OFDM Downlink Scheduling for Delay-Optimality: Many-Channel Many-Source Asymptotics with General Arrival Processes

Manu Sharma and Xiaojun Lin School of Electrical and Computer Engineering, Purdue University West Lafayette, IN, USA. Email: {sharma50, linx}@ecn.purdue.edu

Abstract—We consider the downlink of an OFDM system for supporting a large number of delay-sensitive users. The OFDM scheduling problem can be modeled as a discrete-time multisource multi-server queuing system with time-varying connectivity. For such a system, the Max-Weight policy is known to be throughput-optimal and the Server-Side Greedy (SSG) policy has been recently shown to achieve small queue lengths for i.i.d. arrival processes. However, there is often significant difference between queue-length optimality and delay optimality, and there exist arrival patterns such that algorithms with small queue backlog can still lead to large delay. In this work, we propose a new OFDM scheduling algorithm that gives preference to packets with large delay. Assuming ON-OFF channels, we calculate upper and lower bounds for the delay violation probability for a large class of arrival processes. We discuss the cases where the proposed algorithm is rate-function delay-optimal, and we also show that these bounds can be used to construct an admission control sceme. We substantiate the result via both analysis and simulation.

I. INTRODUCTION

Next generation OFDM-based wireless cellular systems (e.g. WiMax and LTE) are envisioned to provide much higher data rate and larger system capacity. It is conceivable that in the future, both voice, data, and video traffic can be carried on a single packet-based OFDM system, eliminating the need to maintain separate voice networks. An important problem in the realization of this goal is the design of scheduling algorithms that provide low-delay guarantees to delay sensitive voice/video users. In a typical OFDM system, the bandwidth available to the base-station is partitioned into hundreds of orthogonal carriers. A given user can be served by multiple frequency carriers simultaneously, and the allocation of carriers to users can change over every time-slot. How user transmissions should be scheduled over frequency and time will have a significant impact on the delay performance of the system. Morevoer, an efficient admission control algorithm is needed to ensure that the network capacity is fully utilized while meeting the delay requirement of the users. Both of these problems need to be carefully studied.

In this paper, we focus on the down-link OFDM scheduling and admission control problem in a single cell. Arriving packets get queued in the buffer at the base-station before they are transmitted to the users. For the scheduling problem the goal is to minimize the amount of time that any packet spends in the buffer at the base-station. In the literature, it is wellknown that the Max-Weight algorithm is throughput optimal under such a setting, in the sense that it can stabilize the system under the largest set of offered loads. However, it has been observed in [1] [2] [3] that the Max-Weight algorithm can lead to large delays for users. Specifically, although a system can be stabilized by the Max-Weight algorithm, the queue length can be very large. [1] [2] [3] proposed a number of new scheduling algorithms that are efficient in maintaining low queue lengths for all users. They keep queue length small by serving the queues with higher weighted sum and at the same time balancing queues in each time-slot. The authors of [1] [2] [3] use large-deviation tools to study the asymptotic decay-rate of the probability that the queue-length of any user exceed a given threshold, as the number of users and the number of frequency carriers both increase. They show that for Bernoulli arrivals that are i.i.d. across time, the proposed algorithm is rate-function queue-length optimal i.e., they achieve the largest decay-rate for the above queue overflow probability. For more general arrival processes, the algorithms are shown to achieve strictly positive decay-rates for the queue overflow probability

However, simply maintaining low queue-lengths is insufficient for guaranteeing low waiting-time. When the number of arrival packets is constant over time, one may map the decay rate of the queue-overflow probability to that of the delayviolation probability [4]. For general arrival processes, there may not exist such mappings. The discrepancy can be quite large especially when the arrivals are correlated over time. For example, a packet that is present in a queue with low queuelength may have to wait for a long time to get served if few packets are offered to this queue for several time-slots.

In this paper, we directly study the packet delay of OFDM down-link scheduling algorithms under an ON-OFF fading model. We use large-deviations tools and study the asymptotical decay-rate of the probability that the delay of any packet exceeds a threshold, as the number of users and the number of channels both increases. (The precise definition of the above delay-violation probability is given in Section II.) We provide a new OFDM scheduling algorithm and derive a lower bound on the rate-function of the delay-violation probability attained by this algorithm. We also obtain an upper-bound on the ratefunction of the delay-violation probability for any scheduling policy. From these bounds, we can identify the cases where the proposed algorithm is rate-function delay-optimal. Unlike [1], [2], [3], our result holds for a large class of general arrival processes, which may be correlated across time. We further use these bounds to derive a simple threshold policy for admission control. To the best of our knowledge, this is the first work that deals directly with the design and analysis of rate-function delay-optimal scheduling and admission control policies in OFDM wireless cellular systems.

When a large number of users are served by a singleserver queue with fixed capacity, it is easy to see that the delay-optimal policy should serve packets in a First-come First-serve manner. Previously, many-source large-deviations tools have been used to study the delay performance of Firstcome First-serve (FCFS) scheduling policy in such single server queues [5] [6]. Somewhat surprisingly, our analysis indicates that, when the number of users and the number of carriers are large, an OFDM system under ON-OFF fading behaves quite closely to a single-server queue with intermittent connectivity, provided that some conditions on the per-user transmission requirements are satisfied (see Lemma 6 in Section IV). Unfortunately, the FCFS policy no longer satisfies these conditions. Specifically, due to the random connectivity between queues and servers, we may not always be able to serve the set of packets with the highest delay in every timeslot. Hence, we must design a new policy, called DWM (delayweighted matching), that respects the conditions on the peruser transmission requirement.

In summary, the main contributions of this work are,

- We develop a scheduling algorithm, called DWM (delayweighted matching), and obtain a lower bound on the asymptotic decay-rate of its delay-violation probability. By comparing with a related upper bound on the ratefunction for any scheduling policy, we identify cases when the proposed DWM algorithm is rate-function delay-optimal. Our analysis holds for a large class of arrivals processes that may be correlated across time. Further, we develop a simple admission control policy based on the decay rate attained by the DWM algorithm.
- The key insight that emerges from our work is that OFDM systems with a large number of users and channels may be approximately modeled as a single-server queue with intermittent connectivity. While this insight considerably simplifies the analysis and design, its application requires careful consideration of the restrictions imposed by the random nature of channel capacity in a wireless OFDM system. We provide the analytical techniques that successfully address these issues.

II. SYSTEM MODEL

We model the down-link of a single cell in an OFDM system as a discrete-time multi-source multi-server system with stochastic connectivity. There are n frequency sub-carriers each of which is represented by a server. There are $m = \phi n$ users and the base-station maintains a queue for each user. For most of the analysis we will fix ϕ and study the delay violation

probability as m and n increase (defined below). Then, in section VIII when we study the admission control problem, we will vary ϕ and study the largest value of ϕ that can meet a given delay constraint. The arrival process to each queue is assumed to be stationary and ergodic, and i.i.d. across queues. However, arrivals may correlate across time. We assume that time is slotted. Let $a_u(i)$ denote the number of packets that arrive to queue u at time i, and let $a_u(i,j) = \sum_{k=i}^j a_u(k)$ denote the total arrivals to queue u from time i to j. We use \bar{a} to denote $\mathbb{E}[a_u(\cdot)]$. Let $A(i) = \sum_{u=1}^{m} a_u(i)$ denote the cumulative arrivals at time i and $A(i,j) = \sum_{k=i}^{j} A(k)$ denote the cumulative arrivals to all queues from time i to j. To model channel fading, the queue-server connectivity in time-slot i is given by the matrix $\mathbf{C}(i) = [c_{u,s}(i)]_{m \times n}$. We assume an ON-OFF model. When $c_{u,s}(i) = 1$ we say that the queue u is connected to the server s at time i. When $c_{u,s}(i) = 0$ we say that the queue u is disconnected from server s at time *i*. At every time-slot the resource manager or the scheduler at the base station allocates queues to servers. If a connected server s is allocated to a queue u in time-slot i, then one packet from u can be served by the end of the time-slot by server s. In a time-slot, multiple servers may be assigned to a single queue, but each server can be assigned only to one queue. For concreteness we assume that all arrivals occur at the beginning of a time-slot followed by any possible service. We also assume that the average arrival rate \bar{a} falls into the interior of the maximum stability region of the system [7], and hence the system can be made stationary and ergodic under some scheduling algorithm. Further, we assume that each queue has infinite buffer capacity so that no packets are ever dropped.



Fix a queue u at a time-slot i. Over all packets that are present in queue u at time i, let $D_u(i)$ denote the maximum delay starting from time i until all packets in this set are served, Note that if the packets of each queue u are served in a First-come First-serve (FCFS) manner and there is at least one packet that arrives to queue u at time i, then $D_u(i)$ is the maximum delay of all packets that arrive to queue u at time i. Further, this definition allows $D_u(i)$ to be well-defined even when there is no packet arriving to queue u at time i. Let $D(i) = \max_u \{D_u(i)\}$. Hence, D(i) > d if and only if there exists a packet that arrived on or before time i and that has not been served till time i + d. In this paper, we are interested in the delay performance in the large-system regime. Specifically, consider a sequence of systems with a fixed ϕ , but with both the number of users m and the number of servers n increasing proportionally to infinity. Assuming that the system is stationary and ergodic, let

$$I(d) = \lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P}\Big(D(0) > d\Big),$$

whenever the limit exists. I(d) is called the rate-function for delay threshold d, which captures the asymptotic decay-rate (as the system size increases) of the probability that D(i)exceeds the threshold d. One can imagine that a larger value of rate-function would imply a lower probability of packets getting delayed by d time-slots. In fact, for large n we can estimate $\mathbb{P}(D(0) > d) \approx e^{-nI(d)}$, and the estimate becomes better for increasing values of n. Our goal is then to design scheduling algorithms that achieve large values of the delay rate-function I(d). A policy is said to be rate-function delayoptimal if it achieves the maximum value of I(d) that any scheduling algorithm can achieve. Note that the above largen, fixed-d asymptotics are meaningful for the OFDM systems with a large number of users and carriers but requiring small delay.

Before we continue with the system model, we state a technical result that is often referred to in the rest of the paper.

Lemma 1: Let X_i , i = 1, 2, ... be a sequence of binary random variables such that for all i,

$$\mathbb{P}\Big(X_i = 1 \, \Big| \, X_{i'}, i' \neq i \Big) < c(n)e^{-nb},$$

regardless of the values of other random variables $X_{i'}, i' \neq i$, where c(n) is a polynomial in n of finite degree. Let N_1 be such that $c(n) < e^{\frac{nb}{2}}$ for all $n > N_1$. Then, for any 0 < a < 1,

$$\mathbb{P}\Big(\sum_{i=1}^{t} X_i > at\Big) < e^{-\frac{tnab}{3}}$$

for all $n > N := \max\{\frac{12}{ab}, N_1\}$. *Proof:* Please refer to Appendix A.

Additionally, we make the following technical assumptions about the arrival process and the channel states. Recall that $m = \phi n$ for a fixed ϕ .

Assumption 1: Arrivals are bounded. There exists $L < \infty$ such that $a_u(i) \leq L$ for any *i* and *u*.

Assumption 2: Given any $\epsilon > 0$ and $\delta > 0$, there exists T > 0, N > 0, and a positive function $I_B(\epsilon, \delta)$ independent of n and t such that

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{t} \mathbf{1}_{\{|A(i)-\bar{a}m| > \epsilon m\}}}{t} > \delta\Big) < e^{-mtI_B(\epsilon,\delta)}, \quad (1)$$

for all t > T and n > N. For each $\epsilon > 0$ and $\delta > 0$, let $T_B(\epsilon, \delta)$ and $N_B(\epsilon, \delta)$ be one corresponding set of values for such T and N, respectively.

Assumption 3: We assume that the channel process is *i.i.d.*, i.e.,

$$c_{u,s}(i) = \begin{cases} 1 & \text{with probability } q, \\ 0 & \text{with probability } 1 - q, \end{cases}$$

independently across i, u, s.

Remark: Assumption 1 is mild and it states that the arrivals in every time-slot must be bounded above by a finite number L. Assumption 3 has been used in previous work [1] [2] [3] [8]. Although the *i.i.d.* ON-OFF channel model is a simplification, we believe that the insights will also be useful for more general channel models. For instance, in Section VII we will also provide simulation results for Markovian channels. Assumption 2 is very general and captures a large class of arrival processes. The intuition behind Assumption 2 is a statistical multiplexing effect when a large number of sources are multiplexed. The basic tenet of the assumption is that the arrivals to different queues are independent of each other. Recall that the arrivals to every queue may vary around the mean value \bar{a} . In some time-slots, the arrivals to one queue may be higher or lower than \bar{a} . However, when considering a large number of such independent sources, one would expect that the sources with large arrivals would balance the sources with small arrivals so that the sum is close to $\bar{a}\phi n$. Hence, the chance that the sum is far away from the mean $\bar{a}\phi n$ is low, especially when n is large. Further, as long as the temporal correlation of arrivals is short-ranged, the chance that the total arrival over a large interval is far away from the mean also diminishes as the length of the time interval increases. Such intuitive properties are captured in Assumption 2. Specifically, the probability bound on the right hand side of (1) can be made arbitrarily small for sufficiently large nand t. The assumption can be mathematically verified for a large class of arrival processes. We provide here the proof for two special classes, i.e., i.i.d. arrivals and arrivals driven by two-state Markov chains.

Lemma 2: Let $a(\cdot)$ be a packet arrival process such that in every time-slot,

$$a(i) = \begin{cases} r & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

Note that in this case $\bar{a} = pr$. Then, given any $\epsilon > 0$, and $\delta > 0$, there exist T > 0, N > 0, and a positive function $I_B(\epsilon, \delta)$ independent of n and t such that

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{t} \mathbf{1}_{\{|A(i)-\bar{a}m| > \epsilon m\}}}{t} > \delta\Big) < e^{-mtI_B(\epsilon,\delta)}$$

for all t > T and n > N.

Proof: Let $\epsilon_2 = \frac{\epsilon}{r}$. Then, it is clear that, if at any time *i* the fraction of queues that receive *r* arrivals belongs to the interval $(p - \epsilon_2, p + \epsilon_2)$, then $|A(i) - \bar{a}m| < \epsilon m$. Moreover, the probability of this event is no smaller than $1-2e^{-m\min\{D_{KL}(p+\epsilon_2||p),D_{KL}(p-\epsilon_2||p)\}}$, where $D_{KL}(x||y) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}$, is the Kullback-Leibler divergence [9]. We define S(i) to be a sequence of random variables such that S(i) = 1 if $|A(i) - \bar{a}m| > \epsilon m$ and S(i) = 0 otherwise. Then, from Lemma 1 we know that there exists N > 0 such that

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{t} S(i)}{t} > \delta\Big) < e^{-mt\frac{\delta \min\{D_{KL}(p+\epsilon_2||p), D_{KL}(p-\epsilon_2||p)\}}{3}}$$

for all n > N and t > 0. The result then follows.

The next lemma shows that Assumption 2 also holds for an arrival process driven by a two-state Markov chain.

Lemma 3: Let $a(\cdot)$ be a packet arrival process driven by a Markov chain with two states 1 and 2. Assume that whenever the Markov chain is in state i, r_i packets are generated in each time-slot. State transitions occur at the end of time-slots. Suppose that the transition probability of the chain is given

by the matrix, $\begin{bmatrix} 1 - p_1 & p_1 \\ p_2 & 1 - p_2 \end{bmatrix}$. Note that in this case $\bar{a} = \frac{p_2}{p_1 + p_2} r_1 + \frac{p_1}{p_1 + p_2} r_2$. Then, given $\epsilon > 0$, and $\delta > 0$, there exists T, N and a positive

function $I_B(\epsilon, \delta)$ independent of n and t such that

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{t} \mathbf{1}_{\{|A(i) - \bar{a}m| > \epsilon m\}}}{t} > \delta\Big) < e^{-mtI_B(\epsilon, \delta)}$$

for all t > T and n > N.

Proof: Please refer to Appendix B.

and

B. Chernoff Bound and Cramer's Theorem

In the rest of the paper, we will frequently use the following standard results from Probability Theory in our proofs. Let $X_i, 1 \leq i \leq n$ be a sequence of i.i.d. random variables. For any $x > \mathbb{E}[X_i]$, the Chernoff bound states that

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i \ge nx\Big) \le e^{-n[\theta x - \lambda_{X_1}(\theta)]},$$

for any real number $\theta > 0$, (2)

where $\lambda_{X_i}(\theta) = \log \mathbb{E}[e^{\theta X_i}]$ is the cumulant-generating function of X_i . The best bound is obtained by choosing the real number $\theta = \theta^*$ that maximizes $\theta x - \lambda_{X_1}(\theta)$, assuming that θ^* exists. Cramer's Theorem states that the upper bound of (2) is tight in the exponent [9, Chapter 2], i.e.

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P}\Big(\sum_{i=1}^{n} X_i \ge nx\Big) = \theta^* x - \lambda_{X_1}(\theta^*).$$

Note that the cumulative arrivals in our system in any time interval -t + 1 to 0, i.e. $A(-t + 1, 0) = \sum_{u=1}^{m} a_u(-t + 1, 0),$ is just the sum of $m = \phi n$ i.i.d. random variables. Hence, using Cramer's Theorem we have, for any $x \ge 0$,

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P} \Big(A(-t+1,0) \ge n(t+x) \Big)$$
$$= \lim_{m \to \infty} -\frac{\phi}{m} \mathbb{P} \Big(A(-t+1,0) \ge \frac{m}{\phi}(t+x) \Big)$$
$$= \phi \sup_{\theta} \left[\theta \Big(\frac{t+x}{\phi} \Big) - \lambda_{a_u(-t+1,0)}(\theta) \right].$$

where $\lambda_{a_u(-t+1,0)}(\theta) = \log \mathbb{E}[e^{\theta a_u(-t+1,0)}]$ is the cumulantgenerating function of $a_u(-t+1,0)$. For a fixed ϕ , we define the quantity

$$I_A(t,x) := \phi \sup_{\theta} \left[\theta \left(\frac{t+x}{\phi} \right) - \lambda_{a_u(-t+1,0)}(\theta) \right].$$

This quantity is the rate function for the probability that in ttime-slots, the total number of arrivals is greater than or equal to nx + nt. The minimum of $I_A(t, x)$ taken over all positive integer values of t is defined as

$$I_A(x) := \inf_{t>0} I_A(t, x).$$

In our analysis, we will also use another quantity that is closely related to $I_A(x)$. Define

$$I_A^+(t,x) = \lim_{y \to x^+} I_A(t,y).$$

$$I_A^+(x) := \inf_{t>0} I_A^+(t, x).$$

Roughly speaking, $I_A^+(t,x)$ can be interpreted as the ratefunction for the probability that in t time-slots, the total number of arrivals is strictly greater than nx + nt. Clearly, for any value of x where $I_A(t, x)$ is continuous with respect to x, we must have $I_A^+(t,x) = I_A(t,x)$. However, there may be discontinuous points of x such that $I_A^+(t,x) \neq I_A(t,x)$. This potential difference may lead to a gap between the upper and lower bounds that we derive for the rate-function of the delay-violation probability. We will provide more details when we discuss this gap in Section V.

III. AN UPPER-BOUND ON THE RATE FUNCTION

In this section we derive an upper-bound on the rate function I(d) of the delay asymptote for all scheduling algorithms.

Theorem 1: Given the system model as described in Section II, under any scheduling algorithm,

$$\limsup_{n \to \infty} -\frac{1}{n} \log \mathbb{P}\Big(D(0) > d\Big)$$

$$\leq I_0^+ \triangleq \min\{(d+1)I_X, \min_{0 \le c \le d}\{I_A^+(d-c) + cI_X\}$$

where $I_X = \log(\frac{1}{1-q})$. *Proof:* We consider two events \mathcal{E}_1 and \mathcal{E}_2 that imply that D(0) > d.

Event \mathcal{E}_1 : Suppose that there is a packet that is present at queue 1 at time 0. Further, suppose that from time 0 to dqueue 1 is disconnected from all servers. Then, at the end of time-slot d this packet is still in the buffer, hence D(0) > d.

$$\mathbb{P}\Big(\mathcal{E}_1\Big) = (1-q)^{n(d+1)} = e^{-n(d+1)I_X}.$$
(3)

Event \mathcal{E}_2 : Consider the following sequence of events. Fix any $\epsilon > 0$. Choose t such that $I_A^+(t, d-c) < I_A^+(d-c) + \epsilon$. Further, there exists $\delta > 0$ such that $I_A(t, d-c+\delta) \leq I_A^+(t, d-c+\delta)$ $c) + \epsilon \leq I_A^+(d-c) + 2\epsilon$. Suppose that from time -t + 1 to 0 there are greater than or equal to $nt+n(d-c+\delta)$ arrivals to the system. Let the probability of this event be $p_{(d-c)}$. Then from Cramer's Theorem, we know that $\lim_{n\to\infty} -\frac{1}{n}\log p_{(d-c)} \leq$ $I_A(t, d-c+\delta) \leq I_A^+(d-c) + 2\epsilon$. The total service at any time cannot exceed n. Hence, at the end of time 0, there are at least n(d-c) + 1 packets in the buffer (as long as $n \ge \frac{1}{\delta}$). Moreover, at the end of time d - c the buffer must contain at least one packet that arrived before time 0. Without loss of generality, assume that this packet is present in queue 1. Now, assume that queue 1 remains disconnected from all servers in the next c time-slots. This occurs with probability $(1-q)^{cn} = e^{-nc\log\frac{1}{1-q}}$ independently of all past history. Hence, at the end of time d, there is still a packet that arrived before time 0 and that remains in queue 1. Hence, D(0) > d in this case. In other words, the probability $\mathbb{P}(D(0) > d)$ is no smaller than $p_{(d-c)}e^{-ncI_X}$. Since this is true for any $0 \le c \le d$, by taking c that maximizes the above quantity, we have,

$$\mathbb{P}\Big(\mathcal{E}_2\Big) \ge \max_{0 \le c \le d} \{p_{(d-c)}e^{-ncI_X}\}$$

Thus,

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P}\Big(\mathcal{E}_2\Big) \le \min_{0 \le c \le d} \{I_A^+(d-c) + 2\epsilon + cI_X\}.$$

Hence, by picking the more probable event from \mathcal{E}_1 and \mathcal{E}_2 , we have

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P} \Big(D(0) > d \Big) \\\leq \min \{ \min_{c \in \{0, 1...d\}} \{ I_A^+(d-c) + 2\epsilon + cI_X \}, (d+1)I_X \}.$$

As the above equation holds for all $\epsilon > 0$, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log \mathbb{P} \Big(D(0) > d \Big) \\\leq I_0^+ \triangleq \min\{\min_{c \in \{0, 1...d\}} \{ I_A^+(d-c) + cI_X \}, (d+1)I_X \}.$$

IV. VECTOR MATCHING IN BIPARTITE GRAPHS

In section V, we will propose new scheduling algorithms that attain close to the above upper bound I_0^+ for the decay rate. Towards this end, we first study a snapshot of the system at a given time, when each queue has some packets that need to be served. We will study under what conditions an allocation of servers/carriers to the users/queues can be found such that all desired packets can be served. As we will see below, this problem can be viewed as a vector matching problem in a random bipartite graph.

We first introduce some notations that will be used in this section. We use $G[X \cup Y, E]$ to denote a general bipartite graph, where X and Y are two disjoint sets of vertices and E is the set of edges such that every edge $e \in E$ connects a vertex in X to a vertex in Y. For our OFDM system, the set Xcorresponds to the set of m queues, and the set Y corresponds to the set of n servers. According to Assumption 3, each edge between a vertex u in X and a vertex y in Y exists with probability q, independently of other edges. Let $\partial_G(z)$ denote the set of neighbors of vertex z in G. Suppose that V is a set of vertices of G. We define $\partial_G(V) = \bigcup_{z \in V} \partial_G(z)$. $|\partial_G(V)|$ is called the degree of V and denotes the number of distinct neighbors of the vertices of V. If M is a subset of edges of G, i.e., $M \subset E$, then G(M) is called the sub-graph induced by M and consists of all vertices of G and edges present in M.

We review the concept of matching in a bipartite graph, which is well known in Graph Theory [10, Chapter 16]. Given a bipartite graph $G[X \cup Y, E]$, a matching M is a subset of edges such that in the induced sub-graph G(M) the degree of every vertex is at most one. A perfect matching is a matching such that in the induced sub-graph the degree of every vertex is exactly one. Unfortunately, this concept of matching is not very useful in our setting because, if a queue has more than one packets waiting to be served, we would have liked to allocate more than one servers to the queue.

In order to address the above issue, in this section we generalize this idea of matching to vector matching. Let $G[X \cup Y, E]$ be a bipartite graph where the vertices of set X are indexed as $\{x_1, x_2, ..., x_m\}$. Let v be a |X|-dimensional vector whose elements are non-negative integers. If V is a subset of X, then $v(V) = \sum_{\{i:x_i \in V\}} v_i$. Then, a v-matching M is a sub-set M of edges such that:

$$|\partial_{G(M)}(x_i)| \leq v_i$$
, for all $1 \leq i \leq |X|$, and,
 $|\partial_{G_{(M)}}(y)| \leq 1$, for all $y \in Y$.

In other words each vertex x_i in X is matched to at most v_i vertices in Y, but each vertex in Y is matched to at most one vertex in X. Note that a graph may have more than one v-matchings. A *perfect* v-matching is a v-matching (M) such that $|\partial_{G(M)}(x_i)| = v_i$ for all $x_i \in X$. If G admits a perfect v-matching, then it is said to be perfectly v-matched. For our OFDM system, v_i corresponds to the number of packets that queue x_i requests to serve. Thus, a perfect v-matching will correspond to a schedule of servers to queues that serves the requested packets from all queues. Since the edges appear randomly, we will be interested in the probability that a perfect v-matching can be found. The following two lemmas show some useful properties for estimating this probability.

Lemma 4: Let $G[X \cup Y, E]$ be a bipartite graph. Let v be a |X| dimensional vector whose components are non-negative integers. Then G has a perfect v-matching if and only if for every $V \subset X$, $|\partial_G(V)| \ge v(V)$.

Remark: If $v_i = 1$ for all *i*, then the above result is equivalent to the well known Hall's Marriage Theorem in Graph Theory [10]. For a detailed proof of the result please refer to Appendix C.

Lemma 5: Let $G[X \cup Y, E]$ be a random bipartite graph, in which for every pair of $x \in X$ and $y \in Y$ there is an edge between x and y with probability q, independently of other edges. Let v and w be vectors of length |X| with non-negative integer components such that,

1) $w_1 \ge w_2 + 2;$ 2) $v_1 = w_1 - 1; v_2 = w_2 + 1;$ 3) $v_i = w_i$, for $3 \le i \le |X|$.

Then,

$$\mathbb{P}(G \text{ has a perfect } v \text{-matching})$$
$$\geq \mathbb{P}(G \text{ has a perfect } w \text{-matching}).$$

Remark: Note that v and w are the same everywhere except in the first two components. Moreover, v is more balanced than w, i.e., $v_1 + v_2 = w_1 + w_2$ but $|v_1 - v_2| < |w_1 - w_2|$. The above result then states that, if a vector is more balanced, the probability of perfect vector matching is higher. This basic result forms the basis of the next Corollary. For a detailed proof please refer to Appendix D.

In the following results, H is a given positive integer independent of n.

Corollary 1: Let $G[X \cup Y, E]$ be a random bipartite graph in which for every pair of $x \in X$ and $y \in Y$ there is an edge between x and y with probability q, independently of other edges. Let v and w be two vectors of length |X| with non-negative integer components such that,

1)
$$\max_{i}[v_{i}] \leq H;$$

2) $\sum_{i=1}^{|X|} v_{i} \leq \sum_{i=1}^{|X|} w_{i} = n - H;$
3) $w_{i} = \begin{cases} n - (k+1)H, & \text{if } i = 1\\ H, & \text{if } 2 \leq i \leq k+1;\\ 0, & \text{if } i > k+1, \end{cases}$

where $k = \left|\frac{n}{H}\right| - 2$. Then,

 $\mathbb{P}\Big(G \text{ has a perfect } v\text{-matching}\Big)$ $\geq \mathbb{P}\Big(G \text{ has a perfect } w\text{-matching}\Big).$

Corollary 1 can be shown by converting w to v in a sequence of steps such that in every step the new vector becomes more balanced (in the sense of Lemma 5). The result then follows from Lemma 5. For the details of the proof please refer to Appendix E.

Lemma 6: Let $G[X \cup Y, E]$ be a random bipartite graph, in which for every pair of $x \in X$ and $y \in Y$ there is an edge between x and y with probability q, independently of all other edges. Let |X| = m and |Y| = n. Let w be a vector with non-negative integer components with $\sum_{i=1}^{m} w_i \leq n - H$ and $\max_{1 \leq i \leq m} w_i \leq H$, then for some finite value of N_X ,

$$\mathbb{P}\left(G \text{ has a perfect } w \text{-matching}\right) \geq 1 - \left(\frac{n}{1-q}\right)^{7H} e^{-n\log\frac{1}{1-q}}$$

for all $n > N_X$.

Proof: Please refer to Appendix F.

Remark: The decay rate stated in Lemma 6 can not be further improved. To see this, note that if $w_1 = H$, then the probability that the vertex x_1 is connected to less than H vertices in Y is at least $(1 - q)^{n-H+1}$. This event alone will imply that the decay rate of the probability of prefect w-matching cannot be larger than $\log \frac{1}{1-q}$.

Let X denote the set of m source queues, Y denote the set of n servers, and use an edge between u and s if queue u is connected to server s in the OFDM system. According to Lemma 6, as long as $\sum_{i=1}^{m} w_i \leq n-H$ and $\max_{1\leq i\leq m} w_i \leq$ H, with high probability an allocation of servers to queues can be found such that each user i will be able to serve w_i packets. Hence, our OFDM system is very similar to a single-server queue with service rate n-H and intermittent connectivity, provided that the service requirement of each user is bounded by H. In the sequel, we will use this insight to design scheduling algorithms with good delay performance. However, the additional constraint $\max_{1\leq i\leq m} w_i \leq H$ represents a key difference between our OFDM system and a single-server queue. In a single-server queue with service rate n - H, even if one user requests n - H packets and all other users request no packets, all packets can be served in one time-slot. In our OFDM system, however, since each user is connected to only qn servers on average, there usually exists no feasible schedule that can serve that above pattern in one time-slot. Hence, we must impose the additional constraint that $\max_{1 \le i \le m} w_i \le H$. Due to this difference, we will not be able to directly use an algorithm that is optimal for a single-server queue to serve our OFDM system. In the following sections, we will introduce new concepts that address this difficulty.

V. SCHEDULING POLICIES

In this section we propose two scheduling policies that can attain close to the upper bound I_0^+ (given in Theorem 1) on the asymptotic decay rate of the delay violation probability. We assume that in every time-slot the scheduler has perfect knowledge of queue-server connectivity, which is represented by the matrix C(i). Also, it can use the past history of arrival and channel processes.

A. Intuition behind the Proposed Delay-Based Policies

Motivated by Lemma 6, we consider a single-server queue with intermittent connectivity. Specifically, in every time-slot the server is connected with probability $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-nI_X}$, and disconnected otherwise. Whenever the server is connected it can serve $n_0 = n - H$ packets. However, it cannot serve any packets when disconnected. It is not difficult to see that, if we serve packets in such a single-server queue in a FCFS (First-Come First-Serve) manner, then the delay rate-function is optimal, i.e., it is equal to the upper bound I_0^+ given in Theorem 1.

Now from Lemma 6, our OFDM system is in fact quite similar to the single-server queue in the sense that, under suitable restrictions, the probability that n_0 packets may be served in a time-slot is no less than $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-nI_X}$. However, obviously we cannot use a FCFS policy directly, because it may violate the condition of Lemma 6, which translates to the restriction that in a time-slot every user can have no more than H packets to be served.

To circumvent the difficulty, we propose two policies FBS(h) and DWM. The policy FBS(h) approximates the FCFS policy, while respecting the restrictions mentioned above. We will show that there exists a value for the parameter h such that FBS(h) attains a rate-function close to I_0^+ . However, the FBS(h) policy is conservative in nature, and may waste capacity. Further, it may not be throughput-optimal for a finitesize system. Therefore, we propose another policy DWM, which is more aggressive in serving packets and does not waste capacity. We further show that DWM always serves packets ahead of FBS(h) for every arrival process and every value of h, and hence the delay rate-function for DWM must be no smaller than the delay rate-function for FBS(h). Thus, policy DWM can also attain a rate-function close to I_0^+ when the system size n approaches infinity. Further, it will perform much better in medium-sized systems.

B. Policy FBS(h) (Frame Based Scheduling)

This policy serves packets in units of frames. Suppose that a positive integer h is given. Recall that no more than L packets arrive to any user in a time-slot. Let $n_0 = n - Lh$ be the capacity of each frame. In the policy FBS(h) each frame is composed of packets that satisfy the following two conditions:

- 1) The number of packets in the frame is no greater than n_0 (i.e., the capacity of a frame);
- The difference of arrival times of any two packets in the frame must be no larger than h.

As packets arrive in each time-slot, the frames are constructed by filling in the packets sequentially. Specifically, packets belonging to queue 1 are filled before packets belonging to queue 2, and so on. Further, older packets are added before newer packets. We fill each frame until the above conditions cannot be maintained. Then we start a new frame. There might be a frame that is only partially filled at the end of a time-slot. In the next time-slot this frame is filled first, before starting a new frame.

A frame in general may be represented as a vector in \mathbb{Z}^m , where the *u*-th component of the vector represents the number of packets of user *u* in the frame. The policy FBS(*h*) serves the frames in the same order as they are constructed. Further, at most one frame is served in a time-slot. Specifically, let v(i) denote the vector representing the head-of-line frame at time *i*. From the construction of the frame described above and Assumption 1 on the boundedness of the arrival process, we have $\max_{1 \le u \le m} v_u(i) < Lh$. Moreover, $\sum_{u=1}^m v_u(i) \le n_0 = n - Lh$ for all *i*. Note that a frame might contain less than n_0 packets if it is the only frame left or if it was full because of condition 2 (described earlier).

In each time-slot *i* the policy FBS(*h*) tries to schedule the head-of-line frame v(i) for transmission. Let H = hL. We know from Section IV on vector matching that, with probability $1 - \left(\frac{n}{1-q}\right)^{TH} e^{-nI_X}$, the scheduler can transmit the whole frame in a given time-slot. If the policy FBS(*h*) cannot transfer the whole frame, then no packets are scheduled in this time-slot and the scheduler will try again in the next time-slot. Define the random variable $X_F(i) = 1$ if v(i) is successfully transmitted at time *i*, and $X_F(i) = 0$, otherwise.

The following theorem shows that there exists a value of h such that policy FBS(h) attains a rate-function close to I_0^+ .

Theorem 2: If the arrival process satisfies Assumptions 1 and 2 and the channel process satisfies Assumptions 3, then, there exists a value of h for which the scheduling policy FBS(h) obtains the following rate-function

$$\lim_{n \to \infty} \inf \frac{-1}{n} \log \mathbb{P} \Big(D(0) > d \Big)$$

$$\geq I_0 \triangleq \min \{ \min_{c \in \{0,1,\dots,d\}} I_A(d-c) + cI_X, (d+1)I_X \}.$$

Section VI will be devoted to prove Theorem 2. We will also comment on the potential gap between I_0 and the upper bound I_0^+ shortly in Section V-D.

C. Policy DWM (Delay Weighted Matching)

Although policy FBS(h) attains a rate function close to the upper bound I_0^+ in the asymptotic regime when the system size n increases to infinity, it is clearly inefficient. Specifically, policy FBS(h) may not serve any packet in a time-slot, and may waste up to Lh packets in a time-slot even if it serves a frame. Further, it is not throughput-optimal for any finite-size system. As a result, policy FBS(h) may perform poorly when the system size is not very large. In addition, for Theorem 2 to hold, we need to know the value of h in advance. Such a value of h depends on the statistics of the arrivals and channel states, which may be difficult to predict in practice. Next, we propose another policy, called DWM (Delay Weighted Matching), that addresses the above difficulty.

In every time-slot, define the waiting time of every packet as the time that the packet has spent in the buffer. We assign a weight to every packet as follows. If a packet has a waiting time of W and belongs to the queue with index u, then its weight is $W + \frac{u}{n+1}$. Next, construct a bipartite graph $G[X \cup Y, E]$ such that vertices in X correspond to the oldest n packets of every queue and Y is the set of servers. The edge set E is constructed as follows: if u is connected to s, then all vertices that correspond to packets of u are connected to s. The packets to transmit are then determined by a maximum-weight matching algorithm. In the following Lemma we compare policies DWM and FBS(h).

Lemma 7: For any given sample path and for any value of h, by the end of time-slot i, Policy DWM has served every packet that FBS(h) has served.

Proof: Please refer to Appendix G. By Lemma 7, the rate-function of DWM must be no smaller than that of FBS(h). Combinging Theorem 2 with Lemma 7, we conclude that DWM also attains a rate-function close to I_0^+ when the system size n approaches infinity. Further, we would expect DWM to outperform FBS(h) even if the system size is not very large. Note that DWM does not require the value of h in advance, and hence can be readily used even if we do not have prior statistical knowledge of the arrivals and channel states. Finally, the throughput optimality of DWM for finite system-size n can be shown analogously to other maxweight algorithms. We refer to the readers to our more recent work [11] for details.

D. The Gap between I_0 and I_0^+

For a fixed $d \ge 0$, let $I_0^+(d)$ (correspondingly, $I_0(d)$) denote the upper bound (correspondingly, lower bound) given in Theorem 1 (correspondingly, Theorem 2). In other words,

$$I_0^+(d) \triangleq \min\{\min_{c \in \{0,1,\dots,d\}} I_A^+(d-c) + cI_X, (d+1)I_X\}, \quad (4)$$

$$I_0(d) \triangleq \min\{\min_{c \in \{0,1,\dots,d\}} I_A(d-c) + cI_X, (d+1)I_X\}.$$
 (5)

Clearly, they are virtually of the same form except that $I_0^+(d)$ is computed from $I_A^+(\cdot)$ while $I_0(d)$ is computed from $I_A(\cdot)$.

Recall from Section II-B that

$$I_A(x) := \inf_{t>0} I_A(t, x), \ I_A^+(x) := \inf_{t>0} I_A^+(t, x), \ \text{and} \ I_A^+(t, x) = \lim_{y \to x^+} I_A(t, y).$$

Thus, $I_A^+(t, x)$ and $I_A(t, x)$ differ only at the point x where $I_A(t, x)$ is discontinuous with respect to x. Suppose that for some $x_0 > 0$ there is a non-zero probability that

$$a_u(-t+1,0) \ge \frac{t+x_0}{\phi}.$$

Then, the rate function $I_A(t, x)$ must be finite, convex and increasing with respect to x in the range $[0, x_0]$, and hence it must be continuous in $[0, x_0]$. The only possibility of discontinuity is at the right end-point of an interval $[0, x_0]$ such that

$$\mathbb{P}[a_u(-t+1,0) \ge \frac{t+x_0}{\phi}] > 0, \text{ and} \\ \mathbb{P}[a_u(-t+1,0) > \frac{t+x_0}{\phi}] = 0,$$

in which case $I_A(t, x)$ is finite and $I_A^+(t, x) = +\infty$. Thus, we would expect that $I_0(d)$ and $I_0^+(d)$ are identical in most cases. Specifically, suppose that the value of c_0 that attains the minimum for $I_0(d)$ in (5) is such that $I_A(d-c_0) = I_A(t_0, d-c_0)$, and that the function $I_A(t_0, \cdot)$ is continuous at $x = d - c_0$. Then, we must have $I_0(d) = I_0^+(d)$. The following example illustrates one such case.

Example 1: Consider *i.i.d.* arrivals with *maximum* number of packets that arrive in any time slot equal to L. Suppose that $L\phi > 1$ and $d < L\phi - 1$. For any $t \ge 1$ and $0 \le c \le d$, let $x = d-c < L\phi - 1$. Under such a scenario, there exists a non-zero probability that $a_u(-t+1,0)$ can reach $Lt = \frac{t+(L\phi-1)t}{\phi} > \frac{t+x}{\phi}$. Hence, we must have $I_A(t, d-c) = I_A^+(t, d-c)$ for all t and c, and thus $I_0(d) = I_0^+(d)$.

On the other hand, the following example illustrates a case when the two bounds do not meet.

Example 2: Consider *i.i.d.* Bernoulli arrivals such that in each time-slot either one packet arrives to a queue with probability \bar{a} or no packets arrive to the queue with probability $1-\bar{a}$. (Note that L = 1 in this example.) Suppose that $\phi = 1$ and d = 0. Then, we have $I_A(t, 0) = \log 1/\bar{a}$ and $I_A^+(t, 0) = +\infty$. We can verify that in this case $I_0(1) = \min\{\log 1/\bar{a}, I_X\}$ and $I_0^+(1) = I_X$.

Even in the case when $I_0(d) \neq I_0^+(d)$, the following lemma shows that they cannot be too far away from each other.

Lemma 8: For any $d \ge 1$, we must have $I_0(d) \ge I_0^+(d-1)$. *Proof:* Note that

$$I_0^+(d-1) \triangleq \min\{\min_{c \in \{0,1,\dots,d-1\}} I_A^+(d-1-c) + cI_X, dI_X\}.$$
 (6)

For any $t \ge 1$ and $c \in \{0, 1..., d-1\}$, since $I_A(t, x)$ is increasing in x, we have $I_A^+(t, d-1-c) \le I_A(t, d-c)$. Taking infimum over all t, we have $I_A^+(d-1-c) \le I_A(d-c)$. Comparing (5) and (6) term-by-term, we then have

$$I_0(d) \geq \min\{\min_{c \in \{0,1,\dots,d-1\}} I_A(d-c) + cI_X, dI_X\} \\ \geq I_0(d-1).$$

VI. ANALYSIS OF FBS(h)

In this section, we will prove Theorem 2 for policy FBS(h). (By Lemma 7, the same conclusion will then also hold for policy DWM.) We start with a set of equations that capture the frame dynamics in FBS(h). Define F(i) as the number of unserved frames in buffer at time i. Then, we can write a recursive equation for F(i):

$$F(i) = \max\{F(i-1) + \left\lceil \frac{A(i) - R(i-1)}{n_0} \right\rceil - X_F(i), 0\}.$$

$$Z(i) = \begin{cases} [Z(i-1) + 1] \text{mod}(h), & \text{if } A(i) < R(i-1) \\ 1, & \text{if } A(i) > R(i-1) \\ 0, & \text{if } A(i) = R(i-1). \end{cases}$$

$$R(i) = \mathbf{1}_{\{F(i)>0\}} \mathbf{1}_{\{Z(i)>0\}} [R(i-1) - A(i)] \text{mod}(n_0). \quad (7)$$

To explain this set of equations, recall that after each timeslot, the end-of-line frame may be only partially filled and thus can be filled with new arrivals in the next time-slot. We use R(i) to represent the remaining available space in the end-of-line partially-filled frame at the end of time *i*. Hence, A(i) - R(i-1)represents the number of new frames that are n_0 created at time i. Note that if $A(i) \leq R(i-1)$, i.e., the number of arrivals at time *i* is less than the remaining available space in the end-of-line frame at the end of time i - 1, then no new frame is added. $X_F(i)$ represents the number of frames served in time-slot *i*. Notice that a maximum of one frame and hence n_0 packets can be transmitted in a time-slot. The variable Z(i) counts the number of time-slots for which the end-of-line frame has been open. It starts at 1 when a new frame is opened, i.e., when A(i) > R(i-1). Then it is incremented by 1 every time when the number of arrivals A(i) is less than R(i-1). If it reaches h, then this frame is completed and a new frame is started, in which case Z(i) = 0and R(i) = 0. Let v be an m-dimensional vector whose i^{th} component represents the number of packets of queue i in a frame. The construction of the frames ensures that for every frame $v_i \leq hL = H$ and $\sum_{i=1}^{m} v_i \leq n - H = n_0$. Hence, from Lemma 6 we have that $X_F(i) = 1$ with probability no smaller than $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-nI_X}$ in every time-slot.

Let $R_0 = R(i-1)$ be the empty space in the end-of-line frame at the end of time i-1. Further, let $A_F^{R_0}(i,k)$ denote the number of new frames created from time i to k, including any partially-filled frame at time k but excluding any partiallyfilled frame at time i. We use the notation $A_F(i,k)$ to denote $A_F^{R_0}(i,k)$, if $R_0 = 0$. Hence we can write,

$$F(k) = F(i-1) + A_F^{R_0}(i,k) - X_F(i,k),$$

where $X_F(i, k)$ denotes the total number of frames departing from the buffer in the time interval *i* to *k*. That is,

$$X_F(i,k) = \sum_{j=i}^k X_F(j) \mathbb{1}_{\{F(j)>0\}}.$$

In general the equations in (7) are complicated to analyze. However, if the arrival process satisfies some special conditions in a time interval (i, k), then we can derive some useful results as follows.

Lemma 9: Let $A(\cdot)$ be an arrival process to the system. Let R_0 be the empty space in the end-of-line frame at the end of time i-1. Let the arrivals in the interval from i to k be such that

- 1) The buffer never becomes empty in the interval, i.e., F(j) > 0 for all $j \in \{i, i + 1, ..., k\}$.
- 2) For any h − 1 consecutive time-slots in the interval, the cumulative arrivals are greater than or equal to n₀, i.e., ∑_{j=x}^{x+h-2} A(j) ≥ n₀, for any x ∈ {i, i+1, ..., k-h+2}.

 Then the following holds for policy FBS(h),

$$A_F^{R_0}(i,k) = \left\lceil \frac{A(i,k) - R_0}{n_0} \right\rceil,$$

$$R(k) = [R_0 - A(i,k)] \operatorname{mod}(n_0).$$
(8)

Remark: The condition of the Lemma implies that every frame has exactly n_0 packets. The result then follows. For details, please refer to Appendix H.

Corollary 2: Let $A(\cdot)$ be an arrival process such that F(j) > 0 for all $i \le j \le k$ and let $B = \{x_1, ..., x_{|B|}\}$ be a sequence of time-slots in increasing order, belonging to the interval from *i* to *k*, such that in every interval $(x_i + 1, x_{i+1})$, $i \in \{1, 2..., |B| - 1\}$, the condition 2 of Lemma 9 is satisfied. Then,

$$A_F^{R_0}(x_1+1,x_{|B|}) \le \sum_{j=1}^{|B|-1} \left\lceil \frac{A(x_j+1,x_{j+1})}{n_0} \right\rceil$$

Proof: From Lemma 9 we have that

$$A_F^{R_0}(x_1+1,x_{|B|}) = \sum_{j=1}^{|B|-1} \left\lceil \frac{A(x_j+1,x_{j+1}) - R(x_j)}{n_0} \right\rceil.$$

Since $R(\cdot) \ge 0$, it follows that,

$$A_F^{R_0}(x_1+1,x_{|B|}) \le \sum_{j=1}^{|B|-1} \left\lceil \frac{A(x_j+1,x_{j+1})}{n_0} \right\rceil.$$

We are now ready to prove Theorem 2.

Proof of Theorem 2: We first choose the value of h based on the statistics of the arrival process. Let the mean of the arrival process be \bar{a} . We fix $\delta < \frac{2}{3}$ and $\epsilon < \frac{\bar{a}}{2}$. Then, from Assumption 2 on the arrival process, there exists a positive function $I_B(\epsilon, \delta)$, such that for all $n > N_B(\epsilon, \delta)$ and $t > T_B(\epsilon, \delta)$ we have,

$$\mathbb{P}\Big(\frac{\sum_{i=j+1}^{j+t} \mathbf{1}_{\{|A(i)-\bar{a}n|>\epsilon n\}}}{t} > \delta\Big) < e^{-ntI_B(\epsilon,\delta)},$$

for any integer j.

Recall that I_0 is defined in the statement of Theorem 2. We then choose

$$h = \max\left\{T_B(\epsilon, \delta), \left\lceil \frac{1}{\phi(\bar{a} - \epsilon)(1 - \frac{3\delta}{2})} \right\rceil, \left\lceil \frac{2I_0}{\phi I_B(\epsilon, \delta)} \right\rceil\right\} + 1$$
(9)

The reason for choosing such a value of h will become clear later on. Recall that L is the maximum number of packets that can arrive to a queue at any time-slot i and H = hL. Thus, His the maximum number of packets that can arrive to a queue in h time-slots.

Let L(0) be the last time -t before 0 such that the buffer was empty, i.e., D(-t) = 0. Then given that L(0) = -t, the event D(0) > d occurs if and only if the number of frames that arrive in the time interval from -t+1 to 0 is greater than the total number of frames that could be served in -t+1 to d. That is,

$$\{ D(0) > d, L(0) = -t \}$$

= $\{ L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0 \}$

By taking the union over all possible values of L(0) we get,

$$\mathbb{P}\Big(D(0) > d\Big) \le \sum_{t=1}^{\infty} \mathbb{P}\Big(L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0\Big).$$

We now fix any $1 > \hat{p} > \bar{a}$. Then define,

$$t^* := \max\left\{ T_B(\hat{p} - \bar{a}, \frac{1 - \hat{p}}{6(L+2)}), \left\lceil \frac{6}{1 - \hat{p}} \right\rceil, \\ \left\lceil \frac{I_0}{\min\{I_B(\hat{p} - \bar{a}, \frac{1 - \hat{p}}{6(L+2)}), (\frac{1 - \hat{p}}{9})I_X\}} \right\rceil \right\}$$
(10)

and split the summation as,

$$\mathbb{P}\Big(D(0) > d\Big) \\
\leq \sum_{t=1}^{t^*} \mathbb{P}\Big(L(0) = -t, A_F(-t+1,0) - X_F(-t+1,d) > 0\Big) \\
+ \sum_{t=t^*}^{\infty} \mathbb{P}\Big(L(0) = -t, A_F(-t+1,0) - X_F(-t+1,d) > 0\Big). \tag{11}$$

We divide the proof into two parts. In Part 1 we prove that there exists $N_1 > 0$ such that for all $n > N_1$

$$\sum_{t=1}^{t^*} \mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big)$$

$$< c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} e^{-nI_0},$$

where c_1, c_2 are positive constants independent of t and n. Then, in Part 2 we prove that there exists $N_2 > 0$ such that for all $n > N_2$ $\sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big)$ $< 4e^{-nI_0}$

Finally, by substituting both parts into equation (11), we have that there exists $N := \max\{N_1, N_2\}$ such that for all n > N,

$$\sum_{t=1}^{\infty} \mathbb{P}\Big(D(0) > d, L(0) = -t\Big)$$
$$\leq (c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} + 4)e^{-nI_0}.$$

By taking logarithm and limit as n tends to infinity, we get the desired result.

Part 1: Let us denote by \mathcal{E}_t^{α} the set of sample paths in which every h-1 time-slots in the interval -t+1 to 0 see at least n arrivals. Let \mathcal{E}_t^{β} be the set of sample paths in which $\frac{A(-t+1,0)}{n_0} - \sum_{j=-t+1}^d X_F(j) > 0$. Let \mathcal{E}_t be the sample paths such that L(0) = -t and D(0) > d. Then, the following can be shown,

$$\mathcal{E}_t \subset (\mathcal{E}_t^{\alpha})^c \cup \mathcal{E}_t^{\beta}.$$
 (12)

To see this, observe that \mathcal{E}_t is the set of sample paths in which L(0) = -t and $A_F(-t+1,0) - X_F(-t+1,d) > 0$. For all sample paths in the set $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, Lemma 9 holds and hence, $A_F(-t+1,0) = \left\lceil \frac{A(-t+1,0)}{n_0} \right\rceil$. Moreover, it is easy to observe that, for all sample paths in the set $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, $X_F(-t+1,d) = \sum_{j=-t+1}^d X_F(j)$. Hence, for a sample path belonging to $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, we must have $\frac{A(-t+1,0)}{n_0} - \sum_{j=-t+1}^d X_F(j) > 0$. This implies that, $\mathcal{E}_t \cap \mathcal{E}_t^{\alpha} \subset \mathcal{E}_t^{\beta}$. Thus we have, $\mathcal{E}_t = (\mathcal{E}_t \cap \mathcal{E}_t^{\alpha}) \cup (\mathcal{E}_t \cap (\mathcal{E}_t^{\alpha})^c) \subset (\mathcal{E}_t^{\beta} \cup (\mathcal{E}_t^{\alpha})^c)$. Hence, (12) holds. It then follows that,

$$\mathbb{P}(\mathcal{E}_t) \le \mathbb{P}((\mathcal{E}_t^{\alpha})^c) + \mathbb{P}(\mathcal{E}_t^{\beta}).$$
(13)

We now give the intuition behind the analysis of \mathcal{E}_t^{α} and \mathcal{E}_t^{β} . For a detailed proof, please refer to Appendix I.

We note that the event \mathcal{E}_t^{α} implies that every frame formed in the interval from -t + 1 to 0 will have n_0 packets, i.e. all frames served are completely full. It is then obvious that $\mathbb{P}(\mathcal{E}_t^{\alpha})$ depends on h, i.e., it will be large if we increase the maximum time for which any frame can remain open. By choosing an h large enough we can ensure that the probability $\mathbb{P}((\mathcal{E}_t^{\alpha})^c)$ is arbitrarily small. In particular, we can ensure that the rate-function of $\mathbb{P}((\mathcal{E}_t^{\alpha})^c)$ is greater than the rate-function of $\mathbb{P}(\mathcal{E}_t^{\beta})$. The cost that needs to be paid for having a large his the loss in frame-size, which is $n_0 = n - Lh$. Nonetheless, this decrease in frame-size is independent of n and does not affect the performance of the system significantly for large n. Hence, it does not show up in the rate-function. Specifically, for the choice of h in (9), it can be shown that there exists $N_3, c_3 > 0$ such that we have

$$\mathbb{P}(\mathcal{E}_t^{\alpha}) > 1 - c_3 t e^{-nI_0}.$$
(14)

for all $n > N_3$.

It can be seen that the event \mathcal{E}_t^{β} is similar to the buffer overflow event in a single-server queue with intermittent connectivity as described earlier. Recall that as opposed to a single-server queue with constant rate, in every time-slot, with probability approximately $1 - e^{-nI_X}$ the service is equal to n_0 packets, i.e., one frame. Thus, now there can be two factors responsible for \mathcal{E}_t^{β} . Firstly, if the arrival process is bursty, then \mathcal{E}_t^{β} can be caused by a large burst of arrivals in a few timeslots. Secondly, if q is small \mathcal{E}_t^{β} can be caused by a time interval of low service as frames get piled up in the buffer. For moderate values of q, one can expect that the most likely way in which \mathcal{E}_t^{β} occurs is a mixture of bursty arrivals and sluggish service. From large deviations theory we know that the rate-function of \mathcal{E}_t^{β} is determined by the probability of the most likely sample path leading to \mathcal{E}_t^{β} . More formally, it can be shown that there exists $c_2, c_4, N_4 > 0$ such that

$$\mathbb{P}\left(\mathcal{E}_{t}^{\beta}\right) = \mathbb{P}\left(\frac{A(-t+1,0)}{n_{0}} - \sum_{j=-t+1}^{d} X_{F}(j) > 0\right) \\
= \sum_{a=-t}^{d} \mathbb{P}\left(\sum_{j=-t+1}^{d} X_{F}(j) = t+a\right) \mathbb{P}\left(A(-t+1,0) > (t+a)n_{0}\right) \\
\leq (t+d+1) \max_{-t \leq a \leq d} \left\{\mathbb{P}\left(\sum_{j=-t+1}^{d} X_{F}(j) = t+a\right) \\
\times \mathbb{P}\left(A(-t+1,0) > (t+a)n_{0}\right)\right\} \\
\leq c_{4}2^{c_{2}t}\left(\frac{n}{1-q}\right)^{7H} \\
\times e^{-n\min\{(d+1)I_{X},\min_{a \in \{0,1,2...d\}}\{I_{A}(a)+(d-a)I_{X}\}\}} \\
\leq c_{4}2^{c_{2}t}\left(\frac{n}{1-q}\right)^{7H}e^{-nI_{0}},$$
(15)

for all $n > N_4$.

I

Let $c_1 = 2 \max\{c_3, c_4\}$. Substituting (14) and (15) into (13) we then have

$${}^{\mathbb{D}}(\mathcal{E}_t) \le c_3 t e^{-nI_0} + c_4 2^{c_2 t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_0} \\ \le c_1 2^{c_2 t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_0}$$

for all $n > N_1 = \max\{N_3, N_4\}$. Finally, summing over t = 1 to t^* we have,

$$\sum_{t=1}^{t^*} \mathbb{P}\Big(D(0) > d, L(0) = -t\Big) = \sum_{t=1}^{t^*} \mathbb{P}\Big(\mathcal{E}_t\Big)$$
$$\leq c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} e^{-nI_0},$$

for all $n > N_1$.

Part 2: We would like to show that there exists $N_2 > 0$ such that for $n > N_2$

$$\sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t,0) - X_F(-t,d) > 0\Big) < 4e^{-nI_0}$$

We noted earlier that the equations for evolution of $A_F(-t+$ (1,0) are in general complicated. However, if an arrival process satisfies certain conditions then some simple results such as Lemma 9 and Corollary 2 can be obtained. Hence, to analyze $A_F(-t+1,0)$ we first construct an arrival process $\hat{A}(\cdot)$ that satisfies the conditions of Lemma 9 and $\hat{A}_F(-t+1,0) >$ $A_F(-t+1,0)$. We do this by adding some extra arrivals to the process $A(\cdot)$ in some strategic time-slots. The resulting arrival process $\hat{A}(\cdot)$ has the property that $\hat{A}(i) = \hat{p}n$ whenever $A(i) \leq \hat{p}n$ and $\hat{A}(i) = Ln$ whenever $A(i) > \hat{p}n$. (Please refer to Appendix I for the details of how to construct $\hat{A}(\cdot)$.) Hence, the resulting arrival process $\hat{A}(\cdot)$ is in fact very simple. We now get an upper bound on $\hat{A}_F(-t+1,0)$, which, by construction, is also an upper bound on $A_F(-t+1, 0)$.

Let $B = \{b_1, b_2, b_{|B|}\}$ be the set of time-slots in the interval -t + 1 to 0 when $A(i) \ge \hat{p}n$. Then, from Corollary 2 we have that, given L(0) = -t,

$$\begin{split} \hat{A}_{F}(-t+1,0) \\ &\leq \sum_{j=1}^{|B|-1} \left\lceil \frac{\hat{A}(b_{j}+1,b_{j+1}-1)}{n_{0}} \right\rceil + \sum_{j=1}^{|B|} \left\lceil \frac{\hat{A}(b_{j},b_{j})}{n_{0}} \right\rceil \\ &+ \left\lceil \frac{\hat{A}(-t+1,b_{1}-1)}{n_{0}} \right\rceil + \left\lceil \frac{\hat{A}(b_{|B|}+1,0)}{n_{0}} \right\rceil \\ &\leq \frac{n}{n_{0}} [\hat{p}t + (L+2)|B| + 1]. \end{split}$$

From Assumption 2 on the arrival process we know that for large enough n and t, |B| can be made less than an arbitrarily small fraction of t. Further, we can show that for $|B| < \frac{1-\hat{p}}{6(L+2)}t$, $n > \frac{H(2+\hat{p})}{1-\hat{p}}$ and $t > \frac{6}{1-\hat{p}}$, $A_F(-t+1,0) \leq \hat{A}_F(-t+1,0) \leq (\frac{2+\hat{p}}{3})t$. (Please refer to Appendix I for details.) Hence,

$$\mathbb{P}\left(A_{F}(-t+1,0) \ge (\frac{2+\hat{p}}{3})t, L(0) = -t\right) \\
\le \mathbb{P}\left(|B| > \frac{1-\hat{p}}{6(L+2)}t\right) \\
\le e^{-ntI_{B}(\hat{p}-\bar{a},\frac{1-\hat{p}}{6(L+2)})},$$
(16)

for all $n > N_5 = \max\{N_B(\hat{p} - \bar{a}, \frac{1-\hat{p}}{6(L+2)}), \frac{H(2+\hat{p})}{1-\hat{p}}\}$ and $t > T_1 = \max\{T_B(\hat{p} - \bar{a}, \frac{1-\hat{p}}{6(L+2)}), \frac{6}{1-\hat{p}}\}$. Moreover, we know that for each $i, X_F(i) = 1$ with probability greater than $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-nI_X}$ for all $n > N_X$. Hence, using Lemma 1 we have that, there exists $N_6 > N_X$ such that,

$$\mathbb{P}\Big(X_F(-t+1,d) < (\frac{2+\hat{p}}{3})t, L(0) = -t\Big) \\
\leq \mathbb{P}\Big(X_F(-t+1,d) \le (\frac{2+\hat{p}}{3})(t+d), L(0) = -t\Big) \\
\leq e^{-n(t+d)(\frac{1-\hat{p}}{9})I_X} \\
\leq e^{-nt\frac{(1-\hat{p})I_X}{9}},$$
(17)

for all $n > N_6$ and t > 0.

Combining the above two results, from (16) and (17) we have, for all $n > N_7 = \max\{N_5, N_6\}$ and $t > T_1$,

$$\mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big) \\
\leq 1 - (1 - e^{-nt\frac{(1+\hat{p})I_X}{9}})(1 - e^{-ntI_B(\hat{p}-\bar{a},\frac{1-\hat{p}}{6(L+2)})}) \\
\leq 2e^{-tnI_{BX}},$$

where I_{BX} is the minimum of $I_B(\hat{p}-\bar{a},\frac{1-\hat{p}}{6(L+2)})$ and $\frac{(1-\hat{p})I_X}{9}$. Recall that $t^* > \max\{T_1, \frac{I_0}{I_{BX}}\}$. Hence, summing over all

 $t > t^*$ we have, for all $n > N_2 = \max\{N_7, \left|\frac{\log 2}{I_{BX}}\right|\}$

$$\sum_{t=t^{*}}^{\infty} \mathbb{P}\Big(A_{F}(-t,0) - X_{F}(-t,d) > 0, L(0) = -t\Big)$$

$$\leq \sum_{t=t^{*}}^{\infty} \mathbb{P}\Big(A_{F}(-t,0) - X_{F}(-t,d) > 0, L(0) = -t\Big)$$

$$\leq \sum_{t=t^{*}}^{\infty} 2e^{-ntI_{BX}}$$

$$\leq \frac{2e^{-nt^{*}I_{BX}}}{1 - e^{-nI_{BX}}}$$

$$\leq 4e^{-nt^{*}I_{BX}} \quad (\text{as } e^{-nI_{BX}} < \frac{1}{2})$$

$$\leq 4e^{-nI_{0}}.$$

The result of the theorem then follows.

VII. SIMULATION RESULTS

In this section, we compare the performance of the proposed DWM algorithm with the classic Max-Weight (MW) algorithm [7] and the recently-proposed Server-Side-Greedy (SSG) algorithm in [1], [2]. We simulate these algorithms and compare the empirical probabilities that the maximum delay at any given time exceeds a constant d. We consider two settings: (i) when the arrivals are *i.i.d.* across time-slots, and (ii) when arrivals are correlated across time-slots.

In the first setting, the arrivals to every queue are given by the following distribution:

$$a(i) = \begin{cases} 5 & \text{with probability } 0.167, \\ 0 & \text{with probability } 0.833, \end{cases}$$

independently for all time-slots *i*. We run the MW, SSG and DWM algorithms for a system with n = 30 users and m = 30carriers/servers (and hence $\phi = 1$). The user-server connection probability is q = 0.75, so that the system is stable but heavily loaded, i.e. greater than 83.5% of the maximum load. We run the simulation for 10^5 time-slots.

In the second setting, we consider arrivals that are driven by a Markov chain with two states. When the Markov chain is in state 1, 5 packets are generated in each-time slot, and when the chain is in state 2, no packets are generated. Further, state transitions occur at the end of time-slots. The transition probability of the chain is given by the matrix $\begin{bmatrix} .5 & .5 \\ .1 & .9 \end{bmatrix}$. Note that the probability in state 1 is equal to 0.167. Hence, the Markovian arrivals have the same average rate as the *i.i.d.*



Fig. 2. Performance of DWM, MW and SSG for *i.i.d* connectivity. n = 30, q = 0.75, *i.i.d*. and 2-state Markov chain (m.c.) driven arrivals.

arrivals. The user-server connection probability is chosen as q = 0.75. We also consider a system with n = 30 and m = 30 and run the simulation for 10^5 time-slots.

The results are summarized in Fig. 2. As can be seen from the plot, the proposed DWM algorithm performs consistently better than the Max-Weight algorithm and the SSG algorithm. The delay of the Max-Weight algorithm does not go down substantially with increasing d. The queue-length-based SSG performs better than Max-Weight. However, DWM performs even better than SSG, and further reduces the delay-violation probability by orders-of-magnitude. Note that SSG is designed to minimize the queue overflow probability [1], [2]. Our result thus illustrates that small queue length may not always lead to small delay, e.g., if the packets in that queue is not served for a long time. Since DWM directly treats delay, the performance is significantly better. Finally, even though our analytical results focus on the asymptotic limit of large n, for such a medium-sized system (with n = 30 users and m = 30 carriers/servers) the proposed DWM policy already outperforms existing approaches significantly.

Recall that our analytical results require that the channel is *i.i.d.* across time. Nonetheless, we expect that the key insights from our analysis will also be useful under more general settings. Next, we experiment with a setting in which the userserver connectivity for each channel is correlated across time. Specifically, the connectivity is driven by a Markov chain with two states. When the Markov chain is in state 1, the user is connected to that server and when it is in state 2 the user is disonnected from the server. The transition probability of the $\begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix}$. We run the simulation chain is given by the matrix for both the i.i.d. arrival process and the time-correlated arrival process as described earlier. The results are summarized in Fig. 3. We can observe very similar trends for the relative performance of the DWM, SSG, and MW policies as in Fig. 2. In particular, we observe that DWM achieves much better delay performance even for Markovian-correlated channels.

The above results demonstrate how the delay-violation



Fig. 3. Performance of DWM, MW and SSG for Markov-chain driven connectivity. n = 30, *i.i.d.* and 2-state Markov chain (m.c.) driven arrivals.

probability varies with d. Since our analytical results study the asymptotes as n increases, we next plot the probability of delay-violation with increasing n for DWM, and compare it with another fictitious policy called SSQ. Here, SSQ (Single-Server Queue) represents a fictitious system with a single server of capacity n that is fed with the aggregated arrivals of all users. Hence, it provides a lower bound on the deadline violation probability of other algorithms. In these simulations, we take $\phi = 1$ and the arrivals to every queue are given by the following distribution:

$$a(i) = \begin{cases} 5 & \text{with probability } 0.14, \\ 0 & \text{with probability } 0.86, \end{cases}$$

independently for all time-slots *i*. We consider user-server connectivity that is *i.i.d.* across time, users and servers, with probability q = 0.5. The results are reported in Fig. 4 for both d = 1 and d = 2. It can be seen in Fig. 4 that the rate-function obtained from the theoretical result, i.e., Theorem 2, matches well with the simulation result: the slopes of the different curves (for the same value of d) are almost identical, except that the theoretical result is shifted above the empirical curve by a constant amount. This can be attributed to the o(1) terms not captured by I_0 in Theorem 2. Further, the probability of delay-violation under DWM is well approximated by that under SSQ over the entire range of n (from n = 30 to n = 60). Hence, this result confirms our intuition (see Section V-A) that the OFDM system behaves quite similar to a single-server queue under suitable assumptions.

VIII. ADMISSION CONTROL

From Theorem 2, we know the expression for the lower bound on the rate-function achieved by DWM for any value of ϕ . We now consider an OFDM system with a fixed number of channels *n*, but the number of users $m = \phi n$ can vary (i.e., ϕ can vary). Write $I_0(d, \phi)$ as the lower bound I_0 defined in Theorem 2, which now depends on both *d* and ϕ . For large *n*,



Fig. 4. Delay-violation probability versus n for DWM and SSQ. q = 0.5 and i.i.d. arrivals.



Fig. 5. Rate function versus i.i.d. arrivals with R = 5 with varying p

the probability of delay-violation can be approximated by

$$\mathbb{P}\Big(D(0) > d\Big) \approx e^{-nI_0(d,\phi)}.$$
(18)

Given a fixed bandwidth n and a delay-violation constraint of the form $\mathbb{P}(D(0) > d) < \epsilon$, we can use (18) to estimate the number of users (ϕn) that may be admitted in a cell such that

$$e^{-nI_0(d,\phi)} < \epsilon$$

 $\implies I_0(d,\phi) \ge -\frac{\ln \epsilon}{n}.$ (19)

By plotting $I_0(d, \phi)$ versus ϕ , we can then obtain the maximum value of ϕ that satisfies (19).

Example: Consider a OFMDA system with n = 40 orthogonal channels and a delay constraint given by $\mathbb{P}(D(0) > 4 \text{ time-slots}) < e^{-40}$. Thus, we need the decay rate $I_0(4, \phi) \ge -\frac{\ln\epsilon}{n} = 1$.

Consider the following scenario with *i.i.d.* arrivals of the following distribution:

$$a(i) = \begin{cases} 5 & \text{with probability } p, \\ 0 & \text{with probability } 0, \end{cases}$$

where the values of p may be varied. The user-channel connectivity is *i.i.d.* across users and time, and each user-channel pair is on with probability 0.5. In Figure 5, the rate

function $I(d, \phi)$ is plotted as a function of ϕ for d = 4 and different values of p.

From Figure 5, it can be seen that for p = 0.18 the number of users that can be accommodated in the cell is roughly 0.58n = 23, where as for p = 0.10, roughly 0.95n = 38 users may be accommodated in the cell. These values can then be used by the network provider to perform admission control decision and/or provision the network resources.

IX. CONCLUSION

We consider the scheduling problem of the down-link of an OFDM system for supporting a large number of delaysensitive users. Assuming an ON-OFF channel model, we show that when the scale of the system is large, the OFDM system can be approximated by a Single-Server Queue with intermittent connectivity. Inspired by this observation, we first construct the Frame Based Scheduling (FBS(h)) policy that emulates the single-serve queue by accounting for the restrictions placed by the wireless channel in an OFDM system. We then prove that, for a large class of arrival processes, there exists a value of h for which FBS(h) attains a close-tooptimal rate-function for the delay violation probability when the system size approaches infinity. Since FBS(h) may waste capacity and the suitable value of h depends on the arrival process, we then design the Delay Weighted Matching (DWM) scheduling algorithm, which also achieves a close-to-optimal rate-function for the delay-violation probability, independently of the arrival process. Further, the DWM algorithm achieves high throughput and thus performs well even in medium-sized systems. Our simulations indicate that DWM can significantly improve the performance compared to the state-of-art algorithms in the literature. We further show that the analytical results for DWM can be used to determine a simple threshold for admission control.

There are many interesting directions for future work. First, we plan to use the insight gained from DWM to design scheduling algorithms for more general channel models. Second, although DWM achieves close-to-optimal rate-function, it may have a high computational complexity. It would be worthwhile to consider scheduling algorithms that achieve good delay bounds and are of lower complexity. Based on the results in this paper, the more recent work in [11], [12] has studied certain low-complexity algorithms of this type. We refer the readers to these studies for the latest development. Third, in this work we consider the case when all users have similar arrival patterns, channel conditions, and delay requirements. It would be interesting to see how the DWM algorithm can be extended to users with different arrival patterns and channel conditions and with diverse delay requirements.

ACKNOWLEDGMENT

This work has been partially supported by NSF Award CNS-0643145 and a grant from the Purdue Research Foundation. The authors would like to thank Dr. Bo Ji and Prof. Ness B. Shroff at the Ohio State University for stimulating discussions and valuable suggestions on this paper.

References

- S. Bodas, S. Shakkotai, L. Ying, and R. Srikant, "Scheduling in Multi-Channel Wireless Networks: Rate Function Optimality in the Small Buffer Regime," in *Proceedings of ACM SIGMETRICS*, 2009.
- [2] —, "Low-complexity scheduling algorithms for multichannel downlink wireless networks," in *Proceedings of IEEE INFOCOM*, San Diego, CA, March 2010.
- [3] —, "Scheduling for Small Delay in Multi-rate Multi-channel Wireless Networks," in *Proceedings of IEEE INFOCOM*, Shanghai, China, April 2011.
- [4] V. J. Venkataramanan and X. Lin, "On Wireless Scheduling Algorithms for Minimizing the Queue-Overflow Probability," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 788–801, June 2010.
- [5] C. Courcoubetis and R. Weber, "Buffer Overflow Asymptotics for a Buffer Handling Many Traffic Sources," *Journal of Applied Probability*, vol. 33, pp. 886–903, 1996.
- [6] S. Shakkottai and R. Srikant, "Many-sources delay asymptotics with applications to priority queues," in *Proceedings of ACM SIGMETRICS*, 2001, pp. 183–200.
- [7] L. Tassiulas, R. Tassiulas, and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, pp. 466–478, 1993.
- [8] S. Kittipiyakul and T. Javidi, "Delay-Optimal Server Allocation in Multiqueue Multiserver Systems with Time-Varying Connectivities," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 2319– 2333, May 2009.
- [9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applica*tions, 2nd ed. New York: Springer-Verlag, 1998.
- [10] J. Bondy and U. Murty, *Graph Theory*. New York: Springer-Verlag, 2008.
- [11] B. Ji, G. R. Gupta, X. Lin, and N. B. Shroff, "Low-Complexity Scheduling Policies for Achieving Throughput and Delay Optimality in Multi-Channel Wireless Networks," *submitted to IEEE/ACM Transactions on Networking*, 2012, also available at http://arxiv.org/abs/1301.3598.
- [12] —, "Performance of Low-Complexity Greedy Scheduling Policies in Multi-Channel Wireless Networks: Optimal Throughput and Near-Optimal Delay," in *Proceedings of IEEE INFOCOM*, Turin, Italy, April 2013.

APPENDIX A

Proof of Lemma 1: From the choice of $N \ge N_1$ we have, that for all n > N and i,

$$\mathbb{P}\Big(X_i = 1 \left| X_{i'}, i' \neq i \right) < e^{-\frac{nb}{2}},$$

regardless of the values of other random variables $X_{i'}, i' \neq i$. Applying Chernoff bound, we have,

$$\mathbb{P}\left(\sum_{i=1}^{t} X_i > at\right) < e^{-tD_{KL}\left(a\right||e^{-\frac{nb}{2}}\right)}.$$

Now for all n > N, we have

$$D_{KL}(a||e^{-\frac{nb}{2}})$$

$$= a \log\left[\frac{a}{e^{-\frac{nb}{2}}}\right] + (1-a) \log\left[\frac{1-a}{1-e^{-\frac{nb}{2}}}\right]$$

$$= \frac{abn}{2} + a \log a + (1-a) \log\left[\frac{1-a}{1-e^{-\frac{nb}{2}}}\right]$$

$$\geq \frac{abn}{2} + a \log a + (1-a) \log(1-a)$$

$$\geq \frac{abn}{2} - 2 \qquad (\text{as } x \log\frac{1}{x} \le \frac{1}{e} \text{ for all } 1 > x > 0)$$

$$\geq \frac{abn}{3} \qquad (\text{as } n > \frac{12}{ab}).$$
Hence, $\mathbb{P}\left(\sum_{i=1}^{t} X_i > at\right) < e^{-\frac{tnab}{3}} \text{ for all } n > N.$

APPENDIX B

Proof of Lemma 3: Without loss of generality, we state the proof for the case when m = n. By applying the balance equations we have that the expected fraction of time-slots that the chain spends in states 1 and 2 is given by $\pi_1^* = \frac{p_2}{p_1+p_2}$ and $\pi_2^* = \frac{p_1}{p_1+p_2}$ respectively. Hence the average rate \bar{a} for the chain is given by $\pi_1^*r_1 + \pi_2^*r_2$.

Now consider *n* independent copies of this arrival process driven by Markov chains. We denote by $\pi_1(i)$, the fraction of chains in state 1 at time *i*, and we denote by $\pi_2(i)$ the fraction of chains in state 2 at time *i*. Then, the packet arrivals at time *i* is given by $A(i) = n(r_1\pi_1(i) + r_2\pi_2(i))$. Note that at any time $\pi_1(i) + \pi_2(i) = 1$. Hence, A(i) can be written as a function of $\pi_1(i)$ alone. For convenience we denote the interval $[x - \rho, x + \rho]$ on the real line by $\mathcal{B}(x, \rho)$.

It is easy to see that, if $r_1 = r_2$, then $A(i) = \bar{a}n$ for all i. The result then follows trivially. We consider here the case when $r_1 \neq r_2$. Let $\epsilon_1 = \frac{\epsilon}{|r_1 - r_2|}$. It can be seen that whenever $\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1)$ then $A(i) \in \mathcal{B}(n\bar{a}, n\epsilon)$. It suffices to show that given $\epsilon_1 = \frac{\epsilon}{|r_1 - r_2|} > 0$ and $\delta > 0$, there exists N, T > 0 and a positive function $I_{B_1}(\epsilon_1, \delta) = I_B(\epsilon, \delta)$ independent of n and t such that

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{t} \mathbf{1}_{\{\pi_1(i) \notin \mathcal{B}(\pi^*, \epsilon_1)\}}}{t} > \delta\Big) \le e^{-ntI_{B_1}(\epsilon_1, \delta)}$$

for all n > N and t > T.

We divide the proof into two parts. In Part 1, we show that there exists T_1 such that, irrespective of the starting state $\pi_1(j)$, with high probability $\pi(j + T_1) \in \mathcal{B}(\pi_1^*, \epsilon_1)$. More precisely, there exists a T_1 and a positive rate-function $I_1(\epsilon_1)$ such that,

$$\mathbb{P}\Big(\pi_1(j+T_1) \in \mathcal{B}(\pi_1^*,\epsilon_1) \Big| \pi_1(j)\Big) \ge 1 - 4T_1 e^{-nI_1(\epsilon_1)}.$$
(20)

Then, in Part 2 we prove that at any time j, if $\pi_1(j)$ belongs to $\mathcal{B}(\pi_1^*, \epsilon_1)$, then with high probability, for the next several time-slots the system remains in $\mathcal{B}(\pi_1^*, \epsilon_1)$. More precisely, there exists a positive rate-function $I_2(\epsilon_1)$ such that,

$$\mathbb{P}\Big(\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1) \ \forall \ j < i \le j + T_2 \Big| \pi_1(j) \in \mathcal{B}(\pi_1^*, \epsilon_1) \Big) \\
\ge 1 - 4T_2 e^{-nI_2(\epsilon_1)}$$
(21)

for any positive integer T_2 .

Before we prove Part 1 and Part 2, we show why they imply the result of the lemma. To this end, we choose T_2 such that $\frac{T_1}{T_1+T_2} < \frac{\delta}{2}$ and define $T_3 := T_1 + T_2$. Now, combining the above equations, we have that, irrespective of the state of the system at time j,

$$\mathbb{P}\Big(\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1) \ \forall \ j + T_1 \le i \le j + T_3\Big)$$

$$\ge 1 - 4T_3 e^{-nI_3(\epsilon_1)}, \tag{22}$$

where $I_3(\epsilon_1) = \min\{I_1(\epsilon_1), I_2(\epsilon_1)\}$. We divide the interval 1 to t into consecutive frames consisting of T_3 time-slots each. Index these frames by k. Note that if $(t \mod T_3) \neq 0$, then the last frame may contain less than T_3 time-slots. Define $K := \left[\frac{t}{T_3}\right] - 1$. Define the sequence S(k) as follows:

$$S(k) = \begin{cases} 0, & \text{if } \frac{\sum_{j=kT_3+1}^{kT_3+T_3} \mathbf{1}_{\{\pi_1(j)\in\mathcal{B}(\pi^*,\epsilon_1)\}}}{T_3} \ge 1 - \frac{\delta}{2}, \\ 1, & \text{otherwise.} \end{cases}$$

Then, for every value of $0 \le k \le K - 1$ (i.e. for every frame except the last), we have

$$\mathbb{P}\Big(S(k) = 1 \Big| S(k'), k' \neq k\Big) \le 4T_3 e^{-nI_3(\epsilon_1)},$$

regardless of what happens in all other frames $k' \neq k$. Using Lemma 1 in Section II-B, we know that there exists N such that

$$\mathbb{P}\Big(\sum_{k=0}^{K-1} S(k) > \frac{\delta}{2} K\Big) \le e^{-n\frac{\delta K I_3(\epsilon_1)}{6}},$$

for all n > N. Note that for all $t > 3T_3$, we have $K > \frac{2t}{3T_3}$. Hence,

$$\mathbb{P}\Big(\sum_{k=0}^{K-1} S(k) < \frac{\delta}{2}K\Big) \ge 1 - e^{-nt\frac{\delta I_3(\epsilon_1)}{9T_3}},$$

for all n > N and $t > 3T_3$. Note that the event $\sum_{k=1}^{K} S(k) \le \frac{\delta}{2}K$ implies that,

$$\begin{split} \sum_{j=1}^{t} \mathbf{1}_{\{\pi_{1}(j)\in\mathcal{B}(\pi^{*},\epsilon)\}} \\ &\geq (1-\frac{\delta}{2})KT_{3}(1-\frac{\delta}{2}) \\ &\geq (1-\frac{\delta}{2})^{2}(t-T_{3}) \\ &= (1-\delta)t + \frac{\delta^{2}}{4}t - T_{3}(1-\frac{\delta}{2})^{2} \\ &\geq (1-\delta)t \quad \text{(by choosing } t > \frac{T_{3}(1-\frac{\delta}{2})^{2}}{\frac{\delta^{2}}{4}} \text{).} \end{split}$$

Hence, from the above two equations, we have, for $t>T=\max\{\frac{4(1-\frac{\delta}{2})^2}{\delta^2}T_3,3T_3\}$ and n>N

$$\mathbb{P}\Big(\frac{\sum_{j=1}^{t} \mathbf{1}_{\{\pi_1(j) \notin \mathcal{B}(\pi_1^*, \epsilon_1)\}}}{t} \ge \delta\Big) \le e^{-nt \frac{\delta I_3(\epsilon_1)}{9T}},$$

where $\frac{I_3(\epsilon_1)}{9T}$ is a positive function of δ, ϵ, p_1 and p_2 . The result of the lemma then follows.

We now prove Part 1 and Part 2.

Part 1: We define p_m as

$$p_m := \frac{\min\{p_1, 1 - p_1, p_2, 1 - p_2\}}{4}.$$

Take $\epsilon_2 = \min\{\frac{(p_1+p_2)\epsilon_1}{3}, p_m\}$. We define the function $D_m(x)$ as

$$D_m(x) = \min \left\{ D_{KL}(p_1 + x||p_1), D_{KL}(p_2 + x||p_2), \\ D_{KL}(p_1 - x||p_1), \\ D_{KL}(p_2 - x||p_2) \right\},$$

where $D_{KL}(x||y)$ denotes the Kullback Leibler divergence [9] and is given by the formula $D_{KL}(x||y) = x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y}$. Take T_1 such that $|(1 - p_1 - p_2)|^{T_1-1} < \frac{\epsilon_1}{3}$. We use $\pi_{1\to 2}(i)$ to denote the fraction of chains that transition from state 1 to 2 at the end of time-slot *i*. Similarly, we define $\pi_{2\to 1}(i)$ to be the fraction of chains that transition from state 2 to state 1 at the end of time-slot *i*.

By applying the Chernoff bound to the number of chains in state 1 and 2, we have,

$$\mathbb{P}\left(\pi_{1\to2}(i) > \pi_{1}(i)(p_{1} + \frac{\epsilon_{2}}{2}) \middle| \pi_{1}(i)\right) \\
\leq e^{-n\pi_{1}(i)D_{KL}(p_{1} + \frac{\epsilon_{2}}{2}||p_{1})}, \\
\mathbb{P}\left(\pi_{1\to2}(i) < \pi_{1}(i)(p_{1} - \frac{\epsilon_{2}}{2}) \middle| \pi_{1}(i)\right) \\
< e^{-n\pi_{1}(i)D_{KL}(p_{1} - \frac{\epsilon_{2}}{2}||p_{1})}.$$

However, as $\pi_1(i) < 1$ the above equations imply that

$$\mathbb{P}\left(\pi_{1\to2}(i) > \pi_{1}(i)p_{1} + \frac{\epsilon_{2}}{2} \middle| \pi_{1}(i)\right) \\
\leq e^{-n\pi_{1}(i)D_{KL}(p_{1} + \frac{\epsilon_{2}}{2}||p_{1})}, \\
\mathbb{P}\left(\pi_{1\to2}(i) < \pi_{1}(i)p_{1} - \frac{\epsilon_{2}}{2} \middle| \pi_{1}(i)\right) \\
\leq e^{-n\pi_{1}(i)D_{KL}(p_{1} - \frac{\epsilon_{2}}{2}||p_{1})}.$$

Combining the above equations we get,

$$\mathbb{P}\Big(\pi_{1\to2}(i) \in \mathcal{B}(\pi_1(i)p_1, \frac{\epsilon_2}{2}) \Big| \pi_1(i) \Big) \ge 1 - 2e^{-n\pi_1(i)D_m(\frac{\epsilon_2}{2})}.$$
(23)

Similarly,

$$\mathbb{P}\Big(\pi_{2\to 1}(i) \in \mathcal{B}(\pi_2(i)p_2, \frac{\epsilon_2}{2}) | \pi_1(i) \Big) \ge 1 - 2e^{-n\pi_2(i)D_m(\frac{\epsilon_2}{2})}.$$
(24)

Combining (23) and (24) we have

$$\mathbb{P}\Big(\pi_1(i+1) \in \mathcal{B}(\pi_1(i)(1-p_1) + \pi_2(i)p_2, \epsilon_2) \Big| \pi_1(i) \Big) \\
\geq 1 - 4e^{-n \min\{\pi_1(i), \pi_2(i)\}D_m(\frac{\epsilon_2}{2})}.$$
(25)

We now use (23)-(25) to prove the claim (20) of Part 1. We first show that, irrespective of the state of the system at time 0,

$$\mathbb{P}\Big(\pi_1(1) \in [p_m, 1-p_m] \Big| \pi_1(0)\Big) \ge 1 - 4e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$

To see this, note that there can be two cases: (1) $\pi_1(0) \notin [p_m, 1-p_m]$ and (2) $\pi_1(0) \in [p_m, 1-p_m]$.

Case 1: Assume that $\pi_1(0) < p_m$ (The case when $\pi_1(0) > 1 - p_m$ leads to similar analysis). Note that $\pi_2(0) > 1 - p_m$ as $\pi_1(0) + \pi_2(0) = 1$. From (24) we have

$$\begin{split} \mathbb{P}\Big(\pi_{2\to 1}(0) \in \mathcal{B}(\pi_2(0)p_2, \epsilon_2) \Big| \pi_1(0) < p_m \Big) \\ \geq 1 - 2e^{-n\pi_2 D_m(\epsilon_2)} \\ \geq 1 - 2e^{-np_m D_m(\epsilon_2)} \quad (\text{because } \pi_2(0) > p_m). \end{split}$$

This implies that,

$$\mathbb{P}\Big(\pi_1(1) \in (\pi_2(i)p_2 - \epsilon_2, \pi_2(0)p_2 + \epsilon_2 + p_m) \Big| \pi_1(0) < p_m\Big) \\ \ge 1 - 2e^{-np_m D_m(\epsilon_2)}.$$

It can be seen that $\pi_1(0) < p_m$ implies

$$\pi_2(0)p_2 - \epsilon_2$$

$$\geq (1 - p_m)p_2 - \epsilon_2$$

$$\geq (1 - p_m)(4p_m) - \epsilon_2$$

$$\geq p_m,$$

and

$$\pi_2(0)p_2 + \epsilon_2 + p_m$$

$$\leq p_2 + \epsilon_2 + p_m$$

$$\leq 1 - 4p_m + \epsilon_2 + p_m$$

$$\leq 1 - p_m.$$

Hence, the above implies that,

$$\mathbb{P}\Big(\pi_1(1) \in [p_m, 1 - p_m] \Big| \pi_1(0) < p_m\Big)$$

$$\geq 1 - 2e^{-np_m D_m(\epsilon_2)}.$$
(20)

Similarly, it can be shown that,

$$\mathbb{P}\Big(\pi_1(1) \in [p_m, 1 - p_m] \Big| \pi_1(0) > 1 - p_m\Big) \\
\geq 1 - 2e^{-np_m D_m(\epsilon_2)}.$$
(27)

Case 2: $\pi_1(0) \in [p_m, 1 - p_m]$. Note that in this case $\min\{\pi_1(0), \pi_2(0)\} > p_m$. Hence, applying (25), we have

$$\mathbb{P}\Big(\pi_1(1) \in \mathcal{B}(\pi_1(0)(1-p_1) + \pi_2(0)p_2, \epsilon_2) \\ \Big| \pi_1(0) \in [p_m, 1-p_m] \Big) \\ \geq 1 - 4e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$

Further, notice that for $\pi_1(0) \in [p_m, 1 - p_m]$ we have

$$\pi_1(0)(1-p_1) + \pi_2(0)p_2 - \epsilon_2$$

$$\geq 4p_m - \epsilon_2$$

$$\geq p_m,$$

and

$$\pi_1(0)(1-p_1) + \pi_2(0)p_2 + \epsilon_2 \\ \leq 1 - 4p_m + \epsilon_2 \\ \leq 1 - p_m.$$

It follows from the above equations that

$$\mathbb{P}\Big(\pi_1(1) \in [p_m, 1 - p_m] \Big| \pi_1(0) \in [p_m, 1 - p_m] \Big) \\\geq 1 - 4e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$
(28)

Thus, combining (26), (27) from Case 1 and (28) from Case 2, we have

$$\mathbb{P}\Big(\pi_1(1) \in [p_m, 1 - p_m] \,\Big| \pi_1(0)\Big) \ge 1 - 4e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$
 (29)

Now that we have analyzed $\pi_1(1)$, we proceed to study $\pi_1(i), i = 2, ..., T_1$. In the above case 2, we observed that

$$\pi_1(i) \in [p_m, 1 - p_m] \Rightarrow \mathcal{B}(\pi_1(i)(1 - p_1) + \pi_2(i)p_2, \epsilon_2) \in [p_m, 1 - p_m].$$
(30)

Moreover, since $\pi_1(i) + \pi_2(i) = 1$, we have

$$\pi_1(i)(1-p_1) + \pi_2(i)p_2 = \pi_1(i)(1-p_1-p_2) + p_2.$$
 (31)

Then, from equations (25) and (31), we have,

$$\mathbb{P}\Big(\pi_1(i) \in \mathcal{B}(\pi_1(i-1)(1-p_1-p_2)+p_2,\epsilon_2) \\ \Big| \pi_1(i) \in [p_m, 1-p_m] \Big) \ge 1 - 4e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$

Using (30) and recursively applying the above equation for $T_1 - 1$ consecutive time-slots, we have,

$$\mathbb{P}\Big(\pi_1(i) \in \mathcal{B}(\pi_1(i-1)(1-p_1-p_2)+p_2 \\ \text{for all } 2 \le i \le T_1, \epsilon_2) \Big| \pi_1(1) \in [p_m, 1-p_m] \Big) \\ \ge 1 - 4(T_1-1)e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$

6) Moreover, note that if $\pi_1(i) \in \mathcal{B}(\pi_1(i-1)(1-p_1-p_2)+p_2, \epsilon_2)$ for all $2 \le i \le T_1$ and $\pi_1(1) \in [p_m, 1-p_m]$ then,

$$\begin{aligned} \pi_1(T_1) \\ &\in \mathcal{B}\left(\pi_1(1)(1-p_1-p_2)^{T_1-1}+p_2\sum_{j=0}^{T_1-2}(1-p_1-p_2)^j, \\ &\epsilon_2\sum_{j=0}^{T_1-2}(1-p_1-p_2)^j\right) \\ &= \mathcal{B}\left(\frac{p_2}{p_1+p_2}+\pi_1(1)(1-p_1-p_2)^{T_1-1} \\ &-\frac{p_2(1-p_1-p_2)^{T_1-1}}{p_1+p_2}, \frac{\epsilon_2}{p_1+p_2}\left[1-(1-p_1-p_2)^{T_1-1}\right]\right) \\ &\subset \mathcal{B}(\frac{p_2}{p_1+p_2},\epsilon_1) \quad \text{(from the choice of } T_1 \text{ and } \epsilon_2) \\ &= \mathcal{B}(\pi_1^*,\epsilon_1). \end{aligned}$$

Hence, from the above two equations we have

$$\mathbb{P}\Big(\pi_1(T_1) \in \mathcal{B}(\pi_1^*, \epsilon_1) \Big| \pi_1(1) \in [p_m, 1 - p_m] \Big) \\\geq 1 - 4(T_1 - 1) e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$
(32)

Now combining (29) and (32) we have,

$$\mathbb{P}\Big(\pi_1(T_1) \in \mathcal{B}(\pi_1^*, \epsilon_1) \Big| \pi_1(0) \Big) \ge 1 - 4T_1 e^{-np_m D_m(\frac{\epsilon_2}{2})}.$$
(33)

Part 2: Recall that, we wish to show that if $\pi_1(j) \in \mathcal{B}(\pi_1^*, \epsilon_1)$ then, with high probability, the system remains in $\mathcal{B}(\pi_1^*, \epsilon_1)$ for the next T_2 time-slots. Choose $\epsilon_3 = \epsilon_1 \min\{p_1 + p_2, 2 - p_1 - p_2\}$ We know from (25) that,

$$\mathbb{P}\Big(\pi_1(i+1) \in \mathcal{B}(\pi_1(i)(1-p_1) + \pi_2(i)p_2, \epsilon_3)\Big) \\ \geq 1 - 4e^{-n\min\{\pi_1(i), \pi_2(i)\}D_m(\frac{\epsilon_3}{2})}.$$

Note that if $\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1)$ then $\mathcal{B}(\pi_1(i)(1-p_1) + \pi_2(i)p_2, \epsilon_3) \subset \mathcal{B}(\pi_1^*, \epsilon_1)$. To see this, observe that if $\pi_1(i) \in$

 $\mathcal{B}(\pi_1^*, \epsilon_1)$ then,

$$\pi_1(i)(1-p_1) + \pi_2(i)p_2 - \epsilon_3 \\ \ge \pi_1^* - \epsilon_1 |1-p_1-p_2| - \epsilon_3 \\ \ge \pi_1^* - \epsilon_1,$$

and further,

$$\pi_1(i)(1-p_1) + \pi_2(i)p_2 + \epsilon_3 \\ \leq \pi_1^* + \epsilon_1 |1-p_1-p_2| + \epsilon_3 \\ \leq \pi_1^* + \epsilon_1.$$

Hence, if $\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1)$, then $\mathcal{B}(\pi_1(i)(1 - p_1) + \pi_2(i)p_2, \epsilon_3) \in \mathcal{B}(\pi_1^*, \epsilon_1)$. Now, using the above result and (25), we have

$$\mathbb{P}\Big(\pi_1(i+1) \in \mathcal{B}(\pi_1^*, \epsilon_1) \,\Big| \,\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1)\Big) \ge 1 - 4e^{-n\pi_m D_m(\frac{\epsilon_3}{2})} \text{Similarly},$$

where $\pi_m = \min{\{\pi_1^* - \epsilon_1, \pi_2^* - \epsilon_1\}}$. Now, recursively applying the equation above for T_2 consecutive time-slots, we have

$$\mathbb{P}\Big(\pi_1(i) \in \mathcal{B}(\pi_1^*, \epsilon_1) \ \forall \ j < i \le j + T_2 \Big| \pi_1(j) \in \mathcal{B}(\pi_1^*, \epsilon_1) \Big)$$
$$\ge 1 - 4T_2 e^{-n\pi_m D_m(\frac{\epsilon_3}{2})}. \tag{34}$$

APPENDIX C

Proof of Lemma 4: We first show that, if there exists a set $V' \subset X$ such that $|\partial_G(V')| < v(V')$, then G does not have a perfect v-matching. For contradiction suppose that M is a perfect v-matching in G. Let G(M) be the sub-graph induced by M. Then

$$|\partial_{G(M)}(V')| = \sum_{x \in V'} v(x) = v(V').$$

However, as G(M) is a sub-graph of G, the above equation implies that $|\partial_G(V')| \ge v(V')$, which contradicts our assumption on V'.

We now prove that, if $|\partial_G(V)| \ge v(V)$ for every subset V of X, then G contains a perfect v-matching. We construct another bipartite graph $G'(X' \cup Y, E')$. For each vertex x_i in X, we construct v_i copies in X'. Further, if there exists an edge in G between x_i and a vertex in Y, then there exists an edge in G' between each copy of x_i in X' and the corresponding vertex in Y. With this construction, it is easy to see that a perfect v-matching exists in G if and only if a perfect matching exists in G'. Further, since $|\partial_G(V)| \ge v(V)$ for every subset V of X in G, we must have, for every subset V' of X' in G', $|\partial_{G'}(V')| \ge |V'|$. Thus, invoking Hall's Theorem [10] on G', we conclude that G' must have a perfect matching. Hence, G must contain a perfect v-matching.

APPENDIX D

Proof of Lemma 5: Let $N_v(k)$ denote the number of graphs with exactly k edges that have a perfect v-matching. Similarly, let $N_w(k)$ denote the number of graphs with exactly k edges that have a perfect w-matching. Let N(k) be the total number of possible graphs with exactly k edges. According to our system model, for any value of k, all of the graphs with k edges occur with probability $\frac{1}{N(k)} {n \choose k} q^k (1-q)^{nm-k}$. Hence,

$$\mathbb{P}(G \text{ has a perfect } v\text{-matching})$$

$$= \sum_{k=0}^{nm} \mathbb{P}(G \text{ has } k \text{ edges}) \frac{N_v(k)}{N(k)}$$

$$= \sum_{k=0}^{nm} \binom{nm}{k} q^k (1-q)^{nm-k} \frac{N_v(k)}{N(k)}$$

$$\mathbb{P}\Big(G \text{ has a perfect } w\text{-matching}\Big)$$
$$= \sum_{k=0}^{n^2} \mathbb{P}\Big(G \text{ has } k \text{ edges}\Big) \frac{N_w(k)}{N(k)}$$
$$= \sum_{k=0}^{n^2} \binom{nm}{k} q^k (1-q)^{nm-k} \frac{N_w(k)}{N(k)}.$$

From the above two equations it is easy to see that, to prove the result it suffices to show that $N_v(k) \ge N_w(k)$ for all $k \ge 0$. For $k < \sum_{i=1}^{|X|} v_i$, $N_v(k) = N_w(k) = 0$ and hence the statement is trivially true. We now prove this statement for any general $k \ge \sum_{i=1}^{|X|} v_i$.

We first introduce some definitions and notations that will be used in the proof. Let $w_2 = \theta$, $w_1 = \theta + r$, $r \ge 2$. We sequentially assign indices to all subsets of Y that have exactly $2\theta + r$ vertices. Without loss of generality let these indices be from 1 to $\binom{|Y|}{2\theta+r}$. Define $\mathcal{J}(\cdot)$ to be a function that maps every subset of Y with $2\theta + r$ elements to its corresponding index. Let ξ be any |X|-dimensional vector with non negative integer components such that $\xi_1 + \xi_2 = 2\theta + r$. Suppose that G has a prefect ξ -matching. Then, we define $M_{G,\xi}$ as the perfect ξ -matching in G such that $\mathcal{J}(\partial_{G(M_{G,\xi})}(x_1 \cup x_2))$ is the maximum among all perfect ξ -matchings in G. Define $\mathcal{J}_{G,\xi} :=$ $\mathcal{J}(\partial_{G(M_{G,\xi})}(x_1 \cup x_2))$.

We now construct the mapping $\Pi(\cdot)$ from graphs with k edges that have a perfect w-mapping to graphs with k edges that have a perfect v-matching.

- If G has a perfect v-matching and a perfect w-matching, then Π(G) = G. In this case skip, steps 2 and 3.
- If G has a perfect w-matching but not a perfect v-matching, let ξ = [θ + 1, θ + r − 1, w₃...w_{|X|}]. We first form a graph G* which has a perfect ξ-matching. To do this, we choose a subset P of ∂<sub>M_{G,w}(x₁) such that |P| = r − 1. We delete edges between x₁ and all vertices of the set P and construct edges between x₂ and each vertex in P.
 </sub>

 Π(G) is now obtained from G* by swapping the labels of x₁ and x₂.

It is easy to see that $\Pi(G)$ has a perfect *v*-matching. It can also be seen that the resultant graph $\Pi(G)$ and intermediate graph G^* (formed if step 2 is executed) have the same number of edges as *G*. This follows from the fact that we delete and add an equal number of edges.

We would like to show that if any two distinct graphs G'and G'' have perfect w-matchings and have equal number of edges then they map to different graphs, i.e., $\Pi(G') \neq \Pi(G'')$

. This would prove that, for any k > 0, the number of graphs which have a perfect *v*-matching and exactly *k* edges is no smaller than the number of graphs which have *k* edges and a perfect *w*-matching.

Firstly, note that step 1 ensures that if a graph has both a prefect w-matching and a prefect v-matching then it remains unchanged. Moreover, steps 2 and 3 transform a graph with a w-matching but no v-matching to a graph which has a perfect v-matching but does not have a perfect w-matching. Let G' be a graph with both a perfect w-matching and a perfect v-matching and let G'' be a graph with a perfect w-matching but not a perfect v-matching. Then, by the above observation it follows that $\Pi(G') \neq \Pi(G'')$. Further, if G' and G'' are two different graphs with both a w-matching and a v-matching, then $\Pi(G') = G' \neq G'' = \Pi(G'')$.

Secondly, let G' and G'' be two graphs that have a perfect w-matchings but not a perfect v-matchings. Further, assume that $\mathcal{J}_{G',w} \neq \mathcal{J}_{G'',w}$. Then, we show that $\Pi(G') \neq \Pi(G'')$. To prove this, it suffices to show that if a graph G has a perfect wmatching but not a perfect v-matching, then $\mathcal{J}_{\Pi(G),v} = \mathcal{J}_{G,w}$. To see this notice that there exists a matching M in $\Pi(G)$ such that $\mathcal{J}(\partial_{\Pi(G)(M)}(x_1 \cup x_2)) = \mathcal{J}_{G,w}$. For the sake of contradiction, assume that $\mathcal{J}_{\Pi(G),v} > \mathcal{J}_{G,w}$. Then, note that, for the intermediate graph G^* created step 2, we must have $\mathcal{J}_{G^*,\xi} = \mathcal{J}_{\Pi(G),v} > \mathcal{J}_{G,w}$. From construction we know that G^* is formed by deleting all edges between x_1 and P and creating edges between x_2 and all vertices in P, where P is a subset of $\partial_{G(M_{G,w})}(x_1)$. Then, from $M_{G^*,\xi}$ we can form a perfect w-matching, \hat{M} , in G by replacing edges between x_2 and every vertex of P by edges between x_1 and every vertex of P. It follows that $\mathcal{J}(\partial_{G(\hat{M})}(x_1 \cup x_2)) = \mathcal{J}_{\Pi(G),v} > \mathcal{J}_{G,w}$. But this contradicts our assumption on $\mathcal{J}_{G,w}$.

Thirdly, let G' and G'' be two graphs which have a perfect *w*-matching but do not have a perfect *v*-matching. Further, suppose that $\mathcal{J}_{G',w} = \mathcal{J}_{G'',w}$. If G' and G'' differ in any edge of the form (x_1, y) or (x_2, y) , where *y* belongs to the set $Y \setminus \partial_{G(M_{G,w})}(x_1 \cup x_2)$, then $\Pi(G') \neq \Pi(G'')$. To see this observe that we do not alter edges of the form above in the procedure of creating $\Pi(G)$ from *G*.

Hence, we are now left with graphs similar in all the three conditions mentioned below,

- 1) They have a perfect *w*-matching but do not have a perfect *v*-matching.
- 2) They have the same index .
- 3) They are the same everywhere except in the per-

mutation of $\partial_{G(M_{G,w})}(x_1 \cup x_2)$, i.e., for any two graphs G' and G'' of the same type, $\partial_{G'(M_{G',w})}(x_1 \cup x_2) = \partial_{G'(M_{G'',w})}(x_1 \cup x_2)$ but $\partial_{G'(M_{G'',w})}(x_1) \neq \partial_{G''(M_{G'',w})}(x_1)$.

If two graphs are similar in these three ways, then we say that they are of the same type. We would like to show that there exists a particular way to choose the set P such that all of the graphs that are of the same type, map to different graphs under $\Pi(\cdot)$. To see this note that there are $\binom{2\theta+r}{\theta}$ graphs of each type. On the other hand, the number of output graphs possible by an appropriate choice of P is $\binom{2\theta+r}{\theta+1}$. As $r \ge 2$, it is easy to see that $\binom{2\theta+r}{\theta+1} > \binom{2\theta+r}{\theta}$. Hence, the number of output graphs for any type are greater than the number of graphs of that type. Hence, we can always find a mapping such that if G' and G'' are of the same type, then $\Pi(G') \neq \Pi(G'')$. The result then follows.

Appendix E

Proof of Corollary 1: Let Ξ be the set of all |X|dimensional vectors with non-negative integer components, such that for any vector ξ in Ξ , $\sum_{i=1}^{|X|} \xi_i = n - H$ and $\max_{\{1 \le i \le |X|\}} \xi_i \le H$. First, for any vector v with $\sum_{i=1}^{|X|} v_i < n - H$ and $\max_{\{1 \le i \le |X|\}} v_i \le H$, it is easy to see that there exists a vector ξ' in Ξ , such that v is component-wise no greater than ξ' . Hence,

$$\mathbb{P}(G \text{ has a perfect } v \text{-matching})$$
$$\geq \mathbb{P}(G \text{ has a perfect } \xi' \text{-matching})$$

Thus, we only need to show that the result of the corrolary holds for all $v \in \Xi$.

Fix $v \in \Xi$. Due to symmetry of channels and users, the probability of finding a perfect *v*-matching does not depend on the permutation of the components of *v*. Hence, without loss of generality, we assume that the components of the vector *v* are in non-increasing order, i.e., $v_1 \ge v_2 \ge \dots \ge v_{|X|}$. Similarly, the probability of finding a perfect *w*-matching does not depend on the permutation of the components of *w*. Hence, we can rearrange *w* as

$$w_i = \begin{cases} H, & \text{if } 1 \le i \le k \\ n - (k+1)H, & \text{if } i = k+1 \\ 0, & \text{if } i > k+1, \end{cases}$$

where $k = \left\lceil \frac{n}{H} \right\rceil - 2$.

Consider the following algorithm:

1: $\lambda \leftarrow w$ 2: $c \leftarrow 1$ 3: $d \leftarrow |X|$ 4: while d > c do while $\lambda_c > v_c$ and $\lambda_d < v_d$ do 5: $\lambda_c \leftarrow \lambda_c - 1$ 6: $\lambda_d \leftarrow \lambda_d + 1$ 7: 8: end while 9: if $\lambda_c = v_c$ then $c \leftarrow c + 1$ 10:

11: end if 12: if $\lambda_d = v_d$ then 13: $d \leftarrow d - 1$ 14: end if 15: end while

Let us assume that $w_{k+1} > v_{k+1}$. The case when $w_{k+1} \le v_{k+1}$ can be analyzed in a similar way. To see that the above algorithm terminates when $\lambda = v$, observe that initially $H = \lambda_i \ge v_i$ for all $i \le k$ and $0 = \lambda_i \le v_i$ for all i > k+1. Further, $\sum_{i=1}^{|X|} v_i = \sum_{i=1}^{|X|} w_i \Rightarrow \sum_{i=1}^{k+1} (v_i - w_i) = \sum_{i=k+2}^{|X|} (w_i - v_i)$. Hence, after the inner while loop has executed exactly $\sum_{i=1}^{k+1} (v_i - w_i)$ times, c = k + 1 and d = k. The algorithm terminates at this point.

Moreover, in every step of the algorithm, the vector λ changes in exactly 2 components, i.e., λ_c is incremented by 1 and λ_d is decremented by 1. It can be observed that, due to the initial ordering of components, in every step λ becomes more balanced in the sense of Lemma 5. Hence, applying 5 recursively, the result follows.

APPENDIX F

Proof of Lemma 6: Define $k = \left\lceil \frac{n}{H} \right\rceil - 2$. Let v be a |X|dimensional vector with $v_1 = n - (k+1)H$, $v_i = H \forall 2 \le i \le k+1$ and $v_i = 0$ everywhere else. Then by Corollary 1,

$$\mathbb{P}\Big(G \text{ has a perfect } w\text{-matching}\Big)$$
$$\geq \mathbb{P}\Big(G \text{ has a perfect } v\text{-matching}\Big).$$

Suppose that $|\partial_G(x_1)| \geq v_1$. We can then choose a set $A \subset \partial_G(x_1)$ such that $|A| = v_1$. Now construct the graph $\hat{G}[X \cup Y, E]$ from G by deleting all edges between $(X - x_1)$ and A. Let vector $\hat{v} = [0, v_2, v_3, \dots v_n]$. If \hat{G} has a perfect \hat{v} -matching, say \hat{M} , then a perfect v-matching in G can be obtained as the union of \hat{M} and the set of edges between x_1 and A in G. From the above discussion, we conclude that

$$\mathbb{P}(G \text{ has a perfect } v \text{-matching}) \\ \geq \mathbb{P}(|\partial_G(x_1)| \geq v_1) \mathbb{P}(\hat{G} \text{ has a perfect } \hat{v} \text{-matching}).$$
(35)

Next, we will bound both terms in the product on the right hand side. For the first term, we know that,

$$\mathbb{P}\Big(|\partial_G(x_1)| \ge v_1\Big) \\
\ge \mathbb{P}\Big(|\partial_G(x_1)| \ge H\Big) \\
= 1 - \mathbb{P}\Big(|\partial_G(x_1)| < H\Big) \\
\ge 1 - \binom{n}{H-1}(1-q)^{n-H+1} \\
\ge 1 - n^{H-1}(1-q)^{n-H+1} \\
= 1 - \left(\frac{n}{1-q}\right)^{H-1}(1-q)^n \\
\ge 1 - \left(\frac{n}{1-q}\right)^H e^{-n\log\frac{1}{1-q}},$$
(3)

6)

where in the second inequality we have considered all cases where x_1 is not connected to a subset of n - H + 1 servers, and have taken the union bound over all such cases.

Now consider the second term in the product in (35). If \hat{G} has no \hat{v} -matching, then by Lemma 4, there must exist A and B such that

1)
$$A \subseteq X, B \subseteq Y$$
,
2) $|B| = \hat{v}(A) - 1$, and
3) $\partial_{\hat{r}}(A) \subseteq B$.

Hence, by union bound over all possible subsets $A \subseteq X$ and all possible corresponding subsets $B \subseteq Y$, we have

$$\begin{split} \mathbb{P} \Big(\hat{G} \text{ has no } \hat{v} \text{-matching} \Big) \\ &\leq \sum_{\{A \subseteq X\}} \mathbb{P} \Big(|\partial_{\hat{G}}(A)| < \hat{v}(A) \Big) \\ &= \sum_{\{A \subseteq X\}} \mathbb{P} \Big(|\partial_G(A)| \le \hat{v}(A) - 1 \Big) \end{split}$$

Here, we may consider only those A such that for all vertices x_i in A, $\hat{v}_i > 0$. Notice that the maximum size of any such set is k. Hence,

$$\mathbb{P}\left(\hat{G} \text{ has no } \hat{v}\text{-matching}\right)$$

$$\leq \sum_{a=1}^{k} \sum_{\{A \subseteq X, |A|=a\}} \mathbb{P}\left(|\partial_{G}(A)| \leq \hat{v}(A) - 1\right)$$

$$= \sum_{a=1}^{k} \sum_{\{A \subset X, |A|=a\}} \mathbb{P}\left(|\partial_{G}(A)| \leq Ha - 1\right)$$

$$\leq \sum_{a=1}^{k} \binom{k}{a} \binom{n}{n-Ha+1} (1-q)^{a(n-Ha+1)}, \quad (37)$$

where in the last step we have considered all sets A and $Y \setminus B$ such that |A| = a, $|Y \setminus B| = n - Ha + 1$, and the corresponding event that no vertices in A are connected to vertices in $Y \setminus B$. Let $N_1 = \left\lceil \frac{(1-q)H}{q} \right\rceil$. Then for all $n > N_1$, we have

$$n - Ha + 1)$$

$$= \binom{n}{Ha - 1}$$

$$\leq \binom{Hk + 2H}{Ha - 1} \quad (\text{as } n \leq (k + 2)H)$$

$$\leq (Ha)(Hk + 2H)^{2H}\binom{Hk}{Ha}$$

$$\leq \binom{n}{1 - q}(Hk + 2H)^{2H}\binom{Hk}{Ha} \quad (\text{as } Ha \leq n < \frac{n}{1 - q})$$

$$\leq (\frac{n}{1 - q})^{2H + 1}\binom{Hk}{Ha}, \quad (38)$$

where the last step is true because $Hk + 2H \le n + H \le \frac{n}{1-a}$

for all $n > N_1$. Moreover,

$$(1-q)^{a(n-Ha+1)} \leq (1-q)^{a(kH+H-Ha+1)}$$
 (as $n \ge (k+1)H$)
 $\le (1-q)^{Ha(k-a)}.$ (39)

Substituting (38) and (39) in (37), for all $n > N_2 := \max\{N_1, \left\lceil \frac{1}{(1-q)^{2H}} \right\rceil\}$ we have,

$$\mathbb{P}\left(\hat{G} \text{ has no } \hat{v}\text{-matching}\right) \\
\leq \sum_{a=1}^{k} \left(\frac{n}{1-q}\right)^{2H+1} {\binom{k}{a}} {\binom{Hk}{Ha}} (1-q)^{a(n-Ha+1)} \\
\leq \left(\frac{n}{1-q}\right)^{2H+1} (1-q)^{k(H+1)} \\
+ \left(\frac{n}{1-q}\right)^{2H+1} \sum_{a=1}^{k-1} {\binom{k}{a}} {\binom{Hk}{Ha}} (1-q)^{Ha(k-a)} \\
\leq \left(\frac{n}{1-q}\right)^{2H+1} (1-q)^{n-2H} \\
+ 2\left(\frac{n}{1-q}\right)^{2H+1} \sum_{a=1}^{\left\lfloor\frac{k}{2}\right\rfloor} {\binom{k}{a}} {\binom{Hk}{Ha}} (1-q)^{Ha(k-a)} \\
\leq \left(\frac{n}{1-q}\right)^{2H+1} (1-q)^{n} + 2\left(\frac{n}{1-q}\right)^{2H+1} \\
\times \sum_{a=1}^{\left\lfloor\frac{k}{2}\right\rfloor} e^{a\log k+aH\log kH-Ha(k-a)\log \frac{1}{(1-q)}}, \quad (40)$$

where the last step is true because $\frac{1}{(1-q)^{2H}} \leq \frac{n}{(1-q)^{2H-1}} \leq \left(\frac{n}{1-q}\right)^{2H-1}$.

Note that the exponent in the last summation is quadratic in a with positive leading coefficient. Hence, the terms will achieve their maxima at the end-points, i.e. a = 1 or $a = \lceil \frac{k}{2} \rceil$. It turns out that the term corresponding to a = 1 will dominate all other terms. To see this, substituting a = 1, we have,

$$e^{a\log k + aH\log kH - Ha(k-a)\log \frac{1}{(1-q)}} \le n^{(H+1)} \left(\frac{1}{1-q}\right)^{H} e^{-n\log \frac{1}{(1-q)}} \le \left(\frac{n}{1-q}\right)^{H+1} e^{-n\log \frac{1}{(1-q)}}.$$
(41)

for all $n > N_2$. Similarly, by substituting $a = \lfloor \frac{k}{2} \rfloor$, we have

$$e^{a \log k + aH \log kH - Ha(k-a) \log \frac{1}{(1-q)}} \le e^{2n \log n - \frac{n^2}{6H} \log \frac{1}{(1-q)}} \le e^{-n \log \frac{1}{(1-q)}} \times e^{3n \log n - \frac{n^2}{6H} \log \frac{1}{(1-q)}} \le e^{-n \log \frac{1}{(1-q)}},$$
(42)

for all $n > N_3$, where N_3 is the smallest integer such that the exponent in the second step satisfies $3n \log n - \frac{n^2}{6H} \log \frac{1}{(1-q)} < 0$ for all $n > N_3$.

Hence, from (40), (41) and (42), we have

$$\begin{split} & \mathbb{P}\Big(\hat{G} \text{ has no } \hat{v}\text{-matching}\Big) \\ & \leq \Big(\frac{n}{1-q}\Big)^{4H} e^{-n\log\frac{1}{(1-q)}} \\ & + 2\left\lceil \frac{k}{2} \right\rceil \Big(\frac{n}{1-q}\Big)^{3H+2} e^{-n\log\frac{1}{(1-q)}} \\ & \leq \Big(\frac{n}{1-q}\Big)^{4H} e^{-n\log\frac{1}{(1-q)}} \\ & + \Big(\frac{n}{1-q}\Big)^{3H+3} e^{-n\log\frac{1}{(1-q)}} \quad (\text{as } 2\left\lceil \frac{k}{2} \right\rceil \leq \frac{n}{1-q}) \\ & \leq 2\Big(\frac{n}{1-q}\Big)^{6H} e^{-n\log\frac{1}{(1-q)}} \quad (\text{as } H \geq 1) \end{split}$$

for all $n > N_X := \max\{N_2, N_3\}.$

Finally, substituting (36) and (43) into (35), we have

$$\mathbb{P}(G \text{ has a perfect } v \text{-matching})$$

$$\geq \mathbb{P}(\partial_G(x_1) \geq v_1) \mathbb{P}(\hat{G} \text{ has a perfect } \hat{v} \text{-matching})$$

$$\geq 1 - 3\left(\frac{n}{1-q}\right)^{6H} e^{-n\log\frac{1}{1-q}}$$

$$\geq 1 - \left(\frac{n}{1-q}\right)^{7H} e^{-n\log\frac{1}{1-q}}, \qquad (43)$$

for all $n > N_X$.

1

APPENDIX G

We first prove a result that will be used in the proof of Lemma 7.

Lemma 10: Consider a bipartite graph $G[X \cup Y, E]$ such that every vertex in X is assigned a *unique* positive weight. Let the vertices of X in the decreasing order of their weights be $\{x_1, x_2, ..., x_n\}$. Suppose that there exists a matching M that covers $\{x_1, x_2, ..., x_k\}$, i.e., the set of vertices with k largest weights. Then the maximum-weight matching M^* of G covers the set $\{x_1, x_2, ..., x_k\}$.

Proof: Suppose (for contradiction) that there exists a vertex $x_i \in \{x_1, ..., x_k\}$ that is not covered by M^* . Without loss of generality, assume that x_i is such a vertex with the smallest index. In other words, $x_1, ..., x_{i-1}$ are covered by M^* , but x_i is not. Let $u_1 = x_i$. Since $u_1 = x_i$ is covered by the matching M, it must be matched to a node $s_1 \in Y$. Note that s_1 must be covered by the matching M^* : if it was not covered by M^* , we would have been able to add the edge (u_1, s_1) to the matching M^* , which contradicts to the assumption that M^* is the maximum-weight matching. Let $u_2 \in X$ be the vertex matched to s_1 by the matching M^* . Note that u_2 must belong to $\{x_1, \dots, x_{i-1}\}$. (Otherwise, if u_2 was one of the vertices in $\{x_{i+1}, ..., x_n\}$, then instead of matching s_1 to u_2 in M^* , we could have matched s_1 to u_1 , in which case we would have gotten a matching with a larger weight than M^* .) Hence, u_2 is again covered by the matching M. Using a similar argument, we can thus find $s_2 \in Y$ such that u_2 is matched to s_2 by the matching M, and find $u_3 \in X$ such that u_3 is matched to s_2 by the matching M^* . We can continue performing this procedure as long as we can find a vertex u_k in each step that belongs to $\{x_1, ..., x_{i-1}\}$. Further, each of the u_k 's is distinct because they are part of the matching M, and each of the s_k 's is distinct because they are part of the matching M^* . However, since there are only (i - 1) vertices in the set $\{x_1, ..., x_{i-1}\}$, eventually we will run out of vertices from this set. The last s_K vertex must then be matched to a vertex u_{K+1} in $\{x_{i+1}, ..., x_n\}$ by the matching M^* . Then, by matching each s_k to u_k , we obtain a matching with a larger weight than M^* , which contradicts to the assumption that M^* is the maximum-weight matching. Hence, the result of the lemma must hold.

Proof of Lemma 7: Consider two queuing systems, Q_1 and \mathbf{Q}_2 , each consisting of *n* queues and *n* channels. Both systems have the same arrivals and channel realizations. Q_1 uses DWM and \mathbf{Q}_2 uses FBS(h) as the scheduling rule. Suppose that a packet p enters in the system at time t_p at queue q_p . Then, according to the DWM policy, at any time $t \ge t_p$ if packet p is still present in the system, it has a weight $w(p) = t - t_p + \frac{q_p}{n+1}$. For convenience we may make the following assumption: packets that arrive to the same queue in the same time-slot are queued in the order that they arrive, and they must be served in the same order. We assign every packet p an order-index x_p from 1 to L. If a packet is the first to arrive to a queue in a time-slot, then it is given the order-index L. The next packet is given the order-index L-1, and so on. We then redefine the weights of the packets as $\hat{w}(p) = t - t_p + \frac{q_p}{n+1} + \frac{x_p}{(L+1)(n+1)}$. It must be noted that the DWM schedule would not change if we used weights $\hat{w}(\cdot)$ instead of $w(\cdot)$. However, $\hat{w}(\cdot)$ makes the analysis easier as now every packet in the system has a unique weight.

Let $Q_1(t)$ represent the set of packets present in the system \mathbf{Q}_1 at the end of time-slot t and let $Q_2(t)$ represent the set of packets present in the system \mathbf{Q}_2 at the end of time-slot t. Then, it suffices to show that $Q_1(t) \subset Q_2(t)$ for all time t. We denote by A(t) the set of packets that arrive at time t. Let $X_1(t)$ and $X_2(t)$ denote the set of packets that depart the systems \mathbf{Q}_1 and \mathbf{Q}_2 respectively, at time t. Hence,

$$Q_i(t+1) = (Q_i(t) \cup A(t+1)) \setminus X_i(t+1), \quad i = 1, 2.$$

We prove by contradiction. Suppose that $Q_1(t) \not\subseteq Q_2(t)$ for some t. Without loss of generality, assume that τ is the first time slot such that $Q_1(\tau) \not\subseteq Q_2(\tau)$. Hence, there must be a packet, p, such that $p \in Q_1(\tau)$ and $p \notin Q_2(\tau)$. Because τ is the first time for such an event, we must have $p \in X_2(\tau)$.

Let $B_1(v)$ and $B_2(v)$ denote the set of packets in $Q_1(\tau - 1) \cup A(\tau)$ and $Q_2(\tau - 1) \cup A(\tau)$ with weight greater than or equal to v. Then $B_1(v) \subset B_2(v)$ for all v. As p is scheduled by \mathbf{Q}_2 , thus from the definition of FBS(h), we know that all packets in $B_2(w(p))$ must also be scheduled at time t. This is because all packets with a weight greater than w(p) must belong to either of the following categories-

- 1) Packets that arrived before t_p .
- 2) Packets that arrived at t_p but to queues with higher indices.

3) Packets that arrived at t_p to q_p but have a higher orderindex than x_p , i.e. they are queued in before p.

Hence, if p is a part of the head-of-line frame at time t, then all of these packets should belong to the head-of-line frame at time t.

Consider a bipartite graph $G[X \cup Y, E]$ where $X = Q_2(\tau - 1) \cup A(t)$, Y is the set of servers and E denotes the connectivity at time τ . Then, the above statement means that there exists a matching, M, such that every vertex of $B_2(w(p))$ is covered under M. Hence, M must also cover every vertex of $B_1(w(p))$. Now, from the graph we remove the set of vertices corresponding to packets that are present in $Q_2(\tau - 1)$ but are not present in $Q_1(\tau - 1)$. Let the new graph be graph $G'[X' \cup Y', E']$. Note that X' is just $Q_1(\tau - 1) \cup A(\tau)$. Let the restriction of M to G' be M'. Then M' covers all vertices in $B_1(w(p))$. Moreover, notice that $B_1(w(p))$ is the set of packets with the highest weights in G'.

From the definition of DWM, $X_1(\tau)$ is determined by the maximum weight matching in G'. We have shown that there exists a matching M' that covers the set $B_1(w(p))$, i.e., the set of the packets with the highest weights. Then, by Lemma 10, the maximum weight matching covers every vertex in $B_1(w(p))$. However, this contradicts to the claim that p is not scheduled by $\mathbf{Q_1}$ at time τ .

APPENDIX H

Proof of Lemma 9: Referring to equation (7), condition 1 in the statement of Lemma 9 implies that $1_{\{F(j)>0\}} = 1$ for all $j \in \{i, i+1, ..., k\}$ and conditions 1 and 2 together imply that R(j) = 0 only when $[R(j-1) - A(j)] \mod(n_0) = 0$. Hence, for the interval from i to k, we can write the recursive equation for $A_F^{R_0}(i, j)$ as follows: for any $i \leq j \leq k$,

$$A_F^{R_0}(i,j) = A_F^{R_0}(i,j-1) + \left[\frac{A(j) - R(j-1)}{n_0}\right]$$
(44)

$$R(j) = [R(j-1) - A(j)] \mod(n_0), \tag{45}$$

where $A_F^{R_0}(i,i) = 0$ and $R(i-1) = R_0$. We now prove the equality in (8) by induction on k. For k = i the equality in (8) is trivially true. Suppose that the equality in (8) is true for some k = j. Then, we want to show that it is also true for k = j + 1 that satisfies the conditions in the lemma. To see this, note that from (44),

$$\begin{split} A_F^{R_0}(i, j+1) &= A_F^{R_0}(i, j) + \left\lceil \frac{A(j+1) - R(j)}{n_0} \right\rceil \\ &= A_F^{R_0}(i, j) + \left\lceil \frac{A(j+1) - [R_0 - A(i, j)] \operatorname{mod}(n_0)}{n_0} \right\rceil \\ &= A_F^{R_0}(i, j) \\ &+ \left\lceil \frac{A(j+1) - \left\lceil \frac{A(i, j) - R_0}{n_0} \right\rceil n_0 + A(i, j) - R_0}{n_0} \right\rceil \\ &= A_F^{R_0}(i, j) + \left\lceil \frac{A(i, j+1) - R_0}{n_0} \right\rceil - \left\lceil \frac{A(i, j) - R_0}{n_0} \right\rceil \end{split}$$

$$= \left\lceil \frac{A(i,j+1) - R_0}{n_0} \right\rceil$$

Further, from (45),

$$\begin{aligned} R(j+1) &= [R(j) - A(j+1)] \operatorname{mod}(n_0) \\ &= [[R_0 - A(i,j)] \operatorname{mod}(n_0) - A(j+1)] \operatorname{mod}(n_0) \\ &= [[R_0 - A(i,j+1)] \operatorname{mod}(n_0). \end{aligned}$$

Hence, by induction the equality (8) must be true for any k that satisfies the conditions in the lemma.

APPENDIX I

Proof: We first choose the value of h based on the statistics of the arrival process. Let the mean of the arrival process be \bar{a} . We fix $\delta < \frac{2}{3}$ and $\epsilon < \frac{\phi \bar{a}}{2}$. Then, from Assumption 2 on the arrival process, there exists a positive function $I_B(\epsilon, \delta)$, such that for all $n > \frac{N_B(\epsilon, \delta)}{\phi}$ and $t > T_B(\epsilon, \delta)$ we have,

$$\mathbb{P}\Big(\frac{\sum_{i=j+1}^{j+t} \mathbf{1}_{\{|A(i)-\bar{a}\phi n| > \epsilon\phi n\}}}{t} > \delta\Big) < e^{-ntI_B(\epsilon,\delta)},$$

for any integer j.

We then choose

$$h = \max\left\{T_B(\epsilon, \delta), \left\lceil \frac{1}{(\bar{a} - \epsilon)(1 - \frac{3\delta}{2})} \right\rceil, \left\lceil \frac{2I_0}{I_B(\epsilon, \delta)} \right\rceil\right\} + 1.$$
(46)

The reason for choosing such a value of h will become clear later on. Recall that L is the maximum number of packets that can arrive to a queue at any time-slot i and H = hL. Note that H is then the maximum number of packets that can arrive to a queue in h time-slots.

Let L(0) be the last time, -t before 0, when the buffer was empty, i.e., D(-t) = 0. Then given that L(0) = -t, the event D(0) > d occurs if and only if the number of frames that arrive in the time interval from -t + 1 to 0 is greater than the total number of frames that could be served in -t + 1 to d. That is,

$$\left\{ D(0) > d, L(0) = -t \right\}$$

= $\left\{ L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0 \right\}.$

By taking the union over all possible values of L(0) we get,

$$\mathbb{P}\Big(D(0) > d\Big)$$

 $\leq \sum_{t=1}^{\infty} \mathbb{P}\Big(L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0\Big).$

We now fix any $1 > \hat{p} > \phi \bar{a}$. Then define,

$$t^* := \max\left\{ T_B(\hat{p} - \bar{a}, \frac{1 - \hat{p}}{6(L+2)}), \left\lceil \frac{6}{1 - \hat{p}} \right\rceil, \\ \left\lceil \frac{I_0}{\min\{I_B(\hat{p} - \bar{a}, \frac{1 - \hat{p}}{6(L+2)}), (\frac{1 - \hat{p}}{9})I_X\}} \right\rceil \right\}$$
(47)

and split the summation as,

$$\mathbb{P}(D(0) > d) \leq \sum_{t=1}^{t^*} \mathbb{P}(L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0) + \sum_{t=t^*}^{\infty} \mathbb{P}(L(0) = -t, A_F(-t+1, 0) - X_F(-t+1, d) > 0).$$
(48)

We divide the proof into two parts. In Part 1 we prove that there exists $N_1 > 0$ such that for all $n > N_1$

$$\sum_{t=1}^{t^*} \mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big)$$

$$< c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} e^{-nI_0},$$

where c_1, c_2 are positive constants independent of t and n. And in Part 2 we prove that there exists $N_2 > 0$ such that for all $n > N_2$

$$\sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big) \le 4e^{-nI_0}.$$

Finally, by substituting both parts into equation (48), we have that there exists $N := \max\{N_1, N_2\}$ such that for all n > N,

$$\sum_{t=1}^{\infty} \mathbb{P}\Big(D(0) > d, L(0) = -t\Big)$$
$$\leq (c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} + 4)e^{-nI_0}.$$

By taking logarithm and limit as n tends to infinity, we get the desired result.

Part 1: Let us denote by \mathcal{E}_t^{α} the set of sample paths in which every h - 1 time-slots in the interval -t + 1 to 0 see at least n arrivals. Let \mathcal{E}_t^{β} be the set of sample paths in which $\frac{A(-t+1,0)}{n_0} - \sum_{j=-t+1}^d X_F(j) > 0$. And let \mathcal{E}_t be the sample paths such that L(0) = -t and D(0) > d. Then, the following can be shown,

$$\mathcal{E}_t \subset (\mathcal{E}_t^{\alpha})^c \cup \mathcal{E}_t^{\beta}.$$
(49)

To see this, observe that \mathcal{E}_t is the set of sample paths in which L(0) = -t and $A_F(-t+1,0) - X_F(-t+1,d) > 0$. For all sample paths in the set $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, Lemma 9 holds and hence, $A_F(-t+1,0) = \left\lceil \frac{A(-t+1,0)}{n_0} \right\rceil$. Moreover, it is easy to observe that, for all sample paths in the set $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, $X_F(-t+1,d) = \sum_{j=-t+1}^d X_F(j)$. Hence, for a sample path belonging to $\mathcal{E}_t^{\alpha} \cap \mathcal{E}_t$, we must have $\frac{A(-t+1,0)}{n_0} - \sum_{j=-t+1}^d X_F(j) > 0$. This implies that, $\mathcal{E}_t \cap \mathcal{E}_t^{\alpha} \subset \mathcal{E}_t^{\beta}$. Thus we have, $\mathcal{E}_t = (\mathcal{E}_t \cap \mathcal{E}_t^{\alpha}) \cup (\mathcal{E}_t \cap (\mathcal{E}_t^{\alpha})^c) \subset (\mathcal{E}_t^{\beta} \cup (\mathcal{E}_t^{\alpha})^c)$. Hence, (49) holds. It then follows that,

$$\mathbb{P}(\mathcal{E}_t) \le \mathbb{P}((\mathcal{E}_t^{\alpha})^c) + \mathbb{P}(\mathcal{E}_t^{\beta}).$$
(50)

We note that the event \mathcal{E}_t^{α} implies that every frame formed in the interval from -t + 1 to 0 will have n_0 packets, i.e. all frames served are completely full. It is then obvious that $\mathbb{P}(\mathcal{E}_t^{\alpha})$ depends on h, i.e., it will be large if we increase the maximum time for which any frame can remain open. By choosing an h large enough we can ensure that the probability $\mathbb{P}((\mathcal{E}_t^{\alpha})^c)$ is arbitrarily small. In particular, we can ensure that the rate-function of $\mathbb{P}((\mathcal{E}_t^{\alpha})^c)$ is greater than the rate-function of $\mathbb{P}(\mathcal{E}_t^{\beta})$. The cost that needs to be paid for having a large h is the loss in frame-size, which is $n_0 = n - Lh$. But this decrease in frame-size is independent of n and does not affect the performance of the system significantly for large n. Hence, it does not show up in the rate-function. Specifically, for the choice of h in (46), it can be shown that there exists $N_3, c_3 > 0$ such that we have

$$\mathbb{P}(\mathcal{E}_t^{\alpha}) > 1 - c_3 t e^{-nI_0}.$$
(51)

for all $n > N_3$.

To see this, we first divide the time interval from -t+1 to 0 into frames of $\frac{h-1}{2}$ time-slots each. We index these frames from 1 to $\left\lceil \frac{2t}{h-1} \right\rceil$. Then, from assumption 2 on the arrival process, there exists $N_8 > 0$ such that for any frame j we have

$$\mathbb{P}\Big(\frac{\sum_{i=-j}^{-(j-1)} \frac{(h-1)}{2} \mathbf{1}_{\{|A(i)-\bar{a}m| > \epsilon m\}}}{\frac{h-1}{2}} < \delta\Big) \\> 1 - e^{-\frac{m(h-1)I_B(\epsilon,\delta)}{2}} \\\geq 1 - e^{-nI_0} \quad \text{(from (46) as } (h-1) \ge \frac{2I_0}{\phi I_B(\epsilon,\delta)} \ \text{)}$$

for all $n > N_8$. This equation implies that in every frame, with high-probability, the fraction of time-slots when $A(i) < \bar{a}n - \epsilon n$ is less than δ . Taking the union over all frames, we have

$$\mathbb{P}\Big(\frac{\sum_{i=-j\frac{h}{2}+1}^{-(j-1)\frac{h}{2}}\mathbf{1}_{\{|A(i)-\bar{a}m|>\epsilon m\}}}{\frac{h}{2}} < \delta \ \forall \ j \in \{1,2,...\lceil\frac{2t}{h-1}\rceil\}\Big) > 1 - \Big(\frac{2t}{h-1}+1\Big)e^{-nI_0}.$$
(52)

We now show that if the above statement is true, then the sum of arrivals for any h-1 consecutive time-slots in the interval -t+1 to 0 is greater than n. To see this, let k be any integer from the set $\{-t+1, \ldots -h+1\}$. Let us consider the interval from k to k+h-2. Note that this interval can intersect at most three frames of $\frac{h-1}{2}$ time-slots. It then follows that the interval from k to k+h-2 can have at most $\frac{3\delta h}{2}$ time-slots with arrivals less than $\bar{a}m-\epsilon m$. Hence, $A(k,k+h-2) \ge (h-1)(1-\frac{3\delta h}{2})(\bar{a}-\epsilon)\phi n$. From the choice of $h-1 \ge \left\lceil \frac{1}{(\bar{a}-\epsilon)(1-\frac{3\delta}{2})\phi} \right\rceil$, it follows that $A(k,k+h-2) \ge n$. Hence, all sample paths in which $\left\{ \frac{\sum_{i=-j\frac{h}{2}+1}^{-(j-1)\frac{h}{2}} \mathbf{1}_{\{|A(i)-\bar{a}n|>\epsilon n\}}}{\frac{h}{2}} < \delta \ \forall \ j \in \{1,2,\ldots\lceil \frac{2t}{h-1}\rceil\} \right\}$ belong to the set \mathcal{E}_t^{α} . It then follows from (52) that

$$\mathbb{P}(\mathcal{E}_{t}^{\alpha}) > 1 - (\frac{2t}{h} + 1)e^{-nI_{0}} \\ \ge 1 - 3te^{-nI_{0}}$$
(53)

It can be seen that the event \mathcal{E}_t^{β} is similar to the buffer overflow event in a single-server queues with intermittent connectivity as described earlier. Recall that as opposed to a single-server queue with constant rate, in every time-slot, with probability approximately $1 - e^{-nI_X}$ the service is equal to n_0 packets, i.e., one frame. So now there can be two factors responsible for \mathcal{E}_t^{β} . Firstly, if the arrival process is bursty, then \mathcal{E}_t^{β} can be caused by a large burst of arrivals in a few timeslots. Secondly, if q is small \mathcal{E}_t^{β} can be caused by a time interval of low service as frames get piled up in the buffer. For moderate values of q, one can expect that the most likely way in which \mathcal{E}_t^{β} occurs is a mixture of bursty arrivals and sluggish service. From large deviations theory we know that the rate-function of \mathcal{E}_t^{β} is determined by the probability of the most likely sample path leading to \mathcal{E}_t^{β} . More formally, we show that there exists $c_2, c_4, N_4 > 0$ such that

$$\mathbb{P}\left(\mathcal{E}_{t}^{\beta}\right) \leq c_{4} 2^{c_{2}t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_{0}},\tag{54}$$

for all $n > N_4$.

We first derive an upper bound for the probability of a large burst of arrivals in the time from -t + 1 to 0. We know from the Chernoff bound that for any integers t > 0 and $x \ge 0$,

$$\mathbb{P}\Big(A(-t+1,0) \ge n_0(t+x)\Big)$$
$$= \mathbb{P}\Big(A(-t+1,0) \ge (n-H)(t+x)\Big)$$
$$< e^{-n(\theta(t+x)-\lambda_{a_q(-t+1,0)}(\theta))+H(t+x)\theta}$$

Let $\theta_t = \arg \max_{\theta} [\theta(t+x) - \lambda_{a_q(-t+1,0)}(\theta)]$. Let $\theta^* = \max\{\theta_1, \theta_2, ..., \theta_{t^*}\}$. Then for any t in $\{1, 2, ..., t^*\}$,

$$\mathbb{P}\Big(A(-t+1,0) \ge (n-H)(t+x)\Big) \\
\le e^{-n(\theta(t+x)-\lambda_{a_q(-t+1,0)}(\theta))}e^{H(t+x)\theta^*} \\
\le e^{-nI_A(x)}e^{H(t+x)\theta^*}$$
(55)

Note that for large n, the probabilities of the event $\{A(-t+1,0) \ge n_0(t+x)\}$ and $\{A(-t+1,0) \ge n(t+x)\}$ differ by a factor which does not depend on n. Hence, in the large deviations sense, the rate-functions of the these two events is the same.

Now we consider the effect of sluggish service. Specifically, we calculate an upper bound on the probability that, in the interval from -t + 1 packets can be served in exactly t + a time-slots, for some $a \le d$. Recall from Lemma 6 that for all $n > N_X$, the probability of receiving service in each time slot is greater than $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-n\log\frac{1}{1-q}}$. Hence, we have

$$\mathbb{P}\left(\sum_{-t+1}^{d} X_F(i) = t+a\right) \\
\leq \left(\frac{n}{1-q}\right)^{7H} \binom{t+d}{t+a} e^{-(d-a)n\log\frac{1}{1-q}} \\
\leq \left(\frac{n}{1-q}\right)^{7H} 2^{t+d} e^{-(d-a)n\log\frac{1}{1-q}}.$$
(56)

It may be observed that this is monotonic function in a.

Using the results from (56) and (55), we have

$$\begin{split} \mathbb{P}\Big(\mathcal{E}_{t}^{\beta}\Big) \\ &= \mathbb{P}\Big(\frac{A(-t+1,0)}{n_{0}} - \sum_{j=-t+1}^{d} X_{F}(j) > 0\Big) \\ &= \sum_{a=0}^{t+d} \mathbb{P}\Big(\sum_{j=-t+1}^{d} X_{F}(j) = a\Big) \mathbb{P}\Big(A(-t+1,0) > an_{0}\Big) \\ &\leq (t+d+1) \max_{0 \leq a \leq t+d} \Big\{ \mathbb{P}\Big(\sum_{j=-t+1}^{d} X_{F}(j) = a\Big) \\ &\times \mathbb{P}\Big(A(-t+1,0) > an_{0}\Big) \Big\} \\ &\leq (t+d+1) \max\Big\{ \mathbb{P}\Big(A(-t+1,0) > (t+d)n_{0}\Big), \\ \mathbb{P}\Big(X_{F}(-t+1,d) = t-1\Big), \\ &a \in \{0,1...,d-1\} \Big\{ \mathbb{P}\Big(X_{F}(-t+1,d) = t+a\Big) \\ &\times \mathbb{P}\Big(A(-t+1,0) > (t+a)n_{0}\Big) \Big\} \Big\} \\ &\leq (t+d+1) \max\Big\{ 2^{t+d}\Big(\frac{n}{1-q}\Big)^{7H} e^{-n(d+1)I_{X}}, \\ &2^{t+d}e^{(t+d)H\theta^{*}}\Big(\frac{n}{1-q}\Big)^{7H} \\ &\times e^{-n\min_{a \in \{0,1,2...d\}}\{I_{A}(a) + (d-a)I_{X}\}} \Big\} \\ &\leq (t+d+1)2^{(t+d)(\frac{H\theta^{*}}{\log^{2}}+1)}\Big(\frac{n}{1-q}\Big)^{7H} \\ &\times \max\Big\{\Big(\frac{n}{1-q}\Big)^{7H} e^{-n(d+1)I_{X}}, \\ &e^{-n\min_{a \in \{0,1,2...d\}}\{I_{A}(a) + (d-a)I_{X}\}}\Big\} \\ &\leq c_{4}2^{c_{2}t}\Big(\frac{n}{-1}\Big)^{7H} \end{split}$$

$$(1-q)$$

× $e^{-n\min\{(d+1)I_X,\min_{a\in\{0,1,2...d\}}\{I_A(a)+(d-a)I_X\}\}}$

$$\leq c_4 2^{c_2 t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_0},$$
(57)

for all $n > N_4$, where $c_4 = 2^{d(\frac{H\theta^*}{\log 2} + 2)+1}$ and $c_2 = \frac{H\theta^*}{\log 2} + 2$. Let $c_1 = 2 \max\{c_3, c_4\}$. Substituting (53) and (57) into (50)

we then have

$$\mathbb{P}(\mathcal{E}_t) \le c_3 t e^{-nI_0} + c_4 2^{c_2 t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_0} \\ \le c_1 2^{c_2 t} \left(\frac{n}{1-q}\right)^{7H} e^{-nI_0}$$

for all $n > N_1 = \max\{N_3, N_4\}$. Finally, summing over t = 1to t^* we have,

$$\sum_{t=1}^{t^*} \mathbb{P}\Big(D(0) > d, L(0) = -t\Big) = \sum_{t=1}^{t^*} \mathbb{P}\Big(\mathcal{E}_t\Big)$$
$$\leq c_1 t^* 2^{c_2 t^*} \Big(\frac{n}{1-q}\Big)^{7H} e^{-nI_0},$$

for all $n > N_1$.

Part 2: We would like to show that there exists $N_2 > 0$ such that for $n > N_2$

$$\sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t,0) - X_F(-t,d) > 0\Big) < 4e^{-nI_0}.$$

We noted earlier that the equations for evolution of $A_F(-t+$ (1,0) are in general complicated. But if an arrival process satisfies certain conditions then some simple results such as Lemma 9 and Corollary 2 can be obtained. Hence, to analyze $A_F(-t+1,0)$ we first construct an arrival process $\hat{A}(\cdot)$ that satisfies the conditions of Lemma 9 and $A_F(-t+1,0) >$ $A_F(-t+1,0)$. We do this by adding some extra arrivals to the process $A(\cdot)$ in some strategic time-slots. The resulting arrival process $\hat{A}(\cdot)$ has the property that $\hat{A}(i) = \hat{p}n$ whenever $A(i) \leq \hat{p}n$ and $\hat{A}(i) = \phi Ln$ whenever $A(i) > \hat{p}n$. Hence, the resulting arrival process $\hat{A}(\cdot)$ is in fact very simple. We can then get an upper bound on $\hat{A}_F(-t+1,0)$, which, by construction, is also an upper bound on $A_F(-t+1,0)$.

Before constructing $A(\cdot)$, we first consider another arrival process denoted by $A'(\cdot)$ such that for the new arrival process $A'(\cdot)$, every frame formed in the interval -t + 1 to 0 is completely full, i.e., has exactly n_0 packets each. Moreover, $A_{F}(-t+1,0) = A'_{F}(-t+1,0)$. It is then easy to see that from arguments similar to Lemma 9, we have $A'_F(-t+1,0) =$ $\frac{A'(-t+1,0)}{n_0}$. We construct $A'(\cdot)$ as follows,

1: for i = -t + 1 to 0 do if Z(i-1) = h-1 and A(i) < R(i-1) then 2.

3:
$$r \leftarrow R(i-1) - A(i)$$

4: $j \leftarrow 0$
5: while $r > 0$ do
6: if $A(i-j) > \hat{p}n$ then
7: $\delta A = 0$
8: else
9: $\delta A = \min\{r, \hat{p}n - A(i-j)\}$
10: end if
11: $A'(i-j) \leftarrow A(i-j) + \delta A$
12: $r \leftarrow r - \delta A$
13: $j \leftarrow j + 1$
14: end while
15: end if
16: end for

In the above algorithm, step 2 checks to see if at time ithere is a frame that has been open for h time-slots but has not received n_0 packets. Note that if this is true, then the arrivals in the interval i - h + 1 to i must be less than n_0 . In order to make the frame completely full, we add extra arrivals to these time-slots till the frame-size becomes n_0 . While doing so, we do not add packets to time-slots that already have greater than $\hat{p}n$ arrivals, and we do not add more than $\hat{p}n - A(j)$ arrivals to a time-slot j with $A(j) < \hat{p}n$. Note that as $\hat{p} > \phi \bar{a}$ and $h-1 > \frac{1}{\overline{a}}$, hence it follows that $\hat{p}n(h-1) > n > n_0$. This implies that we can always fill frames in this manner.

From the algorithm above, it follows that $A(i) < \hat{p}n$ if and only if $A'(i) \leq \hat{p}n$. Now we construct the arrival process $\hat{A}(\cdot)$ which satisfies the conditions of Lemma 9 as follows

1: for i = -t + 1 to 0 do if $A(i) > \hat{p}n$ then 2: $A(i) \leftarrow \phi Ln$ 3: 4: else 5: $A(i) \leftarrow \hat{p}n$ end if 6: 7: end for

The resulting arrival process $\hat{A}(.)$ is in fact very simple. For most time-slots, $\hat{A}(i) = \hat{p}n$ and then for a few time-slots A(i) = Ln. Notice that whenever $A'(i) \leq \hat{p}n$ then A(i) = $\hat{p}n$ and whenever $A'(i) \geq \hat{p}n$ then $\hat{A}(i) = \phi Ln \geq \hat{p}n$. It follows that $\hat{A}((-t+1,0))$ is greater than A'(-t+1,0). Thus from Lemma 9 we have $\hat{A}_F(-t+1,0) = \left\lceil \frac{\hat{A}(-t+1,0)}{n_0} \right\rceil \geq$ $\left\lceil \frac{A'(-t+1,0)}{n_0} \right\rceil = A'_F(-t+1,0) = A_F(-t+1,0).$ So an upper bound on $\hat{A}_F(-t+1,0)$ is an upper bound on $A_F(-t+1,0)$.

Let $B = \{b_1, b_2, , b_{|B|}\}$ be the set of time-slots in the interval -t + 1 to 0 when $A(i) \ge \hat{p}n$. Then from Corollary 2 we have that, given L(0) = -t,

$$\begin{split} A_{F}(-t+1,0) \\ &\leq \sum_{j=1}^{|B|-1} \left\lceil \frac{\hat{A}(b_{j}+1,b_{j+1}-1)}{n_{0}} \right\rceil + \sum_{j=1}^{|B|} \left\lceil \frac{\hat{A}(b_{j},b_{j})}{n_{0}} \right\rceil \\ &\quad + \left\lceil \frac{\hat{A}(-t+1,b_{1}-1)}{n_{0}} \right\rceil + \left\lceil \frac{\hat{A}(b_{|B|}+1,0)}{n_{0}} \right\rceil \\ &\leq \sum_{j=1}^{|B|-1} \frac{(b_{j+1}-1-b_{j})\hat{p}n}{n_{0}} + \sum_{j=1}^{|B|} \frac{\phi Ln}{n_{0}} \\ &\quad + \frac{(b_{1}-(-t+1)+1)\hat{p}n}{n_{0}} + \frac{(0-b_{|B|})\hat{p}n}{n_{0}} + 2|B| + 1 \\ &\leq \frac{(t-|B|)\hat{p}n}{n_{0}} + |B|\frac{\phi Ln}{n_{0}} + 2|B| + 1 \\ &\leq \frac{n}{n_{0}}[\hat{p}t + (\phi L+2)|B| + 1]. \end{split}$$

From Assumption 2 on the arrival process we know that for large enough n and t, |B| can be made less than an arbitrarily small fraction of t. Further, we can show that for $|B| < \frac{1-\hat{p}}{6(L+2)}t$, $n > \frac{H(2+\hat{p})}{1-\hat{p}}$ and $t > \frac{6}{1-\hat{p}}$, $A_F(-t+1,0) \le \frac{1-\hat{p}}{1-\hat{p}}$ $\hat{A}_F(-t+1,0) \leq (\frac{2+\hat{p}}{3})t$. This follows by substituting the values of t, |B| and n in the equation above,

$$A_{F}(-t+1,0) \leq \hat{A}_{F}(-t+1,0) \leq \frac{n}{n_{0}}[\hat{p}t + (\phi L+2)|B| + 1] \leq \frac{2+\hat{p}}{1+2\hat{p}}[\hat{p} + \frac{(1-\hat{p})}{3}]t \quad (\text{since } (\frac{1-\hat{p}}{6}t > 1) \leq \frac{2+\hat{p}}{1+2\hat{p}} \cdot \frac{1+2\hat{p}}{3}t = \left(\frac{2+\hat{p}}{3}\right)t.$$
(58)

It then follows that,

$$\mathbb{P}\left(A_F(-t+1,0) \ge \left(\frac{2+\hat{p}}{3}\right)t, L(0) = -t\right)$$
$$\le \mathbb{P}\left(|B| > \frac{1-\hat{p}}{6(\phi L+2)}t\right)$$
$$\le e^{-nt\phi I_B\left(\frac{\hat{p}}{\phi}-\bar{a},\frac{1-\hat{p}}{6(\phi L+2)}\right)},$$
(59)

for all $n > N_5 = \max\{N_B(\frac{\hat{p}}{\phi} - \bar{a}, \frac{1-\hat{p}}{6(\phi L+2)}), \frac{H(2+\hat{p})}{1-\hat{p}}\}$ and $t > T_1 = \max\{T_B(\frac{\hat{p}}{\phi} - \bar{a}, \frac{1-\hat{p}}{6(\phi L+2)}), \frac{6}{1-\hat{p}}\}$. Moreover, we know that for each $i, X_F(i) = 1$ with probability greater than $1 - \left(\frac{n}{1-q}\right)^{7H} e^{-nI_X}$ for all $n > N_X$. Hence, using Lemma 1 we have that, there exists $N_6 > N_X$ such that,

$$\mathbb{P}\Big(X_F(-t+1,d) < (\frac{2+\hat{p}}{3})t, L(0) = -t\Big) \\
\leq \mathbb{P}\Big(X_F(-t+1,d) \le (\frac{2+\hat{p}}{3})(t+d), L(0) = -t\Big) \\
\leq e^{-n(t+d)(\frac{1-\hat{p}}{9})I_X} \\
\leq e^{-nt\frac{(1-\hat{p})I_X}{9}},$$
(60)

for all $n > N_6$ and t > 0.

Combining the above two results, from (59) and (60) we have, for all $n > N_7 = \max\{N_5, N_6\}$ and $t > T_1$,

$$\mathbb{P}\Big(A_F(-t+1,0) - X_F(-t+1,d) > 0, L(0) = -t\Big) \\
\leq 1 - (1 - e^{-nt\frac{(1+\hat{p})I_X}{9}})(1 - e^{-ntI_B(\frac{\hat{p}}{\phi} - \bar{a},\frac{1-\hat{p}}{6(\phi L+2)})}) \\
\leq 2e^{-tnI_{BX}},$$

where I_{BX} is the minimum of $I_B(\frac{\hat{p}}{\phi} - \bar{a}, \frac{1-\hat{p}}{6(\phi L+2)})$ and

 $\frac{(1-\hat{p})I_X}{9}.$ Recall that $t^* > \max\{T_1, \frac{I_0}{I_{BX}}\}$. Hence, summing over all $t > t^*$ we have, for all $n > N_2 = \max\{N_7, \left\lceil \frac{\log 2}{I_{BX}} \right\rceil\}$

$$\begin{split} \sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t,0) - X_F(-t,d) > 0, L(0) = -t\Big) \\ &\leq \sum_{t=t^*}^{\infty} \mathbb{P}\Big(A_F(-t,0) - X_F(-t,d) > 0, L(0) = -t\Big) \\ &\leq \sum_{t=t^*}^{\infty} 2e^{-ntI_{BX}} \\ &\leq \frac{2e^{-nt^*I_{BX}}}{1 - e^{-nI_{BX}}} \\ &\leq 4e^{-nt^*I_{BX}} \quad (\text{as } e^{-nI_{BX}} < \frac{1}{2}) \\ &< 4e^{-nI_0}. \end{split}$$

The result of the theorem then follows.