Multi-User Scheduling in Wireless Networks with QoS Constraints

Lei Ying, R. Srikant and G. E. Dullerud
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
{lying, rsrikant, dullerud}@uiuc.edu

***Abstract*— We consider a cellular network consisting of a base station and N receivers. The channel states of the receivers are assumed to be identical and independent of each other. The goal is to compare the throughput of two different scheduling policies (a queue-length-based policy and a greedy scheduling policy) given an upper bound on the queue overflow probability. We consider a multi-state channel model, where each channel is assumed to be in one of L states. Given an upper bound on the queue overflow probability, we obtain a lower bound on the throughput of the queue-length-based policy. For sufficiently large N, the lower bound is shown to be tight, strictly increasing with N, and strictly larger than the throughput of the greedy policy.**

I. Introduction

Consider a cellular network consisting of a base station and N users (receivers), where the base station maintains N separate queues, one corresponding to each user. Assume time is slotted and the channel states of the receivers at each time slot are known at the base station. Then, the base station can decide which queues to serve according to their channel states. In this paper, we assume that the base station operates in a TDMA fashion, i.e., the base station can serve only one queue in each time slot. Two scheduling policies have been widely studied in the literature: (i) the base station serves the user with the best (weighted) channel state (opportunistic scheduling) [6], [2]; or (ii) serve the user with the best queue-length-weighted channel state (queue-length based (QLB) scheduling) [4], [5]. While the QLB scheduling is throughput optimal (i.e., can stabilize any set of user throughput that can be stabilized by any other algorithm), opportunistic scheduling maximizes the total network throughput if all queues are continuously backlogged.

While stability is the first concern of scheduling policies, Quality of Service (QoS) is equally important in applications. In [1], [3], [7], it has been shown that QLB policies are superior to greedy policies when there are QoS constraints in the form of upper bounds on mean delays or buffer overflow probabilities. However, all of these papers only consider simple ON-OFF channel models. In this paper, we investigate the performance of the two different scheduling policies for a general multi-state channel model, where the channel has $L \geq 2$ states. The goal is to show that the QLB scheduling performs better than the greedy scheduling under the QoS constraints. The main result is as follows: (i) Assuming a multi-state channel model and a constant arrival rate in each time slot, under the QLB policy, we compute a lower bound on the large-deviations exponent of the probability that at least one queue in the network exceeds a large threshold. We obtain a lower bound on the maximum network throughput under the buffer overflow constraints, and for large N, the lower bound is tight, strictly increasing, and strictly larger than the throughput of the greedy policy. (ii) We prove that, under the buffer overflow constraint, the throughput under the QLB policy is no less than the throughput of the greedy policy for all N.

II. Basic Model

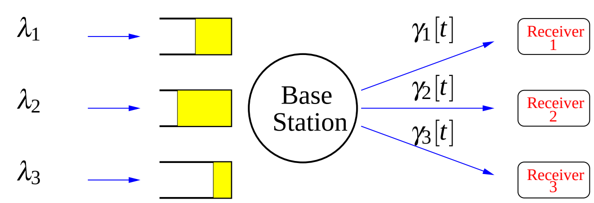

Fig. 1. Single-hop Network

Consider the downlink of a time-slotted cellular network shared by N users, where only one user is allowed to transmit in each time slot. An example with $N = 3$ is shown in Figure 1. Each user is associated with a channel and all channel-state processes $\gamma_i[t]$ are independent and statistically identical. We consider multi-state channels — where each channel has L states $\{0, \ldots, L-1\}$, the probability that the channel is in state l is p_l^c (the superscript "c" indicates it is the probability distribution of the channel states), and we can transmit at most F_l bits/slot when a channel is in state l. Also assume that the arrival rate is equal to λ/N bits/slot for each user and the bits are deposited in the queue at the beginning of each time slot.

We will study the total network throughput of the following scheduling policies:

1) Queue-length based (QLB) policy: Choose user i^* to transmit if

$$i^* \in \arg\max_i \gamma_i[t] Q_i[t],$$

where $Q_i[t]$ is the queue length of user i at the beginning of time slot t.

2) Greedy policy: Choose user i^* if

$$i^* \in \arg\max_i \gamma_i[t].$$

If more than one user has the best channel state, we assume that the base station is equally likely to choose any one of those users.

The QoS constraint considered is

$$\Pr\left(\max_i Q_i(0) > B\right) \leq \varepsilon,$$

where $Q_i(0)$ is the stationary queue length. So this QoS constraint requires the steady-state probability that the queue length is larger than B to be small. Instead of studying this constraint as above, we study the approximation to the constraint given by

$$\theta_B(N,\lambda) := \lim_{B\to\infty} -\frac{1}{B}\log\Pr\left(\max_i Q_i(0) > B\right) \geq \delta, \qquad (1)$$

where the large-deviations exponent θ_B is a function of the number of the users and the total arrival rate. The exponent δ can be related to ε for large B using the approximation $e^{-\delta B} = \varepsilon$. Throughout this paper, we use θ_B to denote the large-deviations exponents, with a superscript to indicate the scheduling policy used ("Greedy" is for the greedy policy and "QLB" is for the queue-length based policy).

III. Queue-Length Based Policy

In this section, we will investigate the performance of the wireless system under the QLB policy. Consider a multi-state-channel system and define the state of the system as the composite state of all the channels. Thus, there are L^N system states. Each system state can be represented as an N-tuple in $\{0,\ldots,L-1\}^N$. To simplify the notation, we use an integer $j\in\{0,\ldots,L^N-1\}$ to represent the system state, and define the system state variable $S(t)$ as follows:

$$S(t) := \sum_{i=0}^{N-1}\gamma_i(t)L^i. \qquad (2)$$

Sometimes, we will also use the N-tuple representation of the system state, and define an N-tuple S^j in $\{0,\ldots,L-1\}^N$ to denote the state j. Let S_i^j be the i^{th} entry of S^j, it is also the state of channel i when the system is in system state j. Further, define a probability vector $\mathbf{p}$ where p_j is the probability the system is in state j, and

$$\bar{\mu} = \sum_{j=0}^{L^N-1}\left(\max_i F_{S_i^j}\right)p_j, \qquad (3)$$

which is the maximum throughput the network can support without QoS constraints. Note that the network is unstable if $\lambda \geq \bar{\mu}$; and large buffer overflow will not happen if $\lambda \leq F_0$. Thus, we assume $F_0 < \lambda < \bar{\mu}$.

For sufficiently large T_s, we define $\mathbf{s}^{(B)}(t)$ on $[-T_s,0]$ using $\mathbf{S}(t)$ on $[0,BT_s]$ as follows:

$$s_j^{(B)}(t) := \frac{1}{B}\sum_{k=0}^{B(T_s+t)} 1_{S(k)=j}$$

for $t=\frac{k}{B}-T_s$ and $k=\{0,\ldots,BT_s\}$, where for values of t which are not of the form k/n, define $s_j^B(t)$ by linear interpolation. Notice that we have scaled and shifted time so that the discrete time units $0,1,\ldots,BT_s$ have now become the continuous time interval $[-T_s,0]$. For each t, the variable $s_j^{(B)}(t)$ is the amount of (scaled) time in the interval $[-T_s,t]$ that the system is in state j. Next, define the system channel rate processes using a L^N-tuple — $\mathbf{u}(t)$, where $\mathbf{u}(t)$ is nonnegative, integrable, and $\sum_{j=0}^{L^N-1}u_j(t)=1$. Further, for B large enough, we have for any $t_1<t_2$

$$\max_j\left|s_j^{(B)}(t_2)-s_j^{(B)}(t_1)-\int_{t_1}^{t_2}u_j(s)\,ds\right| \leq \frac{1}{B}.$$

Next, consider the Kullback-Leibler distance

$$D(\mathbf{u}(t)\|\mathbf{p}) = \sum_{j=0}^{L^N-1}u_j(t)\log\frac{u_j(t)}{p_j}.$$

Refer to $\int_{-T}^0 D(\mathbf{u}(s)||\mathbf{p})\,ds$ as the cost function, and we define following minimum cost problem:

$$\theta_B^{\text{QLB}}(N,\lambda) = \inf_{T,\mathbf{u}}\int_{-T}^0 D(\mathbf{u}(s)||\mathbf{p})\,ds, \qquad (4)$$

where $T\geq 0$, $q_i(-T)=0$ for all i, $\max_i q_i(0)=1$, and the QLB policy is used. We use T^* and $\mathbf{u}^*(t)$ to denote the optimal solution, and use $\mathbf{q}^*(t)$ to denote the corresponding trajectories. It is obvious that the scaled time T_s should be chosen such that $T^*\leq T_s$. It can be shown that T^* is finite for $F_0<\lambda<\bar{\mu}$ (please refer to [8] for details of the proof), so we assume $T_s\geq T^*$ in this paper.

Theorem 1:

$$\theta_B^{\text{QLB}}(N,\lambda) = \lim_{B\to\infty}\frac{-1}{B}\log\Pr(\max_i q_i(0)\geq 1),$$

where $\theta_B^{\text{QLB}}(N,\lambda)$ is defined as (4), and queues are scheduled according to the QLB policy.

Proof: The proof is a straightforward extension of Theorem 6.1 in [3]. ■

Note that the minimum cost problem is intuitively obvious: among all possible channel state trajectories that could lead to overflow, we pick the one that is "closest" to the mean value $\mathbf{p}$. Given $\mathbf{u}(t)$, we call $\int_{-T}^0 D(\mathbf{u}(s)||\mathbf{p})\,ds$ the cost of the trajectory generated by $\mathbf{u}(t)$. Thus, Theorem 1 tells us that the probability of the QoS violation is related to the minimum cost problem (4).

Lemma 2 (Order Property): Given any trajectory, we can find another trajectory that has the same cost and the property such that if $i\geq k$, then

$$q_i(t)\geq q_k(t).$$

Proof: This lemma holds because the channels are symmetric, and the QLB policy is based on the queue lengths and the channel states. So the indices of two queues can be swapped after the two queue lengths become equal, without affecting the cost of the trajectory. Please refer to [8] for details of the proof. ■

Next, define $\mathscr{A}_{M,l}$ to be the set of system states j such that

$$\max_{i\geq N-M} S_i^j = l \geq \max_i S_i^j,$$

and let $\mathscr{P}_{M,l}$ denote the probability that the system state is in $\mathscr{A}_{M,l}$, so

$$\mathscr{P}_{M,l} = \left(\left(\sum_{k=0}^{l} p_k^c\right)^M - \left(\sum_{k=0}^{l-1} p_k^c\right)^M\right)\left(\sum_{k=0}^{l} p_k^c\right)^{N-M} \quad (5)$$

Further, define $\mathscr{A}_M = \left(\bigcup_{l=0}^{L-1} \mathscr{A}_{M,l}\right)^c$, and

$$\mathscr{P}_M = 1 - \sum_{l=0}^{L-1} \mathscr{P}_{M,l}. \quad (6)$$

Now, we define optimization problem OP(M,N,h) and show that the solution of this optimization problem provides us a lower bound to the objective in (4).

$$\text{OP}(M,N,h): \quad C_M^N(h) = \inf_{\mathbf{u},T} TD(\mathbf{u}||\mathbf{p}) \quad (7)$$

$$\text{Subject to}: \quad T\left(\frac{M}{N}\lambda - \sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M,l}} u_j\right)\right) = Mh \quad (8)$$

$$\sum_{j=0}^{L^N-1} u_j = 1, \qquad u_j \geq 0 \quad \forall j. \quad (9)$$

We use $\mathbf{u}_M^*$ and T_M^* to denote the optimal solution of problem (7). Also, from (8), we have

$$C_M^N(h) = \inf_{\mathbf{u}} \frac{(Mh)D(\mathbf{u}||\mathbf{p})}{\frac{M\lambda}{N} - \sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M,l}} u_j\right)} = hC_M^N(1). \quad (10)$$

Theorem 3: For an N-user network with multi-state channels, the objective of the minimum cost problem (4) is lower bounded by $\min_M C_M^N(1)$, thus

$$\theta_B^{\text{QLB}}(N,\lambda) \geq \min_M C_M^N(1).$$

Proof: From Lemma 2, we only need to consider ordered trajectories. Given any ordered trajectory, we segment $[-T,0]$ into small intervals $\{[t_m,t_{m+1})\}$. For each interval $[t_m,t_{m+1})$, there exists an M_m such that for any $t \in (t_m,t_{m+1})$, we have $q_{N-1}(t) = \ldots = q_{N-M_m}(t) > q_{N-M_m-1}(t)$. Define $\mathscr{B}_{M_m}(\mathbf{q}(t))$ to be the set of the system states such that one of users $\{N-M_m,\ldots,N-1\}$ will be scheduled under the QLB policy if the queue lengths are $\mathbf{q}(t)$. Thus, if $q_{N-1}(t_{m+1}) - q_{N-1}(t_m) = h_m$, we have

$$M_m h_m = (t_{m+1}-t_m)\frac{M_m\lambda}{N} - \int_{t_m}^{t_{m+1}} \sum_{j\in\mathscr{B}_{M_m}(\mathbf{q}(s))} F_{l_{j,M_m}} u_j(s)\,ds,$$

where $l_{j,M_m} = \max_{i\geq N-M_m} S_i^j$. Now, suppose the system is in state $j \in \bigcup_{l=0}^{L-1} \mathscr{A}_{M_m,l}$. From the definition of $\mathscr{A}_{M_m,l}$, one of users $\{N-M_m,\ldots,N-1\}$ will be scheduled, so $\bigcup_{l=0}^{L-1} \mathscr{A}_{M_m,l} \subseteq \mathscr{B}_{M_m}(\mathbf{q}(t))$, and $l_{j,M_m} = l$ for $j \in \mathscr{A}_{M_m,l}$. Define an L^N-tuple $\mathbf{K}^m$ such that

$$K_j^m = \frac{1}{t_{m+1}-t_m}\int_{t_m}^{t_{m+1}} u_j(s)\,ds,$$

we have

$$\begin{aligned} M_m h_m &\leq (t_{m+1}-t_m)\frac{M_m\lambda}{N} - \int_{t_m}^{t_{m+1}} \sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M_m,l}} u_j(s)\right) ds \\ &= (t_{m+1}-t_m)\left(\frac{M_m\lambda}{N} - \sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M_m,l}} K_j^m\right)\right). \end{aligned}$$

Choose $\hat{T}_m$ such that

$$M_m h_m = \hat{T}_m\left(\frac{M_m\lambda}{N} - \sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M_m,l}} K_j^m\right)\right), \quad (11)$$

so $\hat{T}_m \leq (t_{m+1}-t_m)$. Note that $D(\mathbf{u}(t)||\mathbf{p})$ is convex in $\mathbf{u}(t)$, from Jensen's inequality, we have

$$\int_{t_m}^{t_{m+1}} D(\mathbf{u}(s)||\mathbf{p})\,ds \geq (t_{m+1}-t_m)D(\mathbf{K}^m||\mathbf{p}) \geq \hat{T}_m D(\mathbf{K}^m||\mathbf{p}).$$

Further, from (11), we know that $\{\hat{T}_m,\mathbf{K}^m\}$ is a feasible solution of OP(M_m,N,h_m), so

$$\int_{t_m}^{t_{m+1}} D(\mathbf{u}(s)||\mathbf{p})\,ds \geq C_{M_m}^N(h_m) \geq \min_M C_M^N(h_m),$$

and

$$\begin{aligned} \int_{-T}^{0} D(\mathbf{u}(s)||\mathbf{p})\,ds &= \sum_m \int_{t_m}^{t_{m+1}} D(\mathbf{u}(s)\|\mathbf{p}) && (12)\\ &\geq \sum_m \min_M C_M^N(h_m) && (13)\\ &= \min_M C_M^N(1), && (14) \end{aligned}$$

where the last equality follows from (10). Thus, $\theta_B^{\text{QLB}}(N,\lambda) \geq \min_M C_M^N(1)$ and the theorem holds. ∎

Suppose $K \in \arg\min_M C_M^N(1)$, the lower bound is tight if the overflow time of the trajectory generated by channel processes $\mathbf{u}_K^*$ is T_K^*. This is not true for the multi-state channel system in general, see a numerical example in [8]. However, we will show that the lower bound is tight for large N.

It is hard to obtain a closed-form expression for $C_M^N(1)$. Instead, we solve the following problem:

$$\text{Rate}(N,M) \quad R_M^N(1) = \inf_{\mathbf{u},T}\left(\frac{N}{T} + \frac{N}{M}\sum_{l=0}^{L-1}\left(F_l \sum_{j\in\mathscr{A}_{M,l}} u_j\right)\right) \quad (15)$$

$$\text{Subject to}: \quad TD(\mathbf{u}||\mathbf{p}) = \theta \quad (16)$$

$$\sum_{j=0}^{L^N-1} u_j = 1, \qquad u_j \geq 0 \quad \forall j. \quad (17)$$

Note that the objective in (15) is the expression for λ obtained from the constraint (8) in problem OP(M,N,h) by letting $h=1$. Further, (16) in Rate(N,M) is a constraint on the objective in OP(M,N,h). Thus, if OP$(M,N,1)$ is viewed as an optimization problem for computing a lower bound on the large-deviations exponent given an arrival rate, then Rate(N,M) can be viewed as an optimization problem for computing a lower bound on the maximum arrival rate that the network can support given a QoS constraint expressed in terms of the large-deviations exponent. It can be verified that the optimal solution $(\mathbf{u}_M^*, T_M^*)$ of (7) is also the optimal solution of (15). Let $\lambda_B^{\text{QLB}}(N,\theta)$ denote the maximum throughput the system can support given the constraint $\Pr(\max_i Q_i(0) > B) \leq e^{-\theta B}$, and define

$$\underline{\lambda}_B^{\text{QLB}}(N,\theta) = \min_M R_M^N(1).$$

From Theorem 3, we have

$$\lambda_B^{\text{QLB}}(N,\theta) \geq \underline{\lambda}_B^{\text{QLB}}(N,\theta).$$

The lower bound is tight if $N \in \arg\min_M R_M^N(1)$, but not so in general. In the next theorem, we obtain a closed-from expression of $\underline{\lambda}_B^{\text{QLB}}(N,\theta)$, and show that it is equal to $\lambda_B^{\text{QLB}}(N,\theta)$ for large N.

Theorem 4: Suppose the QLB policy is used. For the general multi-state channel model, given the queue-overflow constraint $\theta_B^{\text{QLB}}(N,\lambda)=\theta$, the maximum total throughput of the network satisfies

$$\lambda_B^{\text{QLB}}(N,\theta) \geq \min_{1\leq M\leq N}\left(-\frac{N}{\theta}\log\left(\sum_{l=0}^{L-1} e^{-\frac{F_l\theta}{M}}\mathscr{P}_{M,l}+\mathscr{P}_M\right)\right), \quad (18)$$

where $\mathscr{P}_{M,l}$ and $\mathscr{P}_M$ are defined in (5) and (6) respectively. Further, there exists a positive integer N_B^* such that for any $N \geq N_B^*$,

$$\lambda_B^{\text{QLB}}(N,\theta) = -\frac{N}{\theta}\log\left(\sum_{l=0}^{L-1} e^{-\frac{F_l\theta}{N}}\mathscr{P}_{N,l}\right),$$

which implies that

$$\lambda_B^{\text{QLB}}(N,\theta) < \lambda_B^{\text{QLB}}(N+1,\theta),$$

and

$$\lim_{N\to\infty}\lambda_B^{\text{QLB}}(N,\theta) = F_{L-1}.$$

Another consequence is that, for $N \geq N_B^*$, the lower bound on the large-deviations exponent given in Theorem 3 is the large-deviations exponent itself.

Proof: The closed-form expression (18) can be obtained by solving (15) using the Karush-Kuhn-Tucker theorem. It can be shown that

$$u_j^* = \begin{cases} \frac{p_j}{R}e^{-\frac{\theta F_l}{M}} & \text{for } j \in \mathscr{A}_{M,l}; \\ \frac{p_j}{R}, & \text{for } j \in \mathscr{A} \end{cases}, \quad (19)$$

where $R=\sum_{l=0}^{L-1}\mathscr{P}_{M,l}e^{-\frac{F_l\theta}{M}}+\mathscr{P}_M$, and u_j^* is the j^{th} entry of $\mathbf{u}_M^*$.

Next, by analyzing (18), we will show that the lower bound is tight for large N. The main idea is to show that $N \in \arg\min_M R_M^N(1)$ when N is large. Define $\tilde{p}_{-1}=0$, $\tilde{p}_l=\sum_{k=0}^{l}p_k^c$ for $l\geq 0$, and

$$f_B(x) = 1+\sum_{l=0}^{L-1}\left(e^{-\frac{F_l\theta}{x}}-1\right)\tilde{p}_l^N\left(1-\left(\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right)^x\right).$$

We have

$$\underline{\lambda}_B^{\text{QLB}}(N,\theta) \;=\; \min_{1\leq M\leq N}\left(-\frac{N}{\theta}\log f_B(M)\right).$$

Further, it can be shown that the derivative of $f_B(x)$ is

$$\begin{aligned} f_B'(x) \;=\; & \frac{F_0\theta}{x^2}e^{-\frac{F_0\theta}{x}}\tilde{p}_0^N+\sum_{l=1}^{L-1}\frac{e^{-\frac{F_l\theta}{x}}}{x^2}\tilde{p}_l^N\left(F_l\theta-F_l\theta\left(\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right)^x\right. \\ & \left.-x^2\left(\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right)^x\left(1-e^{\frac{F_l\theta}{x}}\right)\log\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right). \end{aligned}$$

Since $\left(\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right)^x$ and $x^2\left(\frac{\tilde{p}_{l-1}}{\tilde{p}_l}\right)^x$ converge to zero when x goes to infinity, there exists $x_B^*>0$, which is independent of N, such that $f_B'(x)>0$ for any $x>x_B^*$. So the lower bound can be rewritten as

$$\begin{aligned} \underline{\lambda}_B^{\text{QLB}}(N,\theta) \;=\; & \min\left\{\min_{1\leq M\leq \min\{x_B^*,N-1\}}\left(-\frac{N}{\theta}(\log f_B(M))\right),\right. \\ & \left.-\frac{N}{\theta}(\log f_B(N))\right\}. \end{aligned}$$

Since $1-e^{-x}\geq 0$ and $\tilde{p}_{L-1}=1$, we have

$$-\frac{N}{\theta}(\log f_B(M)) \geq -\frac{N}{\theta}\log\left(1+\left(e^{-\frac{F_{L-1}\theta}{M}}-1\right)\left(1-\tilde{p}_{L-2}^M\right)\right).$$

Note that x_B^* is independent on N, so for any $M\in[1,x_B^*]$,

$$\lim_{N\to\infty}-\frac{N}{\theta}(\log f_B(M)) = \infty.$$

Further, $\lambda_B^{\text{QLB}}(N,\theta)\leq F_{L-1}$ since F_{L-1} is the maximum service rate one user can receive at each time slot. So there exists N_B^* such that for any $N\geq N_B^*$,

$$\underline{\lambda}_B^{\text{QLB}}(N,\theta) = -\frac{N}{\theta}(\log f_B(N)) = R_N^N(1).$$

From (19), all users are symmetric under $(\mathbf{u}_N^*, T_N^*)$, so $R_N^N(1)$ is achievable. Thus, for any $N\geq N_B^*$, the equality of (18) holds, and

$$\lambda_B^{\text{QLB}}(N,\theta) = -\frac{N}{\theta}\log\left(\sum_{l=0}^{L-1}e^{-\frac{F_l\theta}{N}}\mathscr{P}_{N,l}\right),$$

where $\mathscr{P}_N=0$.

Next, it is easy to show that

$$\lambda_B^{\text{QLB}}(N+1,\theta) > \lambda_B^{\text{QLB}}\left(N+1,\frac{(N+1)\theta}{N}\right) > \lambda_B^{\text{QLB}}(N,\theta),$$

and

$$\lim_{N\to\infty}\lambda_B^{\text{QLB}}(N,\theta) = F_{L-1}.$$

For details of the proof, please refer to [8]. ■

In the theorem above we proved that the lower bound is tight for large N. Also, note that F_{L-1} is the maximum throughput the network can achieve since it is the maximum service rate the user can receive at each time slot.

IV. Greedy Policy

In this section, the greedy policy is considered. Under the greedy policy, the probability $\hat{p}_l$ that a user is in state l and is picked to transmit is given by

$$\hat{p}_l = \frac{1}{N}\left(\left(1-\sum_{j=l+1}^{L-1}p_j^c\right)^N-\left(1-\sum_{j=l}^{L-1}p_j^c\right)^N\right), \quad (20)$$

and the probability that a user is not selected is

$$\hat{p}_{\text{null}} = 1-\sum_{l=0}^{L-1}\hat{p}_l. \quad (21)$$

From the symmetry of the channel states and the greedy scheduling scheme, we know that

$$\Pr(q_i(0)\geq 1) \leq \Pr(\max_i q_i(0)>1) \leq N\Pr(q_i(0)\geq 1),$$

which implies

$$-\lim_{B\to\infty}\frac{1}{B}\log \Pr(q_i(0)\geq 1) = -\lim_{B\to\infty}\frac{1}{B}\Pr(\max_i q_i(0) > 1).$$

Thus, the large-deviations exponent under the greedy policy can be obtained by considering a single-user system with arrival rate λ/N and probability vector $\hat{\mathbf{p}}$, where $\hat{p}_l$ is defined by (20) and (21).

Theorem 5: Suppose the greedy policy is used. Given the buffer-overflow constraint $\theta_B^{\text{Greedy}}(N,\lambda) = \theta$, the maximum throughput the network can support is

$$\lambda_B^{\text{Greedy}}(N,\theta) = -\frac{N\log\left(\sum_{l=0}^{L-1} e^{-F_l\theta}\hat{p}_l + \hat{p}_{\text{null}}\right)}{\theta}.$$

Further, when N goes to infinity,

$$\lim_{N\to\infty}\lambda_B^{\text{Greedy}}(N) \leq \frac{1-e^{-\theta F_{L-1}}}{\theta}.$$

Proof: For details of the proof, please refer to [8]. ■

V. QLB Policy vs Greedy Policy

First, note that $\frac{1-e^{-\theta F_{L-1}}}{\theta} < F_{L-1}$ for $\theta > 0$. So from Theorem 4 and Theorem 5, we can conclude that the throughput of the QLB policy is strictly larger than the throughput of the greed policy for large N. In the next theorem, we further show that the throughput of the QLB policy is no less than the throughput of the greedy policy for all N.

Theorem 6: For an N-user system, the total maximum throughput under the QLB policy is no less than the throughput under the greedy policy:

$$\lambda_B^{\text{QLB}}(N,\theta) \geq \lambda_B^{\text{Greedy}}(N,\theta).$$

Proof: Consider an N-user network with total arrival rate λ. It can be shown that for any $\mathbf{u}(t)$ under the QLB policy, there exits a $\tilde{\mathbf{u}}(t)$ under the greedy policy such that

$$\int_{-T}^{0} D(\mathbf{u}(s)\|\mathbf{p})\,ds \geq \int_{-\tilde{T}}^{0} D(\tilde{\mathbf{u}}(s)\|\tilde{\mathbf{p}})\,ds,$$

where T and $\tilde{T}$ are the overflow times under $(\mathbf{u}(t), \text{QLB policy})$ and $(\tilde{\mathbf{u}}(t), \text{Greedy policy})$ respectively, and $\tilde{p}_{j,i}$ is the probability that user i is selected to transmit under the greedy policy when the system is in state j, i.e.,

$$\tilde{p}_{j,i} = \begin{cases} \frac{1}{m_j}p_j, & \text{if } S_i^j = \max_k S_k^j; \\ 0, & \text{otherwise}, \end{cases}$$

where m_j is the number of channels in state $\max_k S_k^j$ when the system is in state j. So we have

$$\theta_B^{\text{QLB}}(N,\lambda) \geq \theta_B^{\text{Greedy}}(N,\lambda),$$

from which the result follows. For details of the proof, please refer to [8]. ■

Numerical Example: Consider a three-state channel system with $p_0^c = 0.5$, $p_1^c = 0.4$, $p_2^c = 0.1$, $F_0 = 0$, $F_1 = 0.5$, and $F_2 = 1$. Fixing the buffer overflow constraint to be $\theta = 2$, in Figure 2, we plot $\underline{\lambda}_B^{\text{QLB}}(N,\theta)$ and $\lambda_B^{\text{Greedy}}(N,\theta)$ for $N = 1, \ldots, 20$. Numerically, it was observed $\lambda_B^{\text{QLB}}(N,\theta) = \underline{\lambda}_B^{\text{QLB}}(N,\theta)$ for $N \geq 9$. From the numerical example, we can see that the performance of the QLB policy is much better than the performance of the greedy policy.

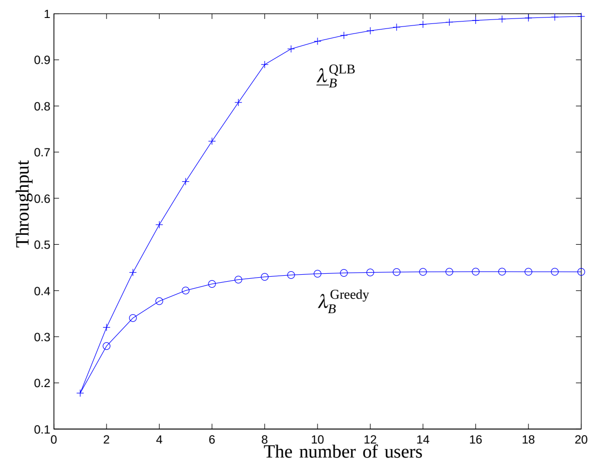

Fig. 2. QLB Policy vs Greedy Policy

VI. Conclusions

In this paper, we extended our earlier large deviations analysis of scheduling policies for ON-OFF wireless channel models to multi-state wireless channel models. We computed a lower bound on the maximum arrival rate that the network can support given a buffer overflow constraint, and showed that this lower bound is tight when the number of users is sufficiently large.

References

[1] A. Ganti, E. Modiano and J. Tsitsiklis. Optimal Transmission Scheduling in Symmetric Communication Models with Intermittent Connectivity, 2004 Preprint.

[2] X. Liu, E. Chong, and N. Shroff. Opportunistic transmission scheduling with resource-sharing constraints in wireless networks. In *IEEE Journal on Selected Areas in Communications*, 19(10):2053 – 2064, October 2001.

[3] S. Shakkottai. Effective Capacity and QoS for Wireless Scheduling, 2004 Preprint.

[4] 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, 1936–1948, December 1992.

[5] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. In *IEEE Transactions on Information Theory*, 39:466-478, March 1993.

[6] P. Viswanath, D. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. In *IEEE Transactions on Information Theory*, 48(6):1277C1294, June 2002.

[7] L. Ying, R. Srikant, and G. E. Dullerud. A Large Deviations Analysis of Scheduling in Wireless Networks. In *Proceedings of 44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, December 2005.

[8] L. Ying, R. Srikant, A. Eryilmaz and G. E. Dullerud. A Large Deviations Analysis of Scheduling in Wireless Networks. Preprint. Available at http://www.ifp.uiuc.edu/~lying/Publications/LD.pdf