \Lambda_{\mathrm{NC}}(kF),$$

since any F-framed policy is a valid but more restrictive kF-framed policy. Then (2) holds with the same δ for any frame size kF. Thus for large enough integer k, the deficit fluid limits associated with the frame size kF satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{D}_l(t) \leq \tau_{\max} - \sigma^*\delta kF \leq -\epsilon < 0$$

for some $\epsilon > 0$. Then using the results from [3] we get that the system is positive recurrent under LDF, and hence stable in the sense of Definition 1. □

LEMMA 3. *For any F,*

$$\Lambda_{\mathrm{NC}}(F) \supseteq \mathrm{int}\left(\Lambda \cap \left(\Lambda - \frac{\tau_{\max}}{F}\mathbf{1}\right)\right),$$

where $\Lambda - \frac{\tau_{\max}}{F}\mathbf{1} = \{\lambda - \frac{\tau_{\max}}{F}\mathbf{1} \colon \lambda \in \Lambda\}$.

PROOF OF LEMMA 3. Note that given the schedules of any causal policy, we can convert them into valid schedules under the F-framed policy by remove those transmissions that serve those packets whose arrival times and transmission times are not in the same frame. For each link, we need to remove at most $\tau_{\max}$ transmissions within a frame of F time slots, which is equivalent to at most $\tau_{\max}/F$ packets per time slot. So the lemma holds. □

We can now proceed to prove Theorem 1.

PROOF OF THEOREM 1. By Lemma 2 and Lemma 3, we have

$$\Lambda_{\mathrm{LDF}} \supseteq \sigma^* \cdot \mathrm{int}\left(\Lambda \cap \left(\Lambda - \frac{\tau_{\max}}{F}\mathbf{1}\right)\right).$$

The theorem holds by letting $F \to \infty$. □

5.2 Proof of Theorem 2

LEMMA 4.

$$\sigma_L^*(F) \geq \frac{\sum_{J \in \mathcal{J}(F)} \pi(J) n(M_{J,L})}{\sum_{J \in \mathcal{J}(F)} \pi(J) N(M_{J,L})},$$

where $n(M) = \min_j \sum_i M_{ij}$ is the minimum number of scheduled links in a maximal schedule of M, and $N(M) = \max_j \sum_i M_{ij}$ is the maximal number of scheduled links in a maximal schedule of M.

The proof is almost the same as the proof of Theorem 3 in [18] and is thus omitted. We bound the ratio between $n(M_{J,L})$ and $N(M_{J,L})$ under the one-hop interference model in the following lemma.

LEMMA 5. *Under the one-hop interference model, for any $J \in \mathcal{J}(F)$ and any $L \subseteq \mathcal{K}$,*

$$\frac{n(M_{J,L})}{N(M_{J,L})} \geq \frac{1}{3}.$$

PROOF OF LEMMA 5. We fix $J \in \mathcal{J}(F)$ and $L \subseteq \mathcal{K}$ and focus on the arrival and maximum delay pattern given by J restricted to the subset of links L. For each link $l \in L$, replace it with n links (each of which has a single packet arrival in the frame) if the total number of packets arriving on l in the frame is $n \geq 2$, leave it along if the total number of packets arriving on l in the frame is 1, and remove it from our consideration if no packet arrives in this frame according to J. We then get a multigraph whose set of links is denoted by $\mathcal{K}'$, where $K' = |\mathcal{K}'|$ equals the total number of packets arriving on L in the original graph according to J, and each link in $\mathcal{K}'$ represents a packet in the original graph with arriving time and deadline given by J. The interference model of $\mathcal{K}'$ inherits from the 1-hop interference model of $\mathcal{K}$; i.e., two links in $\mathcal{K}'$ interference with each other if they share an end node. Let $I(l)$ denote the set of links that interfere with link l in $\mathcal{K}'$, and by convention assume $l \in I(l)$.

A schedule over the multigraph $\mathcal{K}'$ in the frame is represented by a function

$$s \colon \mathcal{K}' \times \{1, 2, \ldots, F\} \to \{0, 1\}$$
$$(i, t) \mapsto s_i(t)$$

with $s_i(t) = 1$ if link $i \in \mathcal{K}'$ is scheduled by s at time slot t, and $s_i(t) = 0$ otherwise. A schedule s is feasible if no two interfering links are scheduled at the same time slot, no link is scheduled before its arriving time or after its deadline, and each link is scheduled at most once during the entire frame. A feasible schedule s is maximal if no more links can be scheduled without breaking the feasibility. We note that a feasible (or maximal, respectively) packet schedule for J over the original set of links $\mathcal{K}$ corresponds to a feasible (or maximal, respectively) schedule over the multigraph $\mathcal{K}'$ given by J. Let $\mathrm{supp}(s)$ be the support of s, i.e., the set of (link, time slot) pairs of scheduled links by s. Let $\|s\| = \sum_i \sum_t s_i(t)$ be the total number of links scheduled by s. We note that $\|s\| = |\mathrm{supp}(s)|$.

Define the one-hop interference neighborhood of the (link, time slot) pair (i, t) to be the interfering links of link i at time slot t, i.e.,

$$I(i,t) = I(i) \times \{t\} \subseteq \mathcal{K}' \times \{1, 2, \ldots, F\}.$$

We now consider another maximal schedule u, and the set of (link, time slot) pairs that are in supp(u) but not in the one-hop interference neighborhoods of (link, time slot) pairs in supp(s), i.e., the set

$$P = \text{supp}(u) \backslash \bigcup_{(i,t)\in \text{supp}(s)} I(i,t).$$

We note that for any (link, time slot) pair $(j, t') \in P$, we must have $(j, \tilde{t}) \in \text{supp}(s)$ for some $\tilde{t} \neq t'$; in other words, link j must be scheduled in s at some time slot other than t'. This holds because otherwise, link j can be added to s at time t without interfering any links in s (note that (j, t') is not in any interference neighborhoods of (link, time slot) pairs in supp(s)). We use an example in Figure 4 to illustrate this point.

From the analysis above, we know that

$$|P| \leq |\text{supp}(s)|.$$

Furthermore, schedule u can have at most two active (link, time slot) pairs in $I(i, t)$ for any $(i, t) \in \text{supp}(s)$ because of the one-hop interference [19, 4, 16, 20]. Therefore, we have

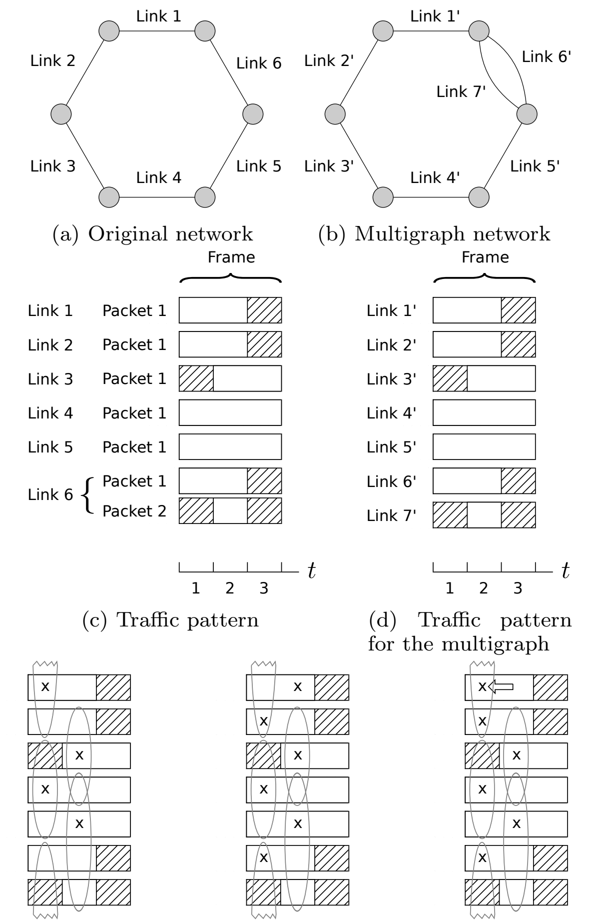
$$\begin{aligned} |\text{supp}(u)| &\leq |P| + \sum_{(i,t)\in\text{supp}(s)} \left|\text{supp}(u) \bigcap I(i,t)\right| \\ &\leq |\text{supp}(s)| + 2|\text{supp}(s)| \\ &= 3|\text{supp}(s)|. \end{aligned}$$

Since s and u are any arbitrary maximal schedules, we have

$$\frac{n(M_{J,L})}{N(M_{J,L})} \geq \frac{1}{3}.$$

□

Theorem 2 is an immediate result of Lemmas 4 and 5, given the one-hop interference model. The proof for general interference is almost identical, and therefore omitted.



(a) Original network (b) Multigraph network

(c) Traffic pattern (d) Traffic pattern for the multigraph

(e) Schedule s (f) Schedule u (g) Modified schedule u'

Figure 4: Consider a six-cycle network with one-hop interference as in (a) and traffic pattern for a frame of 3 time slots as in(c). The network is converted into a multigraph in (b) by dividing the two packets on link 6 into two links with the same end nodes, and the corresponding traffic pattern J is shown in (d). Two maximal schedules, s and u, are given in (e) and (f) with x's denoting the scheduled links, and the one-hop neighborhoods of the scheduled links of s are illustrated by circles. We see that the one-hop neighborhoods of s cover all scheduled links by s, but miss one link scheduled by u. However, as shown in (g), the missed link can be inserted into the one-hop neighborhood in the first time slot so that all scheduled links by u are now covered by the one-hop neighborhoods of s.

6. DISCUSSIONS

In this section we illustrate that the bound in Theorem 2 can be tight in the scenario of collocated networks, where at most one link can be scheduled at each time slot. Notice that in collocated networks any two links interfere with each other, so the interference degree of the network is $\beta = 1$. Hence according to Theorem 2, LDF achieves at least half of the stability region. We now show that there exists an adversarial traffic pattern such that LDF cannot achieve a fraction greater than half of the whole stability region.

Consider a collocated network with two links interfering each other. Suppose the traffic pattern is given as in Figure 5. Assume that the deficits for both links are the same at the beginning of time slot 0. Also assume that when there is a packet arriving to each link (time slots 1, 3, 5, 7, ...), the deficits on both links increase by one with probability $1/2 + \epsilon$ for some small positive ϵ, and remain unchanged with probability $1/2 - \epsilon$. This results in maximum drop rates $p_i = 1/2 - \epsilon$ for $i = 1, 2$. We further assume that when the deficits on the two links are equal, the tie-breaking rule of LDF gives priority to Link 2. Then one can easily see that LDF schedules Link 2 at time slots 1, 5, 9, ..., and schedules Link 1 at time slots 3, 7, 11, ..., while LDF idles at even time slots. Then the average deficit arrival to each link per time slot is $1/4 + \epsilon/2$, and the average deficit departure from each link per time slot is $1/4$. Hence the deficits are not stable under LDF given this traffic pattern. However, one would notice that the optimal scheduler could schedule Link 1 in time slots $4k$ and $4k + 1$ and schedule Link 2 in time slots $4k + 2$ and $4k + 3$, for all positive integer k. Hence the optimal scheduler can stabilize the system when

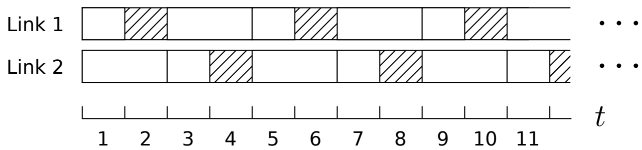


Figure 5: An adversarial traffic pattern for a collocated network with two links. Each blank bar indicates the arriving time and deadline of a real-time packet. The packets arrive on Link 1 at the beginning of time slots 1, 3, 5, 7, . . . , and must be scheduled before the end of time slots 1, 4, 5, 8, ... The packets arrive on Link 2 at the beginning of time slots 1, 3, 5, 7, . . . , and must be scheduled before the end of time slots 2, 3, 6, 7, ...

the maximum drop rates are $p_i = 0$ for $i = 1, 2$. By making ϵ arbitrarily small we can see that the lower bound of 1/2 on the efficiency ratio of LDF is tight in this two-link collocated network.

7. SIMULATIONS

In this section we use simulations to evaluate the stability performance of LDF. Since to the best of our knowledge, neither the stability region nor an optimal scheduling policy has been obtained in the literature, we do not have a benchmark for the stability performance of LDF. As a result, we compare LDF to two other scheduling policies that do not depend on frames and evaluate the performance using simulations. The first simple scheduling policy we consider is RandMax, which randomly chooses a maximal schedule over the links with packets in each time slot. The other one is MaxWeight, which chooses a maximal schedule with the maximum deficit sum over the links with packets in each time slot.

We first considered a 4-link linear network with one-hop interference. We assumed the packet arrival distribution is binomial with number of trials 2 and success probability 0.5, and the maximum delay distribution is uniform over $\{2, 3, 4\}$. This gives us packet arrival rate $\bar{\alpha} = 1$ and mean maximum delay $\bar{\tau} = 3$. We varied the minimum delivery rate to vary the deficit arrival rate. We compare the average deficit sums of the last 1,000 iterations under the three policies, where each simulation is run for 100,000 iterations. The results are shown in Figure 6. As can be observed from the figure, LDF and MaxWeight have similar stability performance, achieving a maximum deficit arrival rate of roughly 0.5 and significantly outperform the simple RandMax policy, which achieves a maximum deficit arrival rate of roughly 0.33. We further remark that for non-real-time traffic, the maximum deficit arrival rate is 0.5. Thus both LDF and MaxWeight have a near-optimal performance in this case.

We also consider a nine-cycle network with two-hop interference, whose non-real-time local-pooling factor is 2/3. The arrival and deadline distributions are the same as the previous case, and the number of iterations is 100,000. The results are shown in Figure 7. Note that in this example, RandMax is still the worst of the three, achieving a maximum deficit arrival rate roughly 0.12, while MaxWeight is slightly better than LDF, both of which achieve a maximum deficit arrival rate roughly 0.16. We note that for non-real-time traffic the maximum deficit arrival rate is 1/3. As we have been trying to convey in this paper, the stability region for the specific packet arrival and deadline distribution is unknown. We only know that the maximum rate for the real-time traffic is $\bar{\lambda} \leq 1/3$. Note that the nine-cycle has an interference degree of 2, so by Theorem 2, LDF has an efficiency ratio of 1/3, which agrees with the simulation result since $0.16 > \frac{1}{3} \times \frac{1}{3} \geq \frac{1}{3}\bar{\lambda}$.

Therefore, both simulations imply good throughput performance of LDF and validate our lower bound on the efficiency ratio.

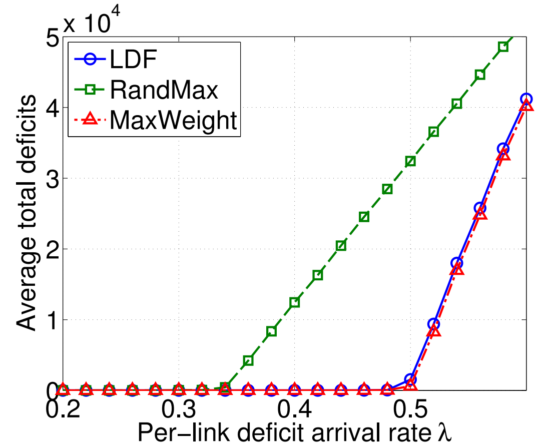


Figure 6: Comparison of the three scheduling policies on a four-linear network with one-hop interference

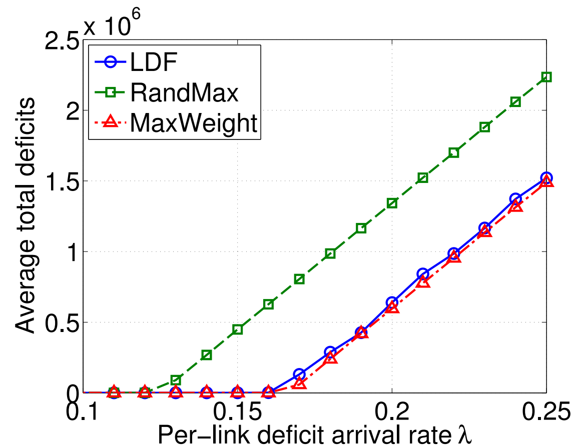


Figure 7: Comparison of the three scheduling policies on a nine-cycle network with two-hop interference

8. CONCLUSIONS

In this paper we considered the problem of scheduling real-time traffic in wireless networks under general stochastic arrivals and deadlines and general interference model. The fraction of delivered packets at a link is required to be no less than a certain threshold. We used deficits to inspect the stability of the system, and studied the stability performance of a scheduling policy that we call the largest-deficit-first (LDF) policy. We proved that the efficiency ratio of LDF can be lower bounded by a quantity that we call the real-time local-pooling factor (R-LPF). Furthermore, we proved that in a network with interference degree β, the R-LPF is at least $1/(\beta+1)$. In particular, for the one-hop interference model, we proved that the R-LPF is at least 1/3.

9. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF) under Grants CNS-1261429, CNS-1264012, and CNS-1262329, in part by the Defense Threat Reduction Agency (DTRA) under Grant HDTRA1-13-1-0030.

10. REFERENCES

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

[2] P. Chaporkar, K. Kar, and S. Sarkar. Throughput guarantees through maximal scheduling in wireless networks. In *Proc. Annu. Allerton Conf. Communication, Control and Computing*, pages 28–30, Monticello, IL, Sept. 2005.

[3] J. G. Dai. On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Ann. Appl. Probab.*, 5(1):49–77, 1995.

[4] J. G. Dai and B. Prabhakar. The throughput of data switches with and without speedup. In *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, volume 2, pages 556–564, Tel Aviv, Israel, Mar. 2000.

[5] A. Dimakis and J. Walrand. Sufficient conditions for stability of longest queue first scheduling: Second-order properties using fluid limits. *Adv. in Appl. Probab.*, 38(2):505–521, June 2006.

[6] I.-H. Hou, V. Borkar, and P. R. Kumar. A theory of QoS for wireless. In *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, pages 486–494, Rio de Janeiro, Brazil, Apr. 2009.

[7] I.-H. Hou and P. R. Kumar. Admission control and scheduling for QoS guarantees for variable-bit-rate applications on wireless channels. In *Proc. ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 175–184, New Orleans, LA, May 2009.

[8] I.-H. Hou and P. R. Kumar. Scheduling heterogeneous real-time traffic over fading wireless channels. In *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, pages 1–9, San Diego, CA, Mar. 2010.

[9] I.-H. Hou and P. R. Kumar. Utility-optimal scheduling in time-varying wireless networks with delay constraints. In *Proc. ACM Int. Symp. Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 31–40, Chicago, IL, Sept. 2010.

[10] J. J. Jaramillo and R. Srikant. Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic. *IEEE/ACM Trans. Netw.*, 19:1125–1136, Aug. 2011.

[11] J. J. Jaramillo, R. Srikant, and L. Ying. Scheduling for optimal rate allocation in ad hoc networks with heterogeneous delay constraints. *IEEE J. Sel. Areas Commun.*, 29(5):979–987, May 2011.

[12] C. Joo, X. Lin, and N. B. Shroff. Performance limits of greedy maximal matching in multi-hop wireless networks. In *Proc. IEEE Conf. Decision and Control (CDC)*, pages 1128–1133, New Orleans, LA, Dec. 2007.

[13] C. Joo, X. Lin, and N. B. Shroff. Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks. *IEEE/ACM Trans. Netw.*, 17:1132–1145, Aug. 2009.

[14] B. Li, C. Boyaci, and Y. Xia. A refined performance characterization of longest-queue-first policy in wireless networks. *IEEE/ACM Trans. Netw.*, 19:1382–1395, Oct. 2011.

[15] B. Li, C. Boyaci, and Y. Xia. Performance guarantee under longest-queue-first schedule in wireless networks. *IEEE Trans. Inf. Theory*, 58:5878–5889, Sept. 2012.

[16] X. Lin and N. B. Shroff. The impact of imperfect scheduling on cross-layer rate control in wireless networks. In *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, volume 3, pages 1804–1814, Miami, FL, Mar. 2005.

[17] R. M. Loynes. The stability of a queue with non-independent inter-arrival and service times. *Math. Proc. Cambridge Philos. Soc.*, 58(3):497–520, July 1962.

[18] A. A. Reddy, S. Sanghavi, and S. Shakkottai. On the effect of channel fading on greedy scheduling. In *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, pages 406–414, Orlando, FL, Mar. 2012.

[19] T. Weller and B. Hajek. Scheduling nonuniform traffic in a packet-switching system with small propagation delay. *IEEE/ACM Trans. Netw.*, 5:813–823, Dec. 1997.

[20] X. Wu and R. Srikant. Regulated maximal matching: A distributed scheduling algorithm for multi-hop wireless networks with node-exclusive spectrum sharing. In *Proc. IEEE Conf. Decision and Control (CDC)*, pages 5342–5347, Seville, Spain, Dec. 2005.