On scheduling for minimizing end-to-end buffer usage over multihop wireless networks

V.J. Venkataramanan and Xiaojun Lin
Department of ECE
Purdue University
Email: {vvenkat,linx}@purdue.edu

Lei Ying
Department of ECE
Iowa State University
Email: leiying@iastate.edu

Sanjay Shakkottai
Department of ECE
The University of Texas at Austin
Email: shakkott@ece.utexas.edu

Abstract

While there has been much progress in designing backpressure based stabilizing algorithms for multihop wireless networks, end-to-end performance (e.g., end-to-end buffer usage) results have not been as forthcoming. In this paper, we study the end-to-end buffer usage (sum of buffer utilization along a flow path) over a network with general topology and with fixed, loop-free routes using a large-deviations approach. We first derive bounds on the best performance that any scheduling algorithm can achieve. Based on the intuition from the bounds, we propose a class of (backpressure-like) scheduling algorithms called $\alpha\beta$-algorithms. We show that the parameters α and β can be chosen such that the system under the $\alpha\beta$-algorithm performs arbitrarily closely to the best possible scheduler (formally the decay rate function for end-to-end buffer overflow is shown to be arbitrarily close to optimal in the large-buffer regime). We also develop variants which have the same asymptotic optimality property, and also provide good performance in the small-buffer regime. Our results are substantiated using both analysis and simulation.

1 Introduction

Scheduling is one of the most critical challenges in multihop wireless network design due to channel fading and wireless interference. A major breakthrough in this area is the back-pressure algorithm proposed in [1], which is throughput optimal, i.e., it can stabilize any traffic flows that can be stabilized by some other algorithms. Significant progress has been made in designing backpressure based algorithms to maximize the network throughput or network utility as a function of the throughput [2–7]. While these results provide stabilizing algorithms with throughput-optimality guarantees, they may not have the desired performance in terms of other end-to-end performance metrics such as end-to-end delay. For example, results in [8, 9] have demonstrated that the classic backpressure algorithm could lead to unnecessarily large delays in multihop networks.

Since many emerging applications of wireless networks, such as wireless mesh networks for public safety, wireless sensor networks for unmanned surveillance, and vehicular networks for accident warnings, require delay constrained communication for desired performance, we are interested in scheduling algorithms that not only guarantee the throughput but also have good delay performance. It has been observed in [10, 11] that scheduling algorithms that do not take queue length information into consideration perform much worse than queue-length-based algorithms. However, the delay analysis for queue-length-based algorithms is challenging because of the dependence of the decision process on the queue length process. Existing results in the literature have focused on order optimality [10], the heavy traffic regimes [12–15] and the large-queue regimes [11, 16–18], but they only consider single-hop flows. In fact, the coupled arrival/departure processes (the departures from the previous hop become the arrivals of current hop) of multihop traffic flows have aggravated the difficulty in analyzing the end-to-end delay performance.

In this paper, we consider a multihop wireless network with general topology and with fixed, loop-free routes. We assume that a node maintains a separate queue for each flow passing through it. To simplify the analysis, we use the end-to-end buffer usage to approximate the end-to-end delay, and study the probability that the end-to-end buffer usage exceeds a certain threshold. Let $X_k^{\text{agg}}(t)$ denote the aggregated queue-length along the route of flow k and λ_k denote the average rate at which packets arrive at the source node of flow k. Mathematically, we are interested in characterizing

$$\mathbf{P}\left[\max_k \frac{X_k^{\text{agg}}(t)}{\lambda_k} \geq B\right]. \tag{1}$$

We call $X_k^{\text{agg}}(t)$ the end-to-end buffer usage of flow k, which is closely related to the end-to-end delay of flow k. For example, assuming that the packets arrive with a constant rate and all queues are FIFO, then $\frac{X_k^{\text{agg}}(t)}{\lambda_k}$ is the delay experienced by the packet of flow k that departs the system at time t (see [20]). In this case, a scheduling algorithm resulting in a small value of (1) guarantees that the probability that the end-to-end delays are larger than B will also be small.

We exploit large-deviations analysis to study this quantity. The main contributions of this paper include:

- We first derive bounds on the best performance that any scheduling algorithm can achieve. In other words, we obtain a $\tilde{\theta}_0$ such that
$$\liminf_{B\to\infty} \frac{1}{B} \log \mathbf{P}\left[\max_k \frac{X_k^{\text{agg}}(t)}{\lambda^k} \geq B\right] \geq -\tilde{\theta}_0$$
holds for any scheduling policy.

- We obtain two fundamental structural properties from this bound: *(i)* considering a single multihop flow, the scheduling algorithm should give preference to links that are closer to the destination; and *(ii)* considering multiple flows competing for a single multi-access channel (e.g., at the downlink of a single cell in a cellular network), the scheduling algorithm should give preference to those flows with the largest ratio of aggregate backlog to arrival rate.

- Based on the structural properties derived from the upper bound, we propose a class of (backpressure-like) scheduling algorithms called $\alpha\beta$-algorithms. Exploiting the large-deviations analysis, we show that the parameters α and β can be chosen such that the system under the $\alpha\beta$-algorithm performs arbitrarily closely to the best possible scheduler (formally the decay rate function for end-to-end buffer overflow probability is shown to be arbitrarily close to optimal in the large-buffer regime).

- Finally, we develop variants of $\alpha\beta$-algorithms (called hybrid $\alpha\beta$-algorithm) that have the same asymptotic optimality property, and also provides good performance in the small-buffer regime. Our simulations demonstrate that the hybrid $\alpha\beta$-algorithm performs better than the classic back-pressure algorithm.

2 System Model

2.1 Traffic Model

We study a multihop network consisting of N nodes, L links, and K multihop flows. Denote by $b(l)$ and $e(l)$ the beginning and ending nodes of link l. The route for each flow is fixed and loop-free. Denote by D^k the number of nodes through which flow k passes, $n^k(i)$ the i^{th} node in flow k's path, $l^k(i)$ the i^{th} link in flow k's path, and $\mathcal{K}_l$ the set of flows that traverse link l. Furthermore, we assume that time is slotted and let $A^k(t)$ denote the amount of data flow k injects to source node $n^k(1)$ at time t. We assume that $A^k(t)$ are identically and independently distributed (i.i.d.) across time slots and $\mathbb{E}[A^k(t)] = \lambda^k$. We assume that the average arrival rates are within the capacity region of the network (hence the system is stationary and ergodic) and that the arrival processes $A^k(t)$'s satisfy a large-deviations principle with rate function $L^k(\cdot)$ as defined in [20].

2.2 Channel Model

We consider a multihop wireless network in this paper, where the links experience interference and fading. Assume that the channel state stays constant over each time-slot, and changes at the beginning of each time slot. We let $C(t)$ denote the channel state at time-slot t, which is i.i.d. across time slots and takes values from $1, \dots, S$ with probabilities $p_1, \dots, p_S$.

Now given that $C(t) = j$, F_j^l denotes the units of data that can be transmitted over link l if there is no interference from other links, and $\mathcal{A}_j$ denotes the collection of subsets of links that do not interfere with each other when transmitting simultaneously. An element of $\mathcal{A}_j$ is called a schedule, which is a length-L binary vector. Consider a schedule $\vec{a} \in \mathcal{A}_j$, then $a_l = 1$ indicates that link l transmits under schedule $\vec{a}$. Note that if $(a_1, \dots, a_l, 1, \dots, a_L)$ is a possible schedule, then so is $(a_1, \dots, a_l, 0, \dots, a_L)$. Define the sets $\hat{\mathcal{E}}_j = \{(a_1 F_j^1, \dots, a_L F_j^L) : \vec{a} \in \mathcal{A}_j\}$ and $\mathcal{E}_j = \{(\gamma_1 F_j^1, \dots, \gamma_L F_j^L) : 0 \le \gamma_i \le a_i \text{ for some } \vec{a} \in \mathcal{A}_j\}$.

2.3 Queueing

Each node maintains a separate queue for each flow. Let $X_i^k(t)$ denote the queue at node i for flow k, and let $E_l^k(C(t), \vec{X}(t))$ denote the units of data of flow k transmitted over link l in time-slot t. We also define $E_l(C(t), \vec{X}(t)) = \sum_{k \in \mathcal{K}_l} E_l^k(C(t), \vec{X}(t))$ to be the net amount of data transmitted over link l. Note that the vector $(E_1(j, \vec{X}(t)), \dots, E_L(j, \vec{X}(t)))$ must belong to the set $\mathcal{E}_j$. The queues for flow k evolve as follows:

$$
\begin{aligned}
X_{n^k(1)}^k(t+1) &= X_{n^k(1)}^k(t) + A^k(t) - E_{l^k(1)}^k(C(t), \vec{X}(t)) \\
X_{n^k(i)}^k(t+1) &= X_{n^k(i)}^k(t) + E_{l^k(i-1)}^k(C(t), \vec{X}(t)) - \\
&\quad E_{l^k(i)}^k(C(t), \vec{X}(t)) \\
&\quad \text{for } i = 2, \dots, D^k - 1 \\
X_{n^k(D^k)}^k(t) &= 0 \text{ for all time } t \\
X_n^k(t) &= 0 \text{ for all other nodes } n \text{ and all time } t
\end{aligned}
$$

Here, we implicitly assume that $E_l^k(C(t), \vec{X}(t))$ cannot be larger than the available amount of data at the node $b(l)$.

3 Objective, Main Results and Intuition

3.1 Objective

The goal of a scheduling algorithm is to determine $E_l^k(C(t), \vec{X}(t))$ subject to fading and interference constraints. In this paper, we are interested in desiging a scheduling algorithm that minimizes the following queue-overflow probability:

$$
\mathbf{P}\left[\max_{k=1,\dots,K} \frac{\sum_{i=1}^{D^k} X_{n^k(i)}^k(0)}{\lambda^k} > B\right], \tag{2}
$$

where $X_{n^k(i)}^k(0)$ denotes the queue-length at the steady state. As we have discussed in the introduction, the quantity in (2) is closely related to the end-to-end delays.

Since it is very difficult to precisely characterize the probability of queue overflow (2) for the general network model that we consider, we use the large-deviations theory to study its asymptotic decay-rate as $B \to \infty$. Specifically, define

$$
-I \triangleq \liminf_{B \to \infty} \frac{1}{B} \log \left(\mathbf{P}\left[\max_{k=1,\dots,K} \frac{(\sum_{i=1}^{D^k} X_{n^k(i)}^k(0))}{\lambda^k} > B\right]\right),
$$

$$
-J \triangleq \limsup_{B \to \infty} \frac{1}{B} \log \left(\mathbf{P}\left[\max_{k=1,\dots,K} \frac{(\sum_{i=1}^{D^k} X_{n^k(i)}^k(0))}{\lambda^k} > B\right]\right).
$$

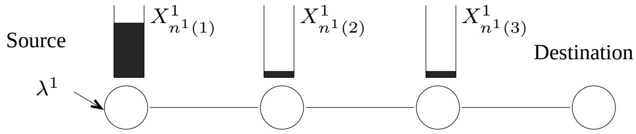

Figure 1: Tandem topology with single flow. The optimal algorithm gives preference to serving links near the destination. Hence most of the buffering occurs at source node.

The significance of studying these quantities lies in the following approximations:

$$\mathbf{P}[M(\vec{X}) > B] \leq e^{-JB+o(B)}$$
$$\mathbf{P}[M(\vec{X}) > B] \geq e^{-IB+o(B)},$$

where $M(\vec{X}) \triangleq \max_{k=1,\ldots,K} \frac{(\sum_{i=1}^{D^k} X^k_{n^k(i)})}{\lambda^k}$, and I and J are upper and lower bounds, respectively, of the asymptotic decay rates.

3.2 Two Structural Principles and The Class of $\alpha\beta$ Algorithms

In section 4.1, we will derive an upper bound on the best decay rate (I). By analyzing this upper bound, we obtain the two basic structural principles:

- Proposition 2 (in Section 4): Considering a tandem network with a single flow, the optimal algorithm should schedule links in such a way that links closer to the destination get preference in service. In other words, all buffering occurs as close to the source node as possible (see Figure 1). This result indicates the following structural principle:

  Principle 1: give preference to scheduling links closer to destination.

- Proposition 3 (in Section 4): Consider a cellular downlink topology where there is a single base station serving K cellular users. The optimal algorithm should schedule the user with the largest value of scaled queue backlog (i.e. $X^k_{n^k(1)}(t)/\lambda^k$) (see Figure 2). This result indicates the following structural principle:

  Principle 2: give preference to serving users with larger value of scaled aggregated queue-length.

Based on the observations above, we propose a class of scheduling algorithms (parametrized by α and β) called $\alpha\beta$-algorithms, which is similar to the back-pressure scheduling algorithm [1], but with different ways of defining link weights.

$\alpha\beta$-scheduling algorithm:

- For each flow k, we define $V^k(\vec{X}) = (\sum_{i=1}^{D^k}(X^k_{n^k(i)})^{(\alpha+1)})^{\frac{1}{\alpha+1}}$, and

$$W^k_l(t) = \frac{(V^k(\vec{X}(t)))^{\beta-\alpha}}{(\lambda_k)^{\beta+1}}\left((X^k_{b(l)}(t))^\alpha - (X^k_{e(l)}(t))^\alpha\right).$$

  The $\alpha\beta$-algorithm assigns a weight $W_l(t)$ to link l such that

$$W_l(t) = \max_{\{k \in \mathcal{K}_l\}} W^k_l(t).$$

- At each time slot, the $\alpha\beta$-scheduling algorithm computes an activation vector $\vec{a}^* \in \mathcal{A}_{C(t)}$ such that the vector $\vec{e}^* \triangleq (a^*_1 F^1_{C(t)}, \ldots, a^*_L F^L_{C(t)})$ satisfies

$$\sum_{l=1}^{L} W_l(t) e^*_l = \max_{\vec{e} \in \hat{\mathcal{E}}_{C(t)}} \sum_{l=1}^{L} W_l(t) e_l.$$

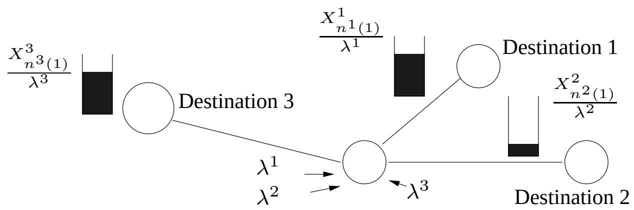

Figure 2: Cellular downlink topology with multiple flows. The optimal algorithm gives preference to serving users with larger value of scaled aggregated queue-length. In this example users 1 and 3 are preferred over user 2.

- If link $a_l^* = 1$, the scheduling algorithm activates link l and serves flow k_l^* with a rate $\min\{F^l_{C(t)}, X^{k_l^*}_{b(l)}(t)\}$, where the flow k_l^* satisfies

$$W_l^{k_l^*}(t) = W_l(t).$$

Note that the $\alpha\beta$-algorithm minimizes the drift of the Lyapunov function $\left(\sum_{k=1}^K (\frac{V^k(\vec{X})}{\lambda^k})^{(\beta+1)}\right)^{\frac{1}{\beta+1}}$ in a fluid scaled sense (see the proof of Proposition 4). When $\alpha = 1$ and $\beta = 1$, the $\alpha\beta$-algorithm is equivalent to the back-pressure algorithm [1].

□

The behavior of the $\alpha\beta$-scheduling algorithm:

- Consider a flow k in the network, and compare $W_l^k(t)$ over each hop l. Consider first the case when all $X_n^k(t) > 0$. Since $(V^k(\vec{X}(t)))^{\beta-\alpha}/(\lambda_k)^{\beta+1}$ is the same for all l, we denote the weight $W_l^k(t)$ to be

$$W_l^k(t) = \kappa\left((X^k_{b(l)}(t))^\alpha - (X^k_{e(l)}(t))^\alpha\right).$$

Note that the weight associated with the last-hop link is

$$W^k_{l^k(D^k-1)}(t) = \kappa(X^k_{n^k(D^k-1)}(t))^\alpha.$$

In the scheduling step, the link with the larger value of $W_l^k(t)$ will have the preference to be activated. When $\alpha \to 0$, the weight of the last-hop link $W^k_{l^k(D^k-1)}(t) = \kappa(X^k_{n^k(D^k-1)}(t))^\alpha$ will dominate the weights, $W_l^k(t)$, of all other links since

$$\lim_{\alpha\to 0}\left((X^k_{b(l)}(t))^\alpha - (X^k_{e(l)}(t))^\alpha\right) = 0,$$

for all links before the last hop, and

$$\lim_{\alpha\to 0}(X^k_{n^k(D^k-1)}(t))^\alpha = 1.$$

Hence, for $\alpha \to 0$, whenever the node $n^k(D^k-1)$ has packets to transmit, it will get preference to do so.

Similarly, if $X^k_{n^k(D^k-1)} = 0$ and $X^k_{n^k(D^k-2)} > 0$, then the weight $W^k_{l^k(D^k-2)}(t)$ will dominate all the weights associated with other links, and node $n^k(D^k-2)$ will get preference in service.

In summary, links closer to the destination get preference in service, so the $\alpha\beta$-scheduling algorithm satisfies Principle 1.

- Now, consider a network where single-hop flows compete for a single multiaccess channel l (e.g., in the downlink of a cell). In this case, the link weight $W_l^k(t)$ simplifies to $(X^k_{n^k(1)}(t))^\beta$ (since the backlog at the destination node is 0). The user with the largest value of $\frac{(X^k_{n^k(1)}(t))^\beta}{(\lambda^k)^{\beta+1}}F^{l^k(1)}_{C(t)}$ will be served. Equivalently, the user with the largest value of $\frac{X^k_{n^k(1)}(t)}{(\lambda^k)^{1+\frac{1}{\beta}}}(F^{l^k(1)}_{C(t)})^{\frac{1}{\beta}}$ will be served. When β is very large, the user with the largest value of $\frac{X^k_{n^k(1)}(t)}{\lambda^k}$ among the set of users with $F^{l^k(1)}_{C(t)} > 0$ will have the priority to be served.

In other words, the users with larger values of scaled backlogs will get preferences to be served, which satisfies Principle 2.

□

In Section 4, we will exploit large-deviations theory to analyze the performance of the class of $\alpha\beta$-scheduling algorithms, and show that this class of algorithms yield the optimal decay rate as $\alpha \to 0$ and $\beta \to \infty$. Letting $\mathbf{P}^{\alpha\beta}$ represent the stationary probability under the $\alpha\beta$-algorithm, and $\mathbf{P}^{\pi}$ be the stationary probability under any scheduling algorithm π, we will prove the following result:

Main Result (Proposition 5): Considering the class of $\alpha\beta$-algorithms, we have

$$\lim_{\alpha\to 0,\beta\to\infty} \limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K}(\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)) > B])$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K}(\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)) > B]),$$

which implies that by choosing α sufficiently small and β sufficiently large, the decay rate of

$$\mathbf{P}\left[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B\right]$$

for the $\alpha\beta$-algorithm can be made arbitrarily close to the optimal decay rate.

4 Analysis

In this section, we will first derive an upper bound on the decay rate (I) achievable by any scheduling algorithm. Then, by studying the bound under some specific topologies, we will obtain the two structure principles (Principles 1 and 2) that lead to the $\alpha\beta$-scheduling algorithm. Finally, we will prove that the class of $\alpha\beta$-algorithms is asymptotically optimal.

4.1 An upper bound on the decay-rate

We consider the optimization problem $\tilde{w}(\vec{\phi}, \vec{f})$ defined in (3) below. The quantity f^k can be interpreted as the long term average rate at which data arrive for flow k, ϕ_j can be interpreted as the long term fraction of time-slots for which the channel is in state j, the quantity $\theta^k_{l,j}$ can be interpreted as the long term fraction of time that service is given to flow k over link l in channel state j, $\gamma_{\vec{e},j}$ can be interpreted as the long term fraction of time that the schedule $\vec{e}$ is used when channel state is j, and x^k_n is the average rate of growth of the backlog of flow k

at node n.

$$
\begin{aligned}
\tilde{w}(\vec{\phi}, \vec{f}) \quad &= \min_{\{x^k_{n^k(i)}, \theta^k_{l^k(i),j}, \gamma_{\vec{e},j}\}} \max_{k=1,\ldots,K} \left(\frac{\sum_{i=1}^{D^k} x^k_{n^k(i)}}{\lambda^k}\right) \\
\text{subject to} \quad & x^k_{n^k(1)} = [f^k \\
& \qquad - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} \theta^k_{l^k(1),j} e_{l^k(1)}]^+ \\
& x^k_{n^k(i)} = [\sum_{j=1}^{S} \phi_j \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} \theta^k_{l^k(i-1),j} e_{l^k(i-1)} \\
& \qquad - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} \theta^k_{l^k(i),j} e_{l^k(i)}]^+ \\
& \qquad \text{for } i = 2, \ldots, D^k - 1 \\
& \sum_{\{\vec{e} \in \mathcal{E}_j\}} \gamma_{\vec{e},j} = 1, \sum_{\{k \in \mathcal{K}_l\}} \theta^k_{l,j} = 1 \\
& \gamma_{\vec{e},j} \geq 0 \text{ for all } \vec{e}, j \\
& \theta^k_{l,j} \geq 0 \text{ for all } k, l, j.
\end{aligned} \tag{3}
$$

For any given $\vec{\phi}$ and $\vec{f}$, the solution to the optimization problem $\tilde{w}(\vec{\phi}, \vec{f})$ is the optimal long term scheduling assignments $\theta^k_{l,j}$ and $\gamma_{\vec{e},j}$ such that the long term average rate of the growth of $M(\vec{X})$ (i.e. $\max_{k=1,\ldots,K}(\frac{\sum_{i=1}^{D^k} x^k_{n^k(i)}}{\lambda^k})$) is minimized. Interpreting this in another way, $\tilde{w}(\vec{\phi}, \vec{f})$ is a lower bound on the average rate of growth of $M(\vec{X})$ for any algorithm. If there exists an optimal algorithm that can attain this lower bound on the average rate of growth, then it must use the long-term scheduling assignment that corresponds to the solution of $\tilde{w}(\vec{\phi}, \vec{f})$. We will soon use this optimization problem to obtain basic insights about the behavior of this *optimal* algorithm.

Note that since $\vec{\lambda}$ is in the capacity region, if $\phi_j = p_j$ and $f^k = \lambda^k$ then the long term average rate of growth of queue backlogs will be zero for any *throughput-optimal* scheduling algorithm (such as the back-pressure algorithm). Hence it will be zero for the *optimal* algorithm. i.e. $\tilde{w}(\vec{p}, \vec{\lambda}) = 0$. Therefore, to make $M(\vec{X})$ exceed a large value B, the long term average channel probability must be $\vec{\phi} \neq \vec{p}$ and/or $\vec{f} \neq \vec{\lambda}$. As is typical of large deviations results, this deviation from the norm is associated with a cost. The cost for $\vec{\phi}$ to deviate from $\vec{p}$ per unit time is given by the relative entropy function $H(\vec{\phi}||\vec{p}) = \sum_{j=1}^{S} \phi_j \log(\frac{\phi_j}{p_j})$. The cost for $\vec{f}$ to deviate from $\vec{\lambda}$ per unit time is given by $L(\vec{f}) \triangleq \sum_{k=1}^{K} L^k(f^k)$. Consider a time-scale of length $\frac{1}{\tilde{w}(\vec{\phi},\vec{f})}$ over which the deviation from the norm is given by $\vec{\phi}$ and $\vec{f}$. Since $\tilde{w}(\vec{\phi}, \vec{f})$ is the average rate of growth for the optimal algorithm, the system must overflow after the time period $\frac{1}{\tilde{w}(\vec{\phi},\vec{f})}$ under all algorithms. Hence the corresponding cost, given by $\frac{1}{\tilde{w}(\vec{\phi},\vec{f})}[H(\vec{\phi}||\vec{p}) + L(\vec{f})]$, provides an upper bound on the minimum cost to overflow for any algorithm. Minimizing this value over $\vec{\phi}$ and $\vec{f}$, we then obtain the tightest upper bound given below. Please refer to [20, Proposition 9] for a more technical and precise discussion.

Define

$$
\tilde{\theta}_0 \quad = \inf_{\{\vec{\phi}, \vec{f} : \tilde{w}(\vec{\phi}, \vec{f}) > 0\}} \frac{1}{\tilde{w}(\vec{\phi}, \vec{f})} [H(\vec{\phi}||\vec{p}) + L(\vec{f})] \tag{4}
$$

Proposition 1 *For any scheduling policy π, we have*

$$
\liminf_{B \to \infty} \frac{1}{B} \log \mathbf{P}^{\pi}[\max_{k=1,\ldots,K} (\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)) \geq B] \geq -\tilde{\theta}_0
$$

Proof: The proof follows from the proof of Proposition 9 in [20] by setting

$$V(\vec{x}) \triangleq \max_{k=1,\ldots,K}\left(\frac{\sum_{i=1}^{D^k} x^k_{n^k(i)}}{\lambda^k}\right). \tag{5}$$

The only assumptions made on $V(\vec{x})$ for the proof to hold is that it be continuous and satisfy the following assumptions:

$V(\vec{x})$ is increasing in each component x_i and
$V(\vec{x}_1 + \vec{x}_2) \leq V(\vec{x}_1) + V(\vec{x}_2)$ for any two vectors $\vec{x}_1 \geq 0$ and $\vec{x}_2 \geq 0$.

It is easy to verify that these assumptions are true for (5). *Q.E.D.*

Proposition 1 shows that the decay rate I of any scheduling algorithm is less than $\tilde{\theta}_0$.

4.2 Insights into the behaviour of optimal algorithm

Let us use a heuristic interpretation of the optimization problem $\tilde{w}(\vec{\phi}, \vec{f})$ to obtain insights into the behavior of the *optimal* algorithm. We will establish the two principles (Proposition 2) and (Proposition 3) by studying a single flow tandem network and a multiflow single-hop cellular downlink network.

4.2.1 Tandem topology (Figure 1)

Consider a tandem topology with a single flow in the network, i.e., $K = 1$ and $L = D^1 - 1$. For simplicity, assume that at most one link can be activated in a time slot. The optimization problem $\tilde{w}(\vec{\phi}, \vec{f})$ simplifies to $\tilde{w}_{\text{tandem}}(\vec{\phi}, \vec{f})$:

$$\begin{aligned}
\tilde{w}_{\text{tandem}}(\vec{\phi}, \vec{f}) \quad &= \quad \min_{\{x^1_{n^1(i)}, \gamma_{\vec{e},j}\}} \frac{\sum_{i=1}^{D^1} x^1_{n^1(i)}}{\lambda^1} \qquad (6)\\
\text{subject to} \quad & \quad x^1_{n^1(1)} = [f^1 - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(1)}]^+ \\
& \quad x^1_{n^1(i)} = [\sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(i-1)} \\
& \qquad\qquad - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(i)}]^+ \\
& \qquad\qquad \text{for } i = 2, \ldots, D^1 - 1 \\
& \quad \sum_{\{\vec{e}\in\mathcal{E}_j\}} \gamma_{\vec{e},j} = 1 \\
& \quad \gamma_{\vec{e},j} \geq 0 \text{ for all } \vec{e},\, j.
\end{aligned}$$

Proposition 2 *One of the optimal solutions to the optimization problem $\tilde{w}_{tandem}(\vec{\phi}, \vec{f})$ has the following property*

$$\begin{aligned}
x^1_{n^1(i)} \quad &= \quad 0 \textit{ for } i = 2, \ldots, D^k - 1 \qquad (7)\\
x^1_{n^1(1)} \quad &\geq \quad 0
\end{aligned}$$

Proof: Since only one link can be active in a time slot, we have

$$\begin{aligned}
\hat{\mathcal{E}}_j \quad &= \quad \{(0, \ldots, F_j^{l^1(i)}, \ldots, 0) : \; i = 1, \ldots, D^k - 1\} \\
& \quad \cup \{(0, \ldots, 0)\}.
\end{aligned}$$

To prove this proposition, it is sufficient to show that for any feasible assignment $x^1_{n^1(i)}, \gamma_{\vec{e},j}$, there exists another assignment $\hat{x}^1_{n^1(i)}, \hat{\gamma}_{\vec{e},j}$ that satisfies the condition (7) and is such that $\sum_{i=1}^{D^1} \hat{x}^1_{n^1(i)} \leq \sum_{i=1}^{D^1} x^1_{n^1(i)}$ (i.e. its objective function value (6) is no greater than the original assignment).

To show this, consider any $i \geq 2$ such that $x^1_{n^1(i)} > 0$. We can reduce the value of $\gamma_{(\ldots,F_j^{l^1(i-1)},\ldots),j}$ and increase the value of $\gamma_{(0,\ldots,0),j}$, i.e., reducing the fraction of time spent on serving link $l^1(i-1)$. This will reduce the value of $\sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(i-1)}$ and

$$x^1_{n^1(i)} = [\sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(i-1)} - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^1(i)}]^+.$$

On the other hand, $x^1_{n^1(i-1)}$ may increase. If $x^1_{n^1(i-1)}$ increases, it will increase by at most the same amount by which $x^1_{n^1(i)}$ reduces as long as $x^1_{n^1(i)} > 0$. We can perform this operation until $x^1_{n^1(i)} = 0$. Hence, in this process, the sum $x^1_{n^1(i-1)} + x^1_{n^1(i)}$ either stays the same or decreases.

Applying the above procedure starting from the node $n^1(D^1 - 1)$ and working backward to node $n^1(2)$, it is easy to see that there exists an assignment $\hat{\gamma}_{\vec{e},j}$ resulting in values $\hat{x}^1_{n^1(i)}; i = 1, \ldots, D^1$ such that $\sum_{i=1}^{D^1} \hat{x}^1_{n^1(i)} \leq \sum_{i=1}^{D^1} x^1_{n^1(i)}$ and $\hat{x}^1_{n^1(i)} = 0$ for $i = 2, \ldots, D^1 - 1$, which leads to the proposition. *Q.E.D.*

The significance of Proposition 2 is that it implies the scheduling algorithm should give preference to scheduling links that are closer to the destination node, and thus significant buffering occurs only at the source node.

4.2.2 Cellular topology (Figure 2)

Consider a cellular downlink with a single base station and K users. The base station can communicate directly with the users. In this network, we have $L = K$ and $D^k = 2$ for $k = 1, \ldots, K$ and $n^1(1) = n^2(1) = \ldots = n^k(1)$. Due to wireless interference, we assume that only one user can be served in a timeslot. The optimization problem $\tilde{w}(\vec{\phi}, \vec{f})$ simplifies to $\tilde{w}_{\text{cellular}}(\vec{\phi}, \vec{f})$:

$$\begin{aligned}
\tilde{w}_{\text{cellular}}(\vec{\phi}, \vec{f}) \quad &= \quad \min_{\{x^k_{n^k(i)}, \gamma_{\vec{e},j}\}} \max_{k=1,\ldots,K} (\frac{x^k_{n^k(1)}}{\lambda^k}) \\
\text{subject to} \quad & \quad x^k_{n^k(1)} = [f^k \\
& \qquad\qquad - \sum_{j=1}^{S} \phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j} e_{l^k(1)}]^+ \\
& \qquad\qquad \text{for } k = 1, \ldots, K \\
& \quad \sum_{\{\vec{e}\in\mathcal{E}_j\}} \gamma_{\vec{e},j} = 1 \\
& \quad \gamma_{\vec{e},j} \geq 0 \text{ for all } \vec{e},\ j.
\end{aligned}$$

Since only one link can be activated in a time slot, we have

$$\hat{\mathcal{E}}_j = \{(0, \ldots, F_j^{l^k(1)}, \ldots, 0) : \ k = 1, \ldots, K\} \cup \{(0, \ldots, 0)\}.$$

Proposition 3 *One of the optimal solutions to the optimization problem $\tilde{w}_{cellular}(\vec{\phi}, \vec{f})$ has the following property: $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} = 0$ for any user r such that $\frac{x^r_{n^r(1)}}{\lambda^r} < \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$.*

Proof: Assume that the optimal solution to $\tilde{w}_{\text{cellular}}(\vec{\phi}, \vec{f})$ is such that there exists some user r such that $\frac{x^r_{n^r(1)}}{\lambda^r} < \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$, $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} > 0$.

We will show that it is possible to maintain the value of $\max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$ while modifying $\gamma_{\vec{e}}$ in such a way that $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} = 0$ for any user r such that $\frac{x^r_{n^r(1)}}{\lambda^r} < \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$. To achieve this, we reduce the value of $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j}$ and increase the value of $\gamma_{(0,\ldots,0),j}$. In other words, we reduce the service given to user r, and increase the time that the scheduler spends idling. Due to this, $\frac{x^r_{n^r(1)}}{\lambda^r}$ will increase. This is done till we end up with either $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} = 0$ and $\frac{x^r_{n^r(1)}}{\lambda^r} < \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$, or $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} \geq 0$ and $\frac{x^r_{n^r(1)}}{\lambda^r} = \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$. Note that the value of $\max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$ is not affected by this procedure.

Therefore, there is an optimal solution that satisfies the property that $\gamma_{(\ldots,F_j^{l^r(1)},\ldots),j} = 0$ for any user r with $\frac{x^r_{n^r(1)}}{\lambda^r} < \max_{k=1,\ldots,K}(\frac{x^k_{n^k(1)}}{\lambda^k})$. *Q.E.D.*

The significance of proposition 3 is that it tells us that the optimal value of $\tilde{w}_{\text{cellular}}(\vec{\phi}, \vec{f})$ is achieved by serving the users with the largest average rate of growth of end-to-end backlog scaled by the average arrival rate. In other words, the optimal scheduling algorithm should give preference to those users with the largest ratio of end-to-end backlog to arrival rate.

4.3 Asymptotic optimality of the class of $\alpha\beta$-algorithms

Based on the two principles above, we propose the class of $\alpha\beta$ scheduling algorithms as described in Section 3.2. We will use the Lyapunov-function-based large deviations approach developed in [20] to show that this class of algorithms is asymptotically optimal as $\alpha \to 0$ and $\beta \to \infty$.

Next, we first show that the $\alpha\beta$ algorithm is large deviations decay rate optimal for minimizing $\mathbf{P}[V(\vec{X}(0)) > B]$, where $V(\cdot)$ is a Lyapunov function defined to be:

$$V(\vec{X}) = \left(\sum_{k=1}^{K}(\frac{V^k(\vec{X})}{\lambda^k})^{(\beta+1)}\right)^{\frac{1}{\beta+1}}. \tag{8}$$

We can show that the $\alpha\beta$-algorithm minimizes the drift of this Lyapunov function in a fluid-sample-path sense (see the proof of Proposition 4 in the Appendix).

Since $\lim_{\alpha\to 0,\beta\to\infty} V(\vec{X}) = \max_{k=1,\ldots,K} \frac{(\sum_{i=1}^{D^k} X^k_{n^k(i)})}{\lambda^k}$, the function $V(\vec{X})$ can be viewed as an approximation of our objective function $M(\vec{X})$ with the parameters α and β controlling the degree of approximation.

Letting $\mathbf{P}^{\alpha\beta}$ represent the stationary probability under the $\alpha\beta$-algorithm and $\mathbf{P}^{\pi}$ be the stationary probability under any scheduling algorithm π, we first have the following proposition.

Proposition 4 *The quantity*

$$\lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B])$$

exists and for any scheduling policy π,

$$\lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) \leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[V(\vec{X}(0)) > B]).$$

Please refer to Appendix A for the proof.

Now we proceed to show that the class of $\alpha\beta$-algorithms is asymptotically optimal in terms of the large deviations decay rate for the stationary probability $\mathbf{P}[M(\vec{X}(0)) > B]$.

Proposition 5 *Considering the $\alpha\beta$-scheduling algorithm, we have*

$$\lim_{\alpha\to 0,\beta\to\infty} \limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K}(\frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k}) > B])$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K}(\frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k}) > B]).$$

Proof: By an application of Holder's inequality, we have the following inequality.

$$\frac{1}{N^{\frac{\alpha}{\alpha+1}}} \sum_{i=1}^{D^k} X^k_{n^k(i)} < (\sum_{i=1}^{D^k} (X^k_{n^k(i)})^{\alpha+1})^{\frac{1}{\alpha+1}} = V^k(\vec{X})$$

Also note that the following inequality holds.

$$\max_{k=1,\ldots,K} \frac{V^k(\vec{X}(t))}{\lambda^k} < (\sum_{k=1}^{K} (\frac{V^k(\vec{X}(t))}{\lambda^k})^{\beta+1})^{\frac{1}{\beta+1}} = V(\vec{X})$$

Combining the two inequalities, we have

$$\frac{1}{N^{\frac{\alpha}{\alpha+1}}} \max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}}{\lambda^k} < V(\vec{X}). \tag{9}$$

Therefore, for the $\alpha\beta$-algorithm,

$$\frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$$
$$< \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > \frac{B}{N^{\frac{\alpha}{\alpha+1}}}])$$

and hence,

$$\limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \tag{10}$$
$$\leq \frac{1}{N^{\frac{\alpha}{\alpha+1}}} \lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]).$$

We also have the following inequality

$$V^k(\vec{X}(t)) = (\sum_{i=1}^{D^k} (X^k_{n^k(i)})^{\alpha+1})^{\frac{1}{\alpha+1}} < \sum_{i=1}^{D^k} X^k_{n^k(i)} \tag{11}$$

and the following inequality

$$V(\vec{X}) = (\sum_{k=1}^{K} \frac{(V^k(\vec{X}))^{\beta+1}}{\lambda^k})^{\frac{1}{\beta+1}} < K^{\frac{1}{\beta+1}} \max_{k=1,\ldots,K} \frac{V^k(\vec{X})}{\lambda^k}. \tag{12}$$

Using these inequalities, we have, for any scheduling policy π,

$$K^{\frac{1}{\beta+1}} \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[V(\vec{X}(0)) > B]) \tag{13}$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]).$$

Since by proposition 4, the $\alpha\beta$-algorithm is optimal in terms of decay rate of $V(\cdot)$, we have

$$\lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) \tag{14}$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[V(\vec{X}(0)) > B])$$

Thus combining equations (10),(13) and (14), we have

$$K^{\frac{1}{\beta+1}} N^{\frac{\alpha}{\alpha+1}}$$
$$\times \limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$$

Therefore, the $\alpha\beta$ algorithms are asymptotically optimal as $\alpha \to 0, \beta \to \infty$. *Q.E.D.*

Proposition 6 *Consider the $\alpha\beta$-scheduling algorithm. The limit*

$$\lim_{\alpha\to 0,\beta\to\infty} \lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$$

exists and is equal to $-\tilde{\theta}_0$.

Remark : This proposition shows that the bound $\tilde{\theta}_0$ on the decay rate as mentioned in Proposition 1 is tight. This is because by choosing α sufficiently small and β sufficiently large, the decay rate of the $\alpha\beta$-algorithm, $\lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$ can be made arbitrarily close to $-\tilde{\theta}_0$.

Proof: First, we will estimate $\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}}{\lambda^k}$ using $V(\vec{X})$. Using inequalities (9), (11) and (12), we can sandwich $\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}}{\lambda^k}$ by $V(\vec{X})$ in the following way

$$\frac{1}{K^{\frac{1}{\beta+1}}} V(\vec{X}) \leq \max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}}{\lambda^k} \leq N^{\frac{\alpha}{\alpha+1}} V(\vec{X}) \tag{15}$$

Therefore, for the $\alpha\beta$-algorithm, we obtain the following relations between the overflow probabilities $\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]$ and $\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]$.

$$\limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \tag{16}$$
$$\leq \frac{1}{N^{\frac{\alpha}{\alpha+1}}} \lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]).$$

and

$$K^{\frac{1}{\beta+1}} \lim_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) \tag{17}$$
$$\leq \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]).$$

Taking the limit as $\alpha \to 0$ and $\beta \to \infty$ on equations (16) and (17), we obtain the following inequalities on the asymptotic performance:

$$\begin{aligned}
&\limsup_{\alpha\to 0,\beta\to\infty} \limsup_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\dots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \\
&\quad \le \limsup_{\alpha\to 0,\beta\to\infty} \lim_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]).
\end{aligned} \tag{18}$$

and

$$\begin{aligned}
&\liminf_{\alpha\to 0,\beta\to\infty} \lim_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) \\
&\quad \le \liminf_{\alpha\to 0,\beta\to\infty} \liminf_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\dots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]).
\end{aligned} \tag{19}$$

Now we will show that the quantity $\lim_{\alpha\to 0,\beta\to\infty}\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B])$ exists and is equal to $-\tilde{\theta}_0$.

Similar to $\tilde{w}(\vec{\phi},\vec{f})$ defined in (3), define $\tilde{w}_1(\vec{\phi},\vec{f})$ as

$$\begin{aligned}
\tilde{w}_1(\vec{\phi},\vec{f}) \quad &= \quad \min_{\{x^k_{n^k(i)},\theta^k_{l^k(i),j},\gamma_{\vec{e},j}\}} V(\vec{x}) \\
\text{subject to} \quad & x^k_{n^k(1)} = [f^k \\
&\qquad - \sum_{j=1}^{S}\phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j}\theta^k_{l^k(1),j} e_{l^k(1)}]^+ \\
& x^k_{n^k(i)} = [\sum_{j=1}^{S}\phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j}\theta^k_{l^k(i-1),j} e_{l^k(i-1)} \\
&\qquad - \sum_{j=1}^{S}\phi_j \sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \gamma_{\vec{e},j}\theta^k_{l^k(i),j} e_{l^k(i)}]^+ \\
&\qquad \text{for } i = 2,\dots,D^k-1 \\
& \sum_{\{\vec{e}\in\mathcal{E}_j\}} \gamma_{\vec{e},j} = 1, \ \sum_{\{k\in\mathcal{K}_l\}} \theta^k_{l,j} = 1 \\
& \gamma_{\vec{e},j} \ge 0 \text{ for all } \vec{e},\ j \\
& \theta^k_{l,j} \ge 0 \text{ for all } k,l,j.
\end{aligned}$$

$\tilde{w}_1(\vec{\phi},\vec{f})$ is an optimization problem similar to $\tilde{w}(\vec{\phi},\vec{f})$, but with a different objective function.

By Proposition 8 of [20], we know that $\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) = -\theta_0$ and by the proof of Proposition 8 in [20], we know that $\theta_0 \ge \tilde{\theta}_1$ where $\tilde{\theta}_1$ is defined as

$$\tilde{\theta}_1 \quad = \quad \inf_{\{\vec{\phi},\vec{f}:\tilde{w}_1(\vec{\phi},\vec{f})>0\}} \frac{1}{\tilde{w}_1(\vec{\phi},\vec{f})}[H(\vec{\phi}||\vec{p}) + L(\vec{f})]$$

Hence we have

$$\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) \le -\tilde{\theta}_1.$$

Combining this with Proposition 9 of [20], we have

$$\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) = -\tilde{\theta}_1. \tag{20}$$

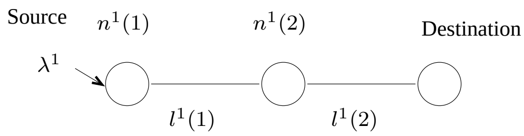

Figure 3: Three node tandem topology

Note that $\tilde{\theta}_0$ and $\tilde{\theta}_1$ are similar except that $\tilde{\theta}_0$ is defined using $\tilde{w}(\vec{\phi},\vec{f})$ and $\tilde{\theta}_1$ using $\tilde{w}_1(\vec{\phi},\vec{f})$.
We will show that as $\alpha \to 0$ and $\beta \to \infty$, $\tilde{w}_1(\vec{\phi},\vec{f}) \to \tilde{w}(\vec{\phi},\vec{f})$ and hence $\tilde{\theta}_1 \to \tilde{\theta}_0$.
From the inequality (15), it is easy to see that

$$\frac{1}{K^{\frac{1}{\beta+1}}}\tilde{w}_1(\vec{\phi},\vec{f}) \leq \tilde{w}(\vec{\phi},\vec{f}) \leq N^{\frac{\alpha}{\alpha+1}}\tilde{w}_1(\vec{\phi},\vec{f}),$$

which implies that

$$K^{\frac{1}{\beta+1}}\tilde{\theta}_1 \geq \tilde{\theta}_0 \geq \frac{1}{N^{\frac{\alpha}{\alpha+1}}}\tilde{\theta}_1.$$

Therefore, $\tilde{\theta}_0 \leq \liminf_{\alpha\to 0,\beta\to\infty} \tilde{\theta}_1 \leq \limsup_{\alpha\to 0,\beta\to\infty} \tilde{\theta}_1 \leq \tilde{\theta}_0$.
Hence, taking the limit as $\alpha \to 0$ and $\beta \to \infty$ in equation (20), we have

$$\lim_{\alpha\to 0,\beta\to\infty} \lim_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B]) = -\tilde{\theta}_0$$

Applying the above result to equations (18) and (19), we have

$$-\tilde{\theta}_0 \leq \liminf_{\alpha\to 0,\beta\to\infty} \liminf_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \leq$$
$$\limsup_{\alpha\to 0,\beta\to\infty} \limsup_{B\to\infty} \frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \leq -\tilde{\theta}_0$$

from which we deduce that $\lim_{\alpha\to 0,\beta\to\infty}\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[\max_{k=1,\ldots,K}\frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])$ exists and is equal to $-\tilde{\theta}_0$ *Q.E.D.*

5 Practical Scheduling algorithms with good delay performance

So far, we have seen that by choosing α sufficiently small and β sufficiently large, we can make the $\alpha\beta$-algorithm have a large deviations decay rate that is arbitrarily close to the optimal decay rate (Proposition 5). Note that large deviations behaviour deals with the regime of large buffer lengths. In this section, we will discuss how to design a hybrid algorithm that has the same large deviations decay rate performance of an $\alpha\beta$-algorithm while at the same time having a better performance in the regime of small buffer lengths. To achieve this goal, the hybrid algorithm emulates the behaviour of 1β-algorithm for small queue backlogs and the behaviour of $\alpha\beta$-algorithm for large queue backlogs.

Consider a three-node, two-link tandem network with one flow as shown in Fig. 3. Since there is only one flow, we will simplify some of the notation by not specifying the flow (e.g. we will use $X_{n(i)}$ instead of $X^k_{n^k(i)}$). Due to interference, only one link can be served. The scheduler must decide whether to serve $l(1)$ or $l(2)$. Consider the $\alpha\beta$-algorithm. The state space (i.e. a plot of $X_{n(1)}$ vs. $X_{n(2)}$) is divided into two regions by the line specified by $((X_{n(1)})^\alpha - (X_{n(2)})^\alpha)F^{l(1)}_{C(t)} = (X_{n(2)})^\alpha F^{l(2)}_{C(t)}$ (see Fig. 5). If the state (i.e. the ordered pair $(X_{n(1)}, X_{n(2)})$) falls in the region above the line, link $l(2)$ will be served. If the state falls in the region below the line, $l(1)$ will be

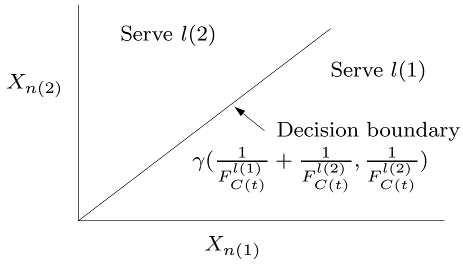

Figure 4: State space plot for 1β-algorithm. γ is a parameter used to define the decision boundary.

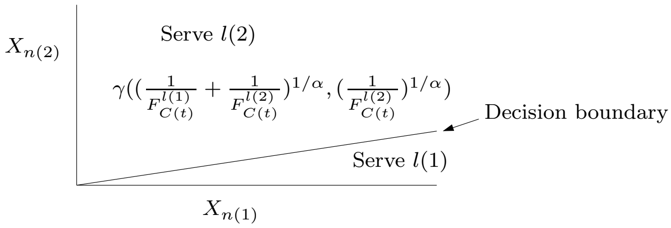

Figure 5: State space plot for $\alpha\beta$-algorithm. γ is a parameter used to define the decision boundary.

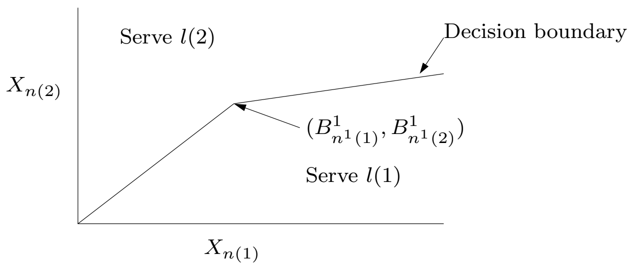

Figure 6: State space plot for hybrid algorithm.

served. In either case, as a consequence of being served, the state will move towards the decision boundary. For 1β-algorithm, we obtain the state space shown in Fig. 4 . For $\alpha\beta$-algorithm (with small α), we obtain the state space shown in Fig. 5. In the case of small α, because the decision line moves toward the x-axis, the state of the system tends to 'squeeze' out towards the right. Hence, for the $\alpha\beta$-algorithm (with small α), the state of the system tends to stay further away from the origin in comparison to the 1β-algorithm. This leads to larger values of $M(\vec{X}) = (X_{n(1)} + X_{n(2)})/\lambda$ (Note that this is not in contradiction to our theoretical result since our theoretical result is a result on the asymptotic rate of decay at large buffer levels).

To overcome this problem, we can construct a hybrid policy which behaves like 1β-algorithm for small buffer lengths and then switches to $\alpha\beta$-algorithm when the buffer lengths are larger. The state space for this hybrid policy is shown in Fig. 6. The decision boundary for the hybrid policy is composed of the decision boundaries shown in Fig. 4 and 5 with the point $(B^1_{n^1(1)}, B^1_{n^1(2)})$ being the point of 'concatenation'. The details of the hybrid algorithm are described next.

5.1 Hybrid algorithms

The hybrid algorithm uses the following function to assign weight to the links.

$$W_l(t) = \max_{\{k \in \mathcal{K}_l\}} (V^k(\vec{X}(t)))^{\beta-\alpha} [Z(X^k_{b(l)}(t)) - Z(X^k_{e(l)}(t))],$$

where for any flow k and i^{th} node in the path of flow k,

$$Z(X^k_{n^k(i)}) = \begin{cases} X^k_{n^k(i)} & \text{if } X^k_{n^k(i)} < B^k_{n^k(i)} \\ [(X^k_{n^k(i)} - B^k_{n^k(i)})^\alpha + B^k_{n^k(i)}] & \text{otherwise.} \end{cases}$$

$B^k_{n^k(i)}$ are threshold values which are constants.

The motivation behind the function $Z(\cdot)$ is as follows. For now, let us assume that we know the values of $B^k_{n^k(i)}$. Since we want the hybrid to emulate 1β-algorithm for small values of backlog, we want $Z(X^k_{n^k(i)}) = X^k_{n^k(i)}$ when $X^k_{n^k(i)}$ is less than a certain threshold $B^k_{n^k(i)}$. Further, since we want to emulate $\alpha\beta$-algorithm with small value of α when the backlog is large, we want $Z(X^k_{n^k(i)}) = (X^k_{n^k(i)})^\alpha$ when $X^k_{n^k(i)} > B^k_{n^k(i)}$. However, a problem with these two choices is that the function $Z(X^k_{n^k(i)})$ is not continuous at $B^k_{n^k(i)}$. Hence we modify the value of $Z(X^k_{n^k(i)})$ for $X^k_{n^k(i)} > B^k_{n^k(i)}$ to $[(X^k_{n^k(i)} - B^k_{n^k(i)})^\alpha + B^k_{n^k(i)}]$.

Now we look at one way of choosing the values $B^k_{n^k(i)}$. Consider again the single flow tandem network of Fig. 3. From the definition of the hybrid policy, we see that when $X_{n(1)} > B^1_{n^1(1)}$ and $X_{n(2)} > B^1_{n^1(2)}$, the decision boundary is determined by the line

$$\begin{aligned} &\Big((X_{n(1)} - B^1_{n^1(1)})^\alpha + B^1_{n^1(1)} \\ &-(X_{n(2)} - B^1_{n^1(2)})^\alpha - B^1_{n^1(2)}\Big) F^{l(1)}_{C(t)} \\ &\qquad = \Big((X_{n(2)} - B^1_{n^1(2)})^\alpha + B^1_{n^1(2)}\Big) F^{l(2)}_{C(t)}. \end{aligned}$$

With some algebra, it can be shown that the line can be expressed as

$$\begin{aligned} &((\gamma(\frac{1}{F^{l(1)}_{C(t)}} + \frac{1}{F^{l(2)}_{C(t)}}) - B^1_{n^1(1)})^{\frac{1}{\alpha}} + B^1_{n^1(1)}, \\ &\qquad (\gamma(\frac{1}{F^{l(2)}_{C(t)}}) - B^1_{n^1(2)})^{\frac{1}{\alpha}} + B^1_{n^1(2)}) \end{aligned} \tag{21}$$

using a parameter γ. Since we want the decision boundary of the hybrid policy to look like figure 6, the line (21) should be of the form $\hat{\gamma}((\frac{1}{F^{l(1)}_{C(t)}} + \frac{1}{F^{l(2)}_{C(t)}})^{\frac{1}{\alpha}}, (\frac{1}{F^{l(2)}_{C(t)}})^{\frac{1}{\alpha}}) + (K_1, K_2)$,where $\hat{\gamma}$ is a parameter and K_1 and K_2 are constants. This provides us with the following solution for $(B^1_{n^1(1)}, B^1_{n^1(2)})$.

$$(B^1_{n^1(1)}, B^1_{n^1(2)}) = \kappa(1/F^{l(1)}_{C(t)} + 1/F^{l(2)}_{C(t)}, 1/F^{l(2)}_{C(t)})$$

where κ is some constant. Generalizing this idea to a single flow tandem network with D^1 nodes, we obtain

$$
\begin{aligned}
&(B^1_{n^1(1)}, B^1_{n^1(2)}, \ldots, B^1_{n^1(D^1-1)}) \\
&= \kappa(\sum_{i=1}^{D^1-1} \frac{1}{F^{l^1(i)}_{C(t)}}, \sum_{i=2}^{D^1-1} \frac{1}{F^{l^1(i)}_{C(t)}}, \ldots, \frac{1}{F^{l^1(D^1-1)}_{C(t)}})
\end{aligned}
$$

Note that the threshold values $B^1_{n^1(i)}$ depend on the channel state $C(t)$. We impose an additional constraint that the sum of the thresholds should be constant, i.e. for any channel state, $\sum_{i=1}^{D^1-1} B^1_{n^1(i)} = B^*$. This gives us $\kappa = \frac{B^*}{\sum_{q=1}^{D^1-1}\sum_{p=q}^{D^1-1}\frac{1}{F^{l^1(p)}_{C(t)}}}$.

Extending this idea to our general system model, we obtain the following expression for $B^k_{n^k(i)}$

$$
B^k_{n^k(i)} = B^* \frac{\sum_{p=i}^{D^k-1}\frac{1}{F^{l^k(p)}_{C(t)}}}{\sum_{q=1}^{D^k-1}\sum_{p=q}^{D^k-1}\frac{1}{F^{l^k(p)}_{C(t)}}}
$$

where B^* is a user-defined parameter.

Since the values $B^k_{n^k(i)}$ are constants, the hybrid algorithm has the same large deviations behaviour as the $\alpha\beta$-algorithm. Formally speaking, we have the following result.

Proposition 7 *Let $\mathbf{P}^{\alpha\beta}_{hyb}$ denote the stationary probability for the hybrid policy with parameters α and β. Then,*

$$
\begin{aligned}
&\lim_{\alpha\to 0,\beta\to\infty} \limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}_{hyb}[\max_{k=1,\ldots,K}(\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)) > B]) \\
&\le \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K}(\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)) > B]).
\end{aligned}
$$

Proof: Similar to the proof of Proposition 4, we can show that the hybrid policy with parameters α, β minimizes the drift of the Lyapunov function $V(\vec{X})$ in a fluid-sample-path sense. This is because the behavior of an algorithm for fluid-sample-paths is determined by its behavior when queue lengths are large. Since the hybrid algorithm behaves like the $\alpha\beta$-algorithm when queue lengths are large, it has the same behavior as the $\alpha\beta$-algorithm as far as fluid-sample paths are concerned.

Then, by using the arguments used in the proof of Proposition 5, we obtain

$$
\begin{aligned}
&K^{\frac{1}{\beta+1}} N^{\frac{\alpha}{\alpha+1}} \\
&\times \limsup_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\alpha\beta}_{\text{hyb}}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B]) \\
&\le \liminf_{B\to\infty} \frac{1}{B} \log(\mathbf{P}^{\pi}[\max_{k=1,\ldots,K} \frac{\sum_{i=1}^{D^k} X^k_{n^k(i)}(0)}{\lambda^k} > B])
\end{aligned}
$$

and hence the result. *Q.E.D.*

6 Simulation

For simulations, we consider the network shown in Fig. 7. The network consists of 10 nodes and 9 links. Due to fading, the capacity of each link takes the value 10 or 0 with a probability of 0.5. The fluctuations of the capacity are *i.i.d* over time and across links. In each time slot, due to interference only one link may be active. There

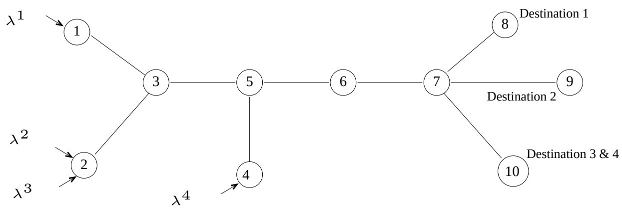

Figure 7: System topology for simulation

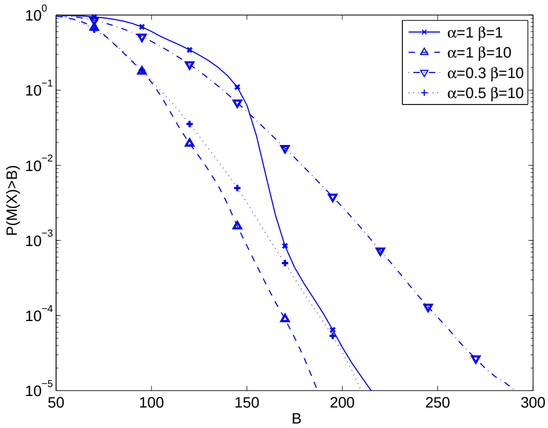

Figure 8: Plot of $\mathbf{P}[M(\vec{X}(0)) > B])$ vs. B for $\alpha\beta$-algorithm for various values of α and β.

are 4 flows traversing the network and data arrives at each flow according to a Bernoulli process. In each time slot, 1 unit of data arrives with a probability of 0.52 and no data arrives otherwise. The arrival process is i.i.d across flows. Flow 1 passes through nodes $1, 3, 5, 6, 7, 8$. Flow 2 passes through nodes $2, 3, 5, 6, 7, 9$. Flow 3 passes through nodes $2, 3, 5, 6, 7, 10$ and flow 4 passes through nodes $4, 5, 6, 7, 10$. The quantity we are interested in is the stationary probability $\mathbf{P}[M(\vec{X}(0)) > B])$ where $M(\vec{X})$ is given by

$$\begin{aligned}\frac{1}{0.52}\max\{&(X_1^1 + X_3^1 + X_5^1 + X_6^1 + X_7^1 + X_8^1),\\ &(X_2^2 + X_3^2 + X_5^2 + X_6^2 + X_7^2 + X_9^2),\\ &(X_2^2 + X_3^2 + X_5^2 + X_6^2 + X_7^2 + X_{10}^2),\\ &(X_4^4 + X_5^4 + X_6^4 + X_7^4 + X_{10}^4)\}\end{aligned}$$

In Fig.8 we present plots of $\mathbf{P}[M(\vec{X}(0)) > B])$ vs. B for the $\alpha\beta$-algorithm with four different choices for the pair of parameters α and β. We see that as expected from theory, $\alpha = 1$, $\beta = 10$ algorithm has a better performance than the $\alpha = 1$, $\beta = 1$ algorithm. However, as we decrease α, we see that $\alpha = 0.5$, $\beta = 10$ does worse than the $\alpha = 1$, $\beta = 10$ algorithm (but still marginally better than $\alpha = 1$, $\beta = 1$ algorithm). Further, $\alpha = 0.3$, $\beta = 10$ does much worse (although at much larger values of B (not shown), the decay rate will be better). In Fig.9 we compare the performance of the $\alpha = 1$, $\beta = 1$ algorithm, $\alpha = 1$, $\beta = 10$ algorithm, hybrid algorithm with $\alpha = 0.5$, $\beta = 10$, $B^* = 50$ and hybrid algorithm with $\alpha = 0.3$, $\beta = 10$, $B^* = 50$. The hybrid algorithms perform much better than either the $\alpha = 1$, $\beta = 1$ or the $\alpha = 1$, $\beta = 10$ algorithm. Also, both hybrid algorithms have similar performance which suggests that there is no advantage to decrease α further. Note that the $\alpha = 1$, $\beta = 1$ algorithm is the same as the well known back-pressure scheduling algorithm [1]. Hence, our simulation results show that the hybrid algorithm (for the case of small α and large β) and the $\alpha\beta$-algorithm (for the case of large β) perform much better than backpressure algorithm.

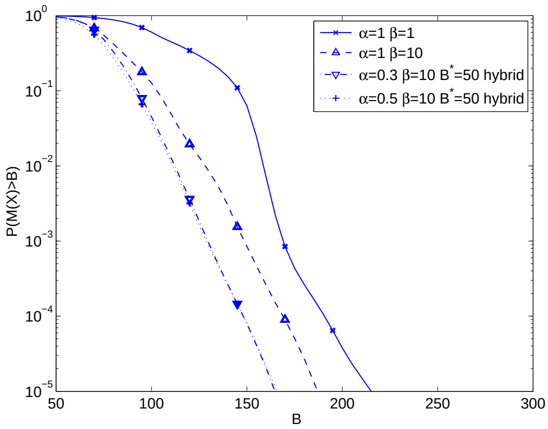

Figure 9: Plot of $\mathbf{P}[M(\vec{X}(0)) > B])$ vs. B for $\alpha\beta$-algorithm and hybrid algorithm for various values of α and β.

7 Conclusion

Using a large deviations framework, we obtain insights into the design of optimal algorithms for minimizing the end-to-end buffer usage in a multiflow multihop wireless network. We propose a class of algorithms (called $\alpha\beta$-algorithms) and variants (called hybrid $\alpha\beta$-algorithm) that can be made to perform arbitrarily close to optimal (in a large deviations sense) by reducing α and increasing β. Through simulations, we show that the class of hybrid algorithms has good performance in the small buffer regime as well. Our result is based on a very general system model and hence can provide insight in a wide range of scheduling scenarios.

A Proof of proposition 4

The result follows by applying Proposition 8 of [20]. To do this, we only need to verify the Assumptions 1-6 stated in [20]. Before we can verify the assumptions, we need to define a norm $||\cdot||$. Define the norm $||\cdot||$ as

$$||\vec{X}|| = \left(\sum_{k=1}^{K} (\frac{V_+^k(\vec{X})}{\lambda^k})^{(\beta+1)}\right)^{\frac{1}{\beta+1}} \quad \text{where} \tag{22}$$

$$V_+^k(\vec{X}) = \left(\sum_{i=1}^{D^k} (|X^k_{n^k(i)}|)^{(\alpha+1)}\right)^{\frac{1}{\alpha+1}}.$$

Lemma 8 *$||\cdot||$ defined in (22) satisfies the properties of a norm. i.e., $||c\vec{X}|| = |c|||\vec{X}||$ for any scalar c, $||\vec{X}|| = 0$ if and only if $\vec{X} = \vec{0}$ and $||\vec{X} + \vec{Y}|| \leq ||\vec{X}|| + ||\vec{Y}||$ for any two vectors $\vec{X}$ and $\vec{Y}$.*

Proof: It is easy to see that $||c\vec{X}|| = |c|||\vec{X}||$. It is also easy to see that $||\vec{X}|| = 0$ if and only if $\vec{X} = \vec{0}$. Next, the triangle inequality follows from application of Minkowski's inequality. By definition, $||\vec{X} + \vec{Y}|| =$

$(\sum_{k=1}^{K}(\frac{V_+^k(\vec{X}+\vec{Y})}{\lambda^k})^{(\beta+1)})^{\frac{1}{\beta+1}}$. We first show that the triangle inequality holds for $V_+^k(\cdot)$.

$$\begin{aligned} V_+^k(\vec{X}+\vec{Y}) &= (\sum_{i=1}^{D^k}(|X^k_{n^k(i)}+Y^k_{n^k(i)}|)^{(\alpha+1)})^{\frac{1}{\alpha+1}} \\ &\leq (\sum_{i=1}^{D^k}(|X^k_{n^k(i)}|+|Y^k_{n^k(i)}|)^{(\alpha+1)})^{\frac{1}{\alpha+1}} \\ &\leq (\sum_{i=1}^{D^k}(|X^k_{n^k(i)}|)^{(\alpha+1)})^{\frac{1}{\alpha+1}} \\ &\qquad +(\sum_{i=1}^{D^k}(|Y^k_{n^k(i)}|)^{(\alpha+1)})^{\frac{1}{\alpha+1}} \\ &= V_+^k(\vec{X})+V_+^k(\vec{Y}) \end{aligned}$$

where the second inequality follows by using the triangle inequality on $|\cdot|$, and the third follows by using Minkowski's inequality.

Using the triangle inequality for $V_+^K(\cdot)$ and Minkowski's inequality, we have the triangle inequality for $||\cdot||$ as follows.

$$\begin{aligned} & ||\vec{X}+\vec{Y}|| \\ =\ & (\sum_{k=1}^{K}(\frac{V_+^k(\vec{X}+\vec{Y})}{\lambda^k})^{(\beta+1)})^{\frac{1}{\beta+1}} \\ \leq\ & (\sum_{k=1}^{K}(\frac{V_+^k(\vec{X})}{\lambda^k}+\frac{V_+^k(\vec{Y})}{\lambda^k})^{(\beta+1)})^{\frac{1}{\beta+1}} \\ \leq\ & (\sum_{k=1}^{K}(\frac{V_+^k(\vec{X})}{\lambda^k})^{(\beta+1)})^{\frac{1}{\beta+1}}+(\sum_{k=1}^{K}(\frac{V_+^k(\vec{Y})}{\lambda^k})^{(\beta+1)})^{\frac{1}{\beta+1}} \\ =\ & ||\vec{X}||+||\vec{Y}||. \end{aligned}$$

Q.E.D.

The assumptions in [20] use certain concepts (stated in detail in [20]) that we will reintroduce in the following section.

A.1 Preliminaries

In this section, we summarize some important concepts necessary to understand the exposition that follows. The following processes play an important role in the large deviations analysis. They are obtained by scaling time and magnitude by a factor B. This scaling is commonly referred to as fluid scaling.

For a fixed B, define the scaled channel state process as (see equation (1) in [20])

$$s_j^B(t)=\frac{1}{B}\sum_{\tau=0}^{B(T+t)}\mathbf{1}_{\{C(\tau)=j\}}$$

for $t=\frac{m}{B}-T, m=0,\ldots,BT$, and by linear interpolation otherwise. Let $\vec{s}^B=[s_j^B, j=1,\ldots,S]$. Define the scaled arrival process as (see equation (4) in [20])

$$a^{k,B}(t)=\frac{1}{B}\sum_{\tau=0}^{B(T+t)}A^k(\tau)$$

for $t = \frac{m}{B} - T, m = 0, \ldots, BT$, and by linear interpolation otherwise. Let $\vec{a}^B = [a^{k,B}, k = 1, \ldots, K]$. Define the scaled backlog process as (see equation (12) in [20])

$$x_n^k(t) = \frac{1}{B} X_n^k(B(T+t))$$

for $t = \frac{m}{B} - T, m = 0, \ldots, BT$, and by linear interpolation otherwise. Let $\vec{x}^B(t) = [x_n^{k,B}(t), n = 1, \ldots, N, k = 1, \ldots, K]$.

We reintroduce the concept of fluid-sample-paths (FSP) (see section 3 of [20]). Given any $T > 0$ and initial condition $\vec{x}^B(-T)$, the scaled channel process $\vec{s}^B(t)$ and scaled arrival process $\vec{a}^B(t)$ determine the evolution of the scaled queue backlog process $\vec{x}^B(t)$ through the behavior of the scheduling algorithm. Define a triple $(\vec{s}^B(t), \vec{a}^B(t), \vec{x}^B(t))$ to consist of such related processes. Consider $B \to \infty$. Since $\vec{s}^B(t)$, $\vec{a}^B(t)$ and $\vec{x}^B(t)$ are Lipschitz continuous, for any sequence of triples $(\vec{s}^B(t), \vec{a}^B(t), \vec{x}^B(t))$ there must exist a subsequence that converges to a limiting triple $(\vec{s}, \vec{a}, \vec{x})$ uniformly over compact intervals. Any such limiting triple is referred to as a fluid-sample-path.

In what follows, we will often refer to derivatives $\frac{d}{dt}\vec{s}(t)$, $\frac{d}{dt}\vec{a}(t)$, $\frac{d}{dt}\vec{x}(t)$ and $\frac{d}{dt}V(\vec{x}(t))$. These derivatives exists almost everywhere due to the Lipschitz continuity of the processes $\vec{s}(t), \vec{a}(t), \vec{x}(t)$ and $V(\vec{x}(t))$. Time t is said to be regular if the derivatives exist at time t. For convenience, when referring to such derivatives, we implicitly assume that the derivative is taken at a *regular* time t. Further, for convenience we consider derivatives to be right derivatives.

Let us now proceed to verify the six assumptions in [20] with the overflow norm $||\cdot||$ and Lyapunov function $V(\cdot)$ defined as in (22) and (8), respectively. We list the assumptions here for reference.

A.2 Restatement of assumptions from [20]

Assumption 1 *The Lyapunov function $V(\vec{x})$, defined for $\vec{x} \geq 0$, satisfies the following:*

(a) $V(\vec{x})$ is a continuous function of $\vec{x}$.

(b) $V(\vec{x}) \geq 0$ for all $\vec{x}$ and $V(\vec{x}) = 0$ if and only if $\vec{x} = 0$.

(c) $V(\vec{x}) \to \infty$ if $||\vec{x}|| \to \infty$.

(d) $\min_{||\vec{x}|| \geq 1} V(\vec{x}) \geq 1$. Further there exists a number $\tilde{C}$ such that $\max_{||\vec{x}|| \leq 1} V(\vec{x}) \leq \tilde{C}$.

(e) For any $\mathcal{B} > 0$, there exists a constant $\mathcal{L}$ that may depend on $\mathcal{B}$, such that for any $||\vec{x}_1|| \leq \mathcal{B}$ and $||\vec{x}_2|| \leq \mathcal{B}$,

$$|V(\vec{x}_1) - V(\vec{x}_2)| \leq \mathcal{L}||\vec{x}_1 - \vec{x}_2||.$$

(f) The following holds (for a fixed arrival rate $\vec{\lambda}$ and a fixed channel state distribution $\vec{p}$ assumed in the system model): For all fluid limits $\vec{x}$ (i.e. fluid sample path with $\frac{d}{dt}\vec{s}(t) = \vec{p}$ and $\frac{d}{dt}\vec{a}(t) = \vec{\lambda}$),

$$\frac{d}{dt}V(\vec{x}(t)) \triangleq \left(\frac{\partial V}{\partial \vec{x}}\right)^T \frac{d\vec{x}}{dt} \leq -\eta V^{\gamma}(\vec{x}(t)), \tag{23}$$

for almost all t, where $0 < \gamma < 1$ and η is a positive constant.

Parts (a)-(c) and (f) of the assumption are typical when using Lyapunov functions to establish stability. Part (f) in particular states that the Lyapunov function must have negative drift when the system channel process and arrival process do not deviate from their normal behaviour. Parts (d) and (e), although not standard, can be made to hold for many Lyapunov functions that have been used for wireless systems, by an appropriate scaling.

Assumption 2 *(a) There exists $\epsilon > 0$ such that for all fluid sample paths FSP$(\vec{s}, \vec{a}, \vec{x})_T$ and for all time t with $||\frac{d}{dt}\vec{s}(t) - \vec{p}|| \leq \epsilon$ and $||\frac{d}{dt}\vec{a}(t) - \vec{\lambda}|| \leq \epsilon$, the following holds:*

$$\frac{d}{dt}V(\vec{x}(t)) \leq -\frac{\eta}{2}V^{\gamma}(\vec{x}(t)),$$

where $0 < \gamma < 1$ and $\eta > 0$ are the same constants as in (23).

(b) *For any $\delta > 0$, there exists $M_1 \geq 0$ such that for all fluid sample paths FSP$(\vec{s}, \vec{a}, \vec{x})_T$ and for all time t with $||\frac{d}{dt}\vec{s}(t) - \vec{p}|| \geq \delta$ or $||\frac{d}{dt}\vec{a}(t) - \vec{\lambda}|| \geq \delta$, the following holds,*

$$\frac{d}{dt}V(\vec{x}(t)) \leq M_1 V^{\gamma}(\vec{x}(t)).$$

Part (a) of this assumption states that if the channel state process and arrival process deviate from their normal behaviour slightly, the Lyapunov function still experiences negative drift. Hence the system is still stable. Part (b) states that even if the channel state process or the arrival process deviates significantly from their normal behaviour, the rate of growth of the Lyapunov function is still controlled. Hence, the system will not blow-up immediately.

Assumption 3 *The Lyapunov function $V(\cdot)$ is linear in scale, i.e., $V(c\vec{x}) = cV(\vec{x})$ for all $c \geq 0$.*

For any arrival rate vector $\vec{f}$ and channel state vector $\vec{\phi}$, define δ_i^k (the difference between the rate of arrival and rate of departure of data at queue X_i^k) as follows. Assume that $\gamma_{\vec{e}}$ is the fraction of time that scheduling vector $\vec{e}$ is used and $\theta_{l,j}^k$ the fraction of service given to flow k over link l in channel state j immediately after time t.

$$\begin{aligned}
\delta_{n^k(1)}^k &= f^k(t) - \sum_{j=1}^{S} \phi_j(t) \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e}} \theta_{l^k(1),j}^k e_l \\
\delta_{n^k(i)}^k &= \sum_{j=1}^{S} \phi_j(t) \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e}} \theta_{l^k(i-1),j}^k e_l - \sum_{j=1}^{S} \phi_j(t) \sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e}} \theta_{l^k(i),j}^k e_l \text{ for } i = 2, \ldots, D^k - 1 \\
\delta_{n^k(D^k)}^k &= 0 \\
\delta_i^k &= 0 \text{ for all other nodes } i
\end{aligned}$$

Define

$$\tilde{V}(\tau, \vec{e} | \vec{x}, \vec{\phi}, \vec{f}) \triangleq V([\vec{x} + \vec{\delta}\tau]^+).$$

Assumption 4 *For any fluid sample path FSP$(\vec{s}, \vec{a}, \vec{x})$, the following holds for all t:*

$$\frac{d}{dt}V(\vec{x}(t)) = \min_{\gamma_{\vec{e}},\, \theta_{l,j}^k} \frac{\partial}{\partial \tau} \tilde{V}(\tau, \vec{e} | \vec{x}(t), \vec{\phi}(t), \vec{f}(t))\Big|_{\tau=0}.$$

where $\vec{\phi}(t) = \frac{d}{dt}\vec{s}(t)$, $\vec{f}(t) = \frac{d}{dt}\vec{a}(t)$, $\sum_{\{\vec{e} \in \hat{\mathcal{E}}_j\}} \gamma_{\vec{e}} = 1$ for $j = 1, \ldots, S$ and $\sum_{k=1}^{K} \theta_{l,j}^k = 1$ for $l = 1, \ldots, L$ and $j = 1, \ldots, S$.

This assumption states that at any point of the fluid-sample-path $(\vec{s}(t), \vec{a}(t), \vec{x}(t))$, the scheduling algorithm minimizes the drift of the Lyapunov function over all scheduling decisions.

Assumption 5 *$V(\vec{x})$ is increasing in each component x_i.*

Assumption 6 *$V(\vec{x}_1 + \vec{x}_2) \leq V(\vec{x}_1) + V(\vec{x}_2)$ for any two vectors $\vec{x}_1 \geq 0$ and $\vec{x}_2 \geq 0$,*

Assumptions 3 and 6 combined imply that Lyapunov function $V(\cdot)$ behaves almost like a norm except that it may not be defined when components of $\vec{x}$ are negative.

A.3 Verification of assumptions

We first verify those assumptions that do not involve the derivative of the Lyapunov function. These are assumptions 1)a)-1)e), 3), 5) and 6).

A.3.1 Verification of assumptions that do not involve the derivative of $V(\vec{X})$:

It is easy to see that assumptions 1)a)-1)d), 3) and 5) hold. Assumption 1)e) follows from the fact that $||\cdot||$ is a norm and the fact that $||\vec{x}|| = V(\vec{x})$ for vector $\vec{x} \geq \vec{0}$. Assumption 6) follows from the fact that $||\cdot||$ satisfies the triangle inequality and that $||\vec{x}|| = V(\vec{x})$ for any vector $\vec{x} \geq \vec{0}$.

A.3.2 Verification of assumptions that involve the derivative of $V(\vec{X})$:

Assumptions involving the derivative of the Lyapunov function are Assumptions 1)f), 2) and 4) which we will verify after deriving the derivative of $V(t)$.

Derivation of the derivative of $V(t)$:

First, we will show that the drift of the Lyapunov function for *fluid-sample-path* FSP$(\vec{s}, \vec{a}, \vec{x})$ is

$$
\begin{aligned}
&\frac{d}{dt}V(\vec{x}(t)) = \\
&\left[\sum_{k=1}^{K}\left(\frac{V^k(\vec{x}(t))}{\lambda^k}\right)^{\beta+1}\right]^{\frac{-\beta}{\beta+1}} \\
&\times\left[\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}\frac{d}{dt}a^k(t)\right. \\
&-\sum_{j=1}^{S}\frac{d}{dt}s_j(t)\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}\left\{\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^{\alpha}\right. \\
&\qquad\left.\left.-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^{\alpha}\right\}e_l\right].
\end{aligned}
\tag{24}
$$

To see this, note that the derivative of $V(t)$ is given by the following limit (recall that we will consider right derivatives for simplicity (Section A.1) and that the queue backlog at the destination node of a flow k, $x^k_{n^k(D^k)}(t)$ is 0 for all time):

$$
\frac{d}{dt}V(t) = \lim_{\delta\to 0^+}\left[\sum_{k=1}^{K}\left(\frac{V^k(\vec{x}(t))}{\lambda^k}\right)^{\beta+1}\right]^{\frac{-\beta}{\beta+1}}\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}\frac{[x^k_{n^k(i)}(t+\delta)-x^k_{n^k(i)}(t)]}{\delta}.
$$

where for fixed δ,

$$
\frac{[x^k_{n^k(i)}(t+\delta)-x^k_{n^k(i)}(t)]}{\delta} = \lim_{B\to\infty}\frac{[x^{k,B}_{n^k(i)}(t+\delta)-x^{k,B}_{n^k(i)}(t)]}{\delta}.
$$

Hence, we consider the summation (for $\delta > 0$)

$$
\left[\sum_{k=1}^{K}\left(\frac{V^k(\vec{x}(t))}{\lambda^k}\right)^{\beta+1}\right]^{\frac{-\beta}{\beta+1}}\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}[x^{k,B}_{n^k(i)}(t+\delta)-x^{k,B}_{n^k(i)}(t)]
\tag{25}
$$

From the queueing equation (2), we obtain the following expression for the scaled process $\vec{x}^B(t)$

$$
\begin{aligned}
x^{k,B}_{n^k(1)}(t+\delta)-x^{k,B}_{n^k(1)}(t) &= \frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}A^k(\tau)-\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}}E^k_{l^k(1)}(j,\vec{X}(\tau))+O\left(\frac{1}{B}\right) \\
x^{k,B}_{n^k(i)}(t+\delta)-x^{k,B}_{n^k(i)}(t) &= \frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}}E^k_{l^k(i-1)}(j,\vec{X}(\tau)) \\
&\quad-\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}}E^k_{l^k(i)}(j,\vec{X}(\tau))+O\left(\frac{1}{B}\right) \\
&\quad\text{for } i=2,\ldots,D^k-1.
\end{aligned}
$$

The $O(\frac{1}{B})$ term arises due to the fact that $\vec{x}(t)$ is linearly interpolated.

Using this in (25), we obtain the following expression

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}[x^{k,B}_{n^k(i)}(t+\delta)-x^{k,B}_{n^k(i)}(t)] \\
= \;& \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}A^k(\tau) \qquad (26)\\
&-\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}}\sum_{k=1}^{K}\sum_{i=1}^{D^k-1}\left[\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i+1)}(t))^{\alpha}\right]\\
&\times E^k_{l^k(i)}(j,\vec{X}(\tau))+O(\frac{1}{B})
\end{aligned}$$

Observe that by rearranging the summation, we have

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k-1}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{n^k(i)}(t))^{\alpha}-(x^k_{n^k(i+1)}(t))^{\alpha}]E^k_{l^k(i)}(j,\vec{X}(\tau)) \qquad (27)\\
= \;& \sum_{l=1}^{L}\sum_{k\in\mathcal{K}_l}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{b(l)}(t))^{\alpha}-(x^k_{e(l)}(t))^{\alpha}]E^k_l(j,\vec{X}(\tau)).
\end{aligned}$$

Now we will use the fact that the convergence of the scaled processes $(\vec{s}^B,\vec{a}^B,\vec{x}^B)$ to the fluid sample path $(\vec{s},\vec{a},\vec{x})$ is uniform over compact intervals and that the queue process and the channel processes are Lipschitz continuous. We will need the following result.

Lemma 9 *Let $f(\cdot)$ be a continous function. Fix $\delta>0$. Then, $f(\vec{x}^B(s))\to f(\vec{x}(s))$ uniformly over the set $s\in[t,t+\delta]$.*

Proof: Since $\vec{x}^B(s)$ converges to $\vec{x}(s)$ u.o.c (uniformly over compact intervals), we have $\sup_{s\in[t,t+\delta]}|\vec{x}^B(s)-\vec{x}(s)|\to 0$ as $B\to\infty$. In other words, for any $\epsilon_1>0$, there exists B_1 such that for any $B>B_1$, $\sup_{s\in[t,t+\delta]}|\vec{x}^B(s)-\vec{x}(s)|<\epsilon_1$. Since $\vec{x}(s)$ is Lipschitz, it is bounded for $s\in[t,t+\delta]$. Hence, there is a set C that is closed, bounded and contains the set $\cup_{B>B_1}\{\vec{x}^B(s):s\in[t,t+\delta]\}\cup\{\vec{x}(s):s\in[t,t+\delta]\}$. The function $f(\cdot)$ is uniformly continous over C since C is compact. Therefore, for any $\epsilon>0$, there exists $\epsilon_1>0$ such that $|f(\vec{x}^B(s))-f(\vec{x}(s))|<\epsilon$ whenever $|\vec{x}^B(s)-\vec{x}(s)|<\epsilon_1$ and $\vec{x}^B(s),\vec{x}(s)\in C$. Since $|\vec{x}^B(s)-\vec{x}(s)|<\epsilon_1$ and $\vec{x}^B(s),\vec{x}(s)\in C$ for $B>B_1$, we can conclude that for any $\epsilon>0$, there exists B_1 such that for any $B>B_1$, $\sup_{s\in[t,t+\delta]}|f(\vec{x}^B(s))-f(\vec{x}(s))|<\epsilon$. *Q.E.D.*

Lemma 10 *Under the $\alpha\beta$-algorithm, there exists $B^1>0$ and $\delta_1>0$ such that for $B>B^1$ and time-slots $\tau\in[\lfloor B(T+t)\rfloor,\lfloor B(T+t+\delta_1)\rfloor-1]$*

$$\begin{aligned}
&\sum_{l=1}^{L}\sum_{k\in\mathcal{K}_l}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{b(l)}(t))^{\alpha}-(x^k_{e(l)}(t))^{\alpha}]E^k_l(j,\vec{X}(\tau))\\
= \;& \sum_{l=1}^{L}\tilde{W}_l(t)E_l(j,\vec{X}(\tau)) \qquad (28)
\end{aligned}$$

where

$$\tilde{W}_l(t)\triangleq\max_{k\in\mathcal{K}_l}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{b(l)}(t))^{\alpha}-(x^k_{e(l)}(t))^{\alpha}]$$

Remark: Since the $\alpha\beta$-algorithm chooses one of the users with the largest value of $\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{b(l)}(t))^{\alpha}-(x^k_{e(l)}(t))^{\alpha}]$ for service, it is natural to expect that when $\tau=B(T+t)$, the above equation must hold. The key

result in this lemma is to show that the equation holds even when τ deviates slightly from $B(T+t)$ by belonging to an interval $[\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_1)\rfloor - 1]$.
Proof: Consider link l. If

$$\frac{(V^{\hat{k}}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{\hat{k}})^{\beta+1}}[(x^{\hat{k}}_{b(l)}(t))^{\alpha} - (x^{\hat{k}}_{e(l)}(t))^{\alpha}] \quad < \quad \max_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k}_{b(l)}(t))^{\alpha} - (x^{k}_{e(l)}(t))^{\alpha}],$$

then by continuity of $\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k}_{b(l)}(t))^{\alpha} - (x^{k}_{e(l)}(t))^{\alpha}]$, there exists $\delta_1, \epsilon_1 > 0$ such that for $s \in [t, t+\delta_1]$, we have

$$\frac{(V^{\hat{k}}(\vec{x}(s)))^{\beta-\alpha}}{(\lambda^{\hat{k}})^{\beta+1}}[(x^{\hat{k}}_{b(l)}(s))^{\alpha} - (x^{\hat{k}}_{e(l)}(s))^{\alpha}] \quad < \quad \max_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}(s)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k}_{b(l)}(s))^{\alpha} - (x^{k}_{e(l)}(s))^{\alpha}] - \epsilon_1.$$

Now, by Lemma 9, we have that there exists $B^1 > 0$ such that for any $B > B^1$ and $s \in [t, t+\delta_1]$

$$\frac{(V^{\hat{k}}(\vec{x}^B(s)))^{\beta-\alpha}}{(\lambda^{\hat{k}})^{\beta+1}}[(x^{\hat{k},B}_{b(l)}(s))^{\alpha} - (x^{\hat{k},B}_{e(l)}(s))^{\alpha}] \quad < \quad \max_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}^B(s)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k,B}_{b(l)}(s))^{\alpha} - (x^{k,B}_{e(l)}(s))^{\alpha}].$$

Therefore, during timeslots $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_1)\rfloor - 1]$,

$$\frac{(V^{\hat{k}}(\vec{X}(\tau)))^{\beta-\alpha}}{(\lambda^{\hat{k}})^{\beta+1}}[(X^{\hat{k}}_{b(l)}(\tau))^{\alpha} - (X^{\hat{k}}_{e(l)}(\tau))^{\alpha}] \quad < \quad \max_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{X}(\tau)))^{\beta-\alpha}}{(\lambda_{k})^{\beta+1}}[(X^{k}_{b(l)}(\tau))^{\alpha} - (X^{k}_{e(l)}(\tau))^{\alpha}]$$

and hence flow $\hat{k}$ will not be activated over link l if link l is scheduled for transmission. Further, one of the flows in $\text{argmax}_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}^B(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k,B}_{b(l)}(t))^{\alpha} - (x^{k,B}_{e(l)}(t))^{\alpha}]$, will be given the complete service $E_l(j, \vec{X}(\tau))$. Hence we have

$$\begin{aligned}
&\sum_{l=1}^{L}\sum_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k}_{b(l)}(t))^{\alpha} - (x^{k}_{e(l)}(t))^{\alpha}]E^k_l(j, \vec{X}(\tau)) \\
= &\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}[(x^{k}_{b(l)}(t))^{\alpha} - (x^{k}_{e(l)}(t))^{\alpha}]E_l(j, \vec{X}(\tau)) \\
= &\sum_{l=1}^{L}\tilde{W}_l(t)E_l(j, \vec{X}(\tau))
\end{aligned}$$

where $(E_1(j, \vec{X}(\tau)), \ldots, E_L(j, \vec{X}(\tau))) \in \mathcal{E}_j$.

Q.E.D.

Therefore, using (27) and lemma 10, (26) becomes (for $B > B^1$ and $\delta < \delta_1$)

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}[x^{k,B}_{n^k(i)}(t+\delta) - x^{k,B}_{n^k(i)}(t)] \\
= &\sum_{k=1}^{K}\frac{(V^{k}(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^{k})^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}A^k(\tau) \\
&-\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}}\sum_{l=1}^{L}\tilde{W}_l(t)E_l(j, \vec{X}(\tau)) + O(\frac{1}{B})
\end{aligned} \tag{29}$$

Lemma 11 *There exists B^4 and δ_4 such that for $B > B^4$ and $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_4)\rfloor - 1]$, we have*

$$\sum_{l=1}^{L}\tilde{W}_l(t)E_l(j, \vec{X}(\tau)) = \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\tilde{W}_l(t)e_l.$$

Remark: Since the $\alpha\beta$-algorithm schedules links (from the set of non-interfering schedules) such that $\sum_{l=1}^{L} \tilde{W}_l(t)e_l$ is maximized, it is natural to expect that when $\tau = B(T+t)$, the above equation would hold. The result of the lemma is stronger, stating that the equation holds even when τ deviates slightly, belonging to $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_4)\rfloor - 1]$.

Proof: Define the set $\acute{\mathcal{L}} = \{l : \tilde{W}_l(t) \neq 0\}$. Choose $\epsilon_2 = \min\{\frac{|\tilde{W}_l(t)|}{2} : l \in \acute{\mathcal{L}}\}$. Due to continuity of $\tilde{W}_l(t)$, there exists $\delta_2 > 0$ such that for $s \in [t, t+\delta_2]$, the following is true $\tilde{W}_l(s) > \epsilon_2$ if $\tilde{W}_l(t) > 0$ or $\tilde{W}_l(s) < -\epsilon_2$ if $\tilde{W}_l(t) < 0$. By application of Lemma 9,

$$\max_{k\in\mathcal{K}_l} \frac{(V^k(\vec{x}^B(s)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x_{b(l)}^{k,B}(s))^\alpha - (x_{e(l)}^{k,B}(s))^\alpha] \to \tilde{W}_l(s)$$

uniformly over the set $s \in [t, t+\delta_2]$. This implies that there exists B^2 such that for $B > B^2$, and time-slot $\tau \in [B(T+t), B(T+t+\delta_2)]$,

$$\frac{1}{B^\beta} \max_{k\in\mathcal{K}_l} \frac{(V^k(\vec{X}(\tau)))^{\beta-\alpha}}{(\lambda_k)^{\beta+1}}[(X_{b(l)}^{k}(\tau))^\alpha - (X_{e(l)}^{k}(\tau))^\alpha] > \epsilon_2 \text{ or } < -\epsilon_2$$

This implies that for $B > B^2$, $E_l(j, \vec{X}(\tau))$ is either 0 or F_j^l for $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_2)\rfloor - 1]$ (i.e. there is enough backlog in queue $X_{b(l)}^k$ that the entire capacity of link l is utilized in these time-slots).

Our strategy will be to show that if $\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l$ then there is a time slot τ where, in the pre-limit system, we have $\sum_{l=1}^{L} W_l(\tau)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} W_l(\tau)e_l$. This is contradictory to the behaviour of the $\alpha\beta$-algorithm during this time-slot τ and hence $\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l$ cannot hold.

Consider $B > B^2$ and $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_2)\rfloor - 1]$. If $\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l$, then there exists $\epsilon_3 > 0$ (since $E_l(j, \vec{X}(\tau))$ is either 0 or F_j^l for $l \in \acute{\mathcal{L}}$) such that

$$\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l - \epsilon_3.$$

Choose $\epsilon_4 < \frac{\epsilon_3}{4\sum_{l=1}^{L}\sum_{j=1}^{S} F_j^l}$. Due to continuity of $\tilde{W}_l(t)$, there exists $\delta_3 > 0$ such that for $s \in [t, t+\delta_3]$ and $l \in \mathcal{L}$, $|\tilde{W}_l(s) - \tilde{W}_l(t)| < \epsilon_4$. Therefore, for $B > B^2$, $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_2)\rfloor - 1]$ and $s \in [t, t+\delta_3]$ we have

$$\sum_{l=1}^{L} \tilde{W}_l(s)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(s)e_l - \frac{\epsilon_3}{2} \tag{30}$$

Define

$$\tilde{W}_l^B(t) \triangleq \max_{k\in\mathcal{K}_l} \frac{(V^k(\vec{x}^B(t)))^{\beta-\alpha}}{(\lambda_k)^{\beta+1}}[(x_{b(l)}^{k,B}(t))^\alpha - (x_{e(l)}^{k,B}(t))^\alpha].$$

Since $\tilde{W}_l(t) = \max_{k\in\mathcal{K}_l} \frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda_k)^{\beta+1}}[(x_{b(l)}^{k}(t))^\alpha - (x_{e(l)}^{k}(t))^\alpha]$ is a continuous function of $\vec{x}(t)$, by Lemma 9 we have that for any $\epsilon_5 > 0$, there exists B^3 such that for any $B > B^3$, for any $l \in \mathcal{L}$, $\sup_{s\in[t,t+\delta_3]} |\tilde{W}_l^B(s) - \tilde{W}_l(s)| < \epsilon_5$. Therefore, for $B > B^3$ we have $\sum_{l=1}^{L} \tilde{W}_l^B(s)E_l(j, \vec{X}(\tau)) \leq \sum_{l=1}^{L} \tilde{W}_l(s)E_l(j, \vec{X}(\tau)) + \sum_{l=1}^{L} \epsilon_5 E_l(j, \vec{X}(\tau))$.

Hence, by (30), for any $B > \max\{B^2, B^3\}$, $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_2)\rfloor - 1]$ and $s \in [t, t+\delta_3]$, we have

$$\sum_{l=1}^{L} \tilde{W}_l^B(s)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(s)e_l - \frac{\epsilon_3}{2} + \sum_{l=1}^{L} \epsilon_5 E_l(j, \vec{X}(\tau)).$$

Further, since for $B > B^3$ we have $\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(s)e_l < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l^B(s)e_l + \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \epsilon_5 e_l$, it follows that

$$\sum_{l=1}^{L} \tilde{W}_l^B(s)E_l(j, \vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l^B(s)e_l + \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \epsilon_5 e_l - \frac{\epsilon_3}{2} + \sum_{l=1}^{L} \epsilon_5 E_l(j, \vec{X}(\tau)).$$

Choose $\epsilon_5 < \frac{\epsilon_3}{4\sum_{l=1}^{L}\sum_{j=1}^{S} F_j^l}$. Then, we have

$$\sum_{l=1}^{L} \tilde{W}_l^B(s)E_l(j,\vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l^B(s)e_l.$$

Hence, we have shown that for $B > \max\{B^2, B^3\}$, $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_2)\rfloor - 1]$ and $s \in [t, t+\delta_3]$, if $\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j,\vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l$, then $\sum_{l=1}^{L} \tilde{W}_l^B(s)E_l(j,\vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l^B(s)e_l$. This implies that for $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\min\{\delta_2,\delta_3\})\rfloor - 1]$,

$$\sum_{l=1}^{L} \tilde{W}_l^B(\frac{\tau}{B} - T)E_l(j,\vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l^B(\frac{\tau}{B} - T)e_l.$$

That is,

$$\sum_{l=1}^{L} W_l(\tau)E_l(j,\vec{X}(\tau)) < \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} W_l(\tau)e_l.$$

However, this is contradictory to the behaviour of the scheduling algorithm during timeslot τ.

Hence, it must be that for $B > \max\{B^2, B^3\}$ and $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\min\{\delta_2,\delta_3\})\rfloor - 1]$

$$\sum_{l=1}^{L} \tilde{W}_l(t)E_l(j,\vec{X}(\tau)) = \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l.$$

Q.E.D.

Let $B^5 = \max\{B^1, B^4\}$ and $\delta_5 = \min\{\delta_1, \delta_4\}$. Combining (27), lemma 10 and lemma 11, we have for $B > B^5$ and $\tau \in [\lfloor B(T+t)\rfloor, \lfloor B(T+t+\delta_5)\rfloor - 1]$,

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k-1} \frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}[(x^k_{n^k(i)}(t))^\alpha - (x^k_{n^k(i+1)}(t))^\alpha]E^k_{l^k(i)}(j,\vec{X}(\tau)) \\
= \quad & \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l.
\end{aligned}$$

Therefore, for $\delta < \delta_5$ and $B > B^5$, by (29) we have

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k} \frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^\alpha[x^{k,B}_{n^k(i)}(t+\delta) - x^{k,B}_{n^k(i)}(t)] \qquad (31)\\
= \quad & \sum_{k=1}^{K} \frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^\alpha \frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1} A^k(\tau) \\
& -\frac{1}{B}\sum_{\tau=\lfloor B(T+t)\rfloor}^{\lfloor B(T+t+\delta)\rfloor-1}\sum_{j=1}^{S}\mathbf{1}_{\{C(\tau)=j\}} \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l + O(\frac{1}{B}) \\
= \quad & \sum_{k=1}^{K} \frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^\alpha[a^{k,B}(t+\delta) - a^{k,B}(t)] \\
& -\sum_{j=1}^{S}[s_j^B(t+\delta) - s_j^B(t)] \max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}} \sum_{l=1}^{L} \tilde{W}_l(t)e_l + O(\frac{1}{B}).
\end{aligned}$$

Substituting (31) into (25) and letting $B \to \infty$, we get

$$
\begin{aligned}
&(\sum_{k=1}^{K}(\frac{V^k(\vec{x}(t))}{\lambda^k})^{\beta+1})^{\frac{-\beta}{\beta+1}}\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}[x^k_{n^k(i)}(t+\delta)-x^k_{n^k(i)}(t)]\\
=\;&(\sum_{k=1}^{K}(\frac{V^k(\vec{x}(t))}{\lambda^k})^{\beta+1})^{\frac{-\beta}{\beta+1}}\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}[a^k(t+\delta)-a^k(t)]\\
&-\sum_{j=1}^{S}[s_j(t+\delta)-s_j(t)]\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^{\alpha}-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^{\alpha})e_l.
\end{aligned}
$$

Finally, as $\delta \to 0^+$, we have

$$
\begin{aligned}
&\frac{d}{dt}V(\vec{x}(t)) = (\sum_{k=1}^{K}(\frac{V^k(\vec{x}(t))}{\lambda^k})^{\beta+1})^{\frac{-\beta}{\beta+1}}\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}\left[\lim_{\delta\to 0^+}\frac{x^k_{n^k(i)}(t+\delta)-x^k_{n^k(i)}(t)}{\delta}\right]\\
=\;&(\sum_{k=1}^{K}(\frac{V^k(\vec{x}(t))}{\lambda^k})^{\beta+1})^{\frac{-\beta}{\beta+1}}\left[\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}\frac{d}{dt}a^k(t)\right.\\
&\left.-\sum_{j=1}^{S}\frac{d}{dt}s_j(t)\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^{\alpha}-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^{\alpha})e_l\right].
\end{aligned}
$$

Verification of assumption 4:

We have

$$
\begin{aligned}
&\frac{\partial}{\partial\tau}\tilde{V}(\tau,\vec{e}|\vec{x}(t),\vec{\phi}(t),\vec{f}(t))\\
=\;&\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{\partial}{\partial x^k_i}\left(V(\vec{x}(t))\right)[(\delta^k_i)^+\mathbf{1}_{\{x^k_i(t)=0\}}+\delta^k_i\mathbf{1}_{\{x^k_i(t)>0\}}]\\
=\;&\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{\partial}{\partial x^k_{n^k(i)}}\left(V(\vec{x}(t))\right)[(\delta^k_{n^k(i)})^+\mathbf{1}_{\{x^k_{n^k(i)}(t)=0\}}+\delta^k_{n^k(i)}\mathbf{1}_{\{x^k_{n^k(i)}(t)>0\}}].
\end{aligned}
$$

From the definition of $V(\cdot)$ given in (8), we have the following expression for the partial derivative

$$
\begin{aligned}
&\frac{\partial}{\partial x^k_{n^k(i)}}\left(V(\vec{x}(t))\right)\\
=\;&(V(\vec{x}(t)))^{-\beta}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{\partial}{\partial\tau}\tilde{V}(\tau,\vec{e}|\vec{x}(t),\vec{\phi}(t),\vec{f}(t))\\
=\;&(V(\vec{x}(t)))^{-\beta}\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^{\alpha}\delta^k_{n^k(i)}.
\end{aligned}
\tag{32}
$$

Since by definition

$$
\begin{aligned}
\delta^k_{n^k(1)} &= f^k(t)-\sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_l\\
\delta^k_{n^k(i)} &= \sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_l-\sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_l \text{ for } i=2,\ldots,D^k-1,
\end{aligned}
$$

the summation $\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^\alpha\delta^k_{n^k(i)}$ simplifies to

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{D^k}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^\alpha\delta^k_{n^k(i)}\\
= &\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^\alpha f^k(t)\\
&-\sum_{k=1}^{K}\sum_{i=1}^{D^k-1}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i)}(t))^\alpha-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(i+1)}(t))^\alpha)\sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_l\\
= &\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^\alpha f^k(t)\\
&-\sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\sum_{l=1}^{L}\sum_{k=1}^{\mathcal{K}_l}\theta^k_{l,j}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^\alpha-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^\alpha)e_l.
\end{aligned}\tag{33}$$

Observe that since $\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}=1$, $\sum_{\{k\in\mathcal{K}_l\}}\theta^k_{l,j}=1$ the following inequality holds.

$$\begin{aligned}
&\sum_{j=1}^{S}\phi_j(t)\sum_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\gamma_{\vec{e}}\sum_{l=1}^{L}\sum_{k=1}^{\mathcal{K}_l}\theta^k_{l,j}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^\alpha-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^\alpha)e_l \qquad (34)\\
\le\ &\sum_{j=1}^{S}\phi_j(t)\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^\alpha-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^\alpha)e_l.
\end{aligned}$$

Using (33) and (34) in (32), we obtain

$$\begin{aligned}
&\frac{\partial}{\partial\tau}\tilde{V}(\tau,\vec{e}|\vec{x}(t),\vec{\phi}(t),\vec{f}(t))|_{\tau=0}\\
\ge\ &\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^\alpha f^k(t)\\
&-\sum_{j=1}^{S}\phi_j(t)\max_{\{\vec{e}\in\hat{\mathcal{E}}_j\}}\sum_{l=1}^{L}\max_{k\in\mathcal{K}_l}(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^\alpha-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^\alpha)e_l.
\end{aligned}$$

The right-hand-side of the above inequality is the drift of the $\alpha\beta$-algorithm (24). Hence the $\alpha\beta$-algorithm minimizes the quantity $\frac{\partial}{\partial\tau}\tilde{V}(\tau,\vec{e}|\vec{x}(t),\vec{\phi}(t),\vec{f}(t))|_{\tau=0}$ over all choices $\gamma_{\vec{e}}$, $\theta^k_{l,j}$.

Verification of assumptions 1)f) and 2):

We use the following consequence of the assumption that the average arrival rate $\vec{\lambda}$ belongs to the interior of the capacity region. We assume that there exists $\tilde{\epsilon}>0$ and real numbers $1>\gamma_{\vec{e}}>0$ (fraction of weight given to service vector $\vec{e}$), $1>\theta^k_{l,j}>0$ (fraction of service given to flow k over link l in state j) such that $\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\le 1$, $\sum_{k=1}^{K}\theta^k_{l,j}\le 1$ and

$$\begin{aligned}
\lambda^k-\sum_{j=1}^{S}p_j\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_l &< -\tilde{\epsilon}\\
\sum_{j=1}^{S}p_j\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_{l^k(i-1)}-\sum_{j=1}^{S}p_j\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)} &< -\tilde{\epsilon}\text{ for } i=2,\ldots,D^k-1
\end{aligned}$$

.

Using the ideas used in verification of Assumption 4), we can show that

$$
\begin{aligned}
\frac{d}{dt}V(t) &\leq \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{n^k(1)}(t))^{\alpha}\frac{d}{dt}a^k(t) \\
&\quad -\sum_{j=1}^{S}\frac{d}{dt}s_j(t)\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\sum_{l=1}^{L}\sum_{k\in\mathcal{K}_l}\theta^k_{l,j}\Big(\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{b(l)}(t))^{\alpha}-\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}(x^k_{e(l)}(t))^{\alpha}\Big)e_l \\
&= \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}\Bigg\{(x^k_{n^k(1)}(t))^{\alpha}\frac{d}{dt}a^k(t)-\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}(x^k_{n^k(1)})^{\alpha}\left[\frac{d}{dt}s_j(t)\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_{l^k(1)}\right] \\
&\quad +\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}\sum_{i=2}^{D^k-1}(x^k_{n^k(i)})^{\alpha}\left[\frac{d}{dt}s_j(t)\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_{l^k(i-1)}-\frac{d}{dt}s_j(t)\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)}\right]\Bigg\}.
\end{aligned}
$$

If $||\frac{d}{dt}\vec{s}(t)-\vec{p}|| \leq \epsilon$ and $||\frac{d}{dt}\vec{a}(t)-\vec{\lambda}|| \leq \epsilon$, we further have

$$
\begin{aligned}
\frac{d}{dt}V(t) &\leq \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}\Bigg\{(x^k_{n^k(1)}(t))^{\alpha}\left[(\lambda^k+\epsilon)-\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_{l^k(1)}+\epsilon S\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_{l^k(1)}\right] \\
&\quad +\sum_{i=2}^{D^k-1}(x^k_{n^k(i)})^{\alpha}\Bigg[\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_{l^k(i-1)}+\epsilon S\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_{l^k(i-1)} \\
&\quad -\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)}+\epsilon S\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)}\Bigg]\Bigg\}. \qquad (35)
\end{aligned}
$$

Choose a real number $\Gamma > 1$ such that for all $k = 1,\ldots,K$, $i = 1,\ldots,D^k-1$, $S\sum_{\{\vec{e}\in\mathcal{E}_j\}}\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)} < \Gamma$. Then (35) simplifies to

$$
\begin{aligned}
\frac{d}{dt}V(t) &\leq \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}\Bigg\{(x^k_{n^k(1)}(t))^{\alpha}\left[\lambda^k-\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(1),j}e_{l^k(1)}\right] \\
&\quad +\sum_{i=2}^{D^k-1}(x^k_{n^k(i)})^{\alpha}\Bigg[\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(i-1),j}e_{l^k(i-1)} \\
&\quad -\sum_{j=1}^{S}\sum_{\{\vec{e}\in\mathcal{E}_j\}}p_j\gamma_{\vec{e}}\theta^k_{l^k(i),j}e_{l^k(i)}\Bigg]+\sum_{i=1}^{D^k-1}(x^k_{n^k(i)})^{\alpha}2\epsilon\Gamma\Bigg\} \\
&\leq \sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}\sum_{i=1}^{D^k-1}(x^k_{n^k(i)})^{\alpha}(-\tilde{\epsilon}+2\epsilon\Gamma). \qquad (36)
\end{aligned}
$$

To simplify (36), we will use the following inequality

$$
\sum_{i=1}^{D^k}(x^k_{n^k(i)})^{\alpha} \geq \max_{\{i=1,\ldots,D^k\}}(x^k_{n^k(i)})^{\alpha} \geq \frac{1}{(D^k)^{\frac{\alpha}{\alpha+1}}}\Big(\sum_{i=1}^{D^k}(x^k_{n^k(i)})^{\alpha+1}\Big)^{\frac{\alpha}{\alpha+1}} \geq \frac{1}{N^{\frac{\alpha}{\alpha+1}}}(V^k(\vec{x}(t)))^{\alpha}. \qquad (37)
$$

We will also use the following inequality

$$
\sum_{k=1}^{K}\Big(\frac{V^k(\vec{x}(t))}{\lambda_k}\Big)^{\beta} \geq \frac{1}{K^{\frac{\beta}{\beta+1}}}\Big(\sum_{k=1}^{K}\Big(\frac{V^k(\vec{x}(t))}{\lambda_k}\Big)^{\beta+1}\Big)^{\frac{\beta}{\beta+1}} = \frac{1}{K^{\frac{\beta}{\beta+1}}}(V(\vec{x}(t)))^{\beta}. \qquad (38)
$$

Combining (37) and (38), we obtain

$$\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta-\alpha}}{(\lambda^k)^{\beta+1}}\sum_{i=1}^{D^k-1}(x^k_{n^k(i)})^{\alpha} \geq \frac{1}{N^{\frac{\alpha}{\alpha+1}}}\sum_{k=1}^{K}\frac{(V^k(\vec{x}(t)))^{\beta}}{(\lambda^k)^{\beta+1}} \geq \frac{1}{N^{\frac{\alpha}{\alpha+1}}K^{\frac{\beta}{\beta+1}}\max_{\{k=1,\ldots,K\}}\lambda^k}(V(\vec{x}(t)))^{\beta}. \tag{39}$$

Now, applying (39) in (36), we obtain

$$\frac{d}{dt}V(t) \leq (V(\vec{x}(t)))^{\beta}\frac{-\tilde{\epsilon}+2\epsilon\Gamma}{N^{\frac{\alpha}{\alpha+1}}K^{\frac{\beta}{\beta+1}}\max_{\{k=1,\ldots,K\}}\lambda^k}.$$

To verify Assumption 1)f), set $\epsilon = 0$. We have

$$\frac{d}{dt}V(t) \leq (V(\vec{x}(t)))^{\beta}\frac{-\tilde{\epsilon}}{N^{\frac{\alpha}{\alpha+1}}K^{\frac{\beta}{\beta+1}}\max_{\{k=1,\ldots,K\}}\lambda^k}.$$

To verify Assumption 2), set $\epsilon = \frac{\tilde{\epsilon}}{4\Gamma}$, we have

$$\frac{d}{dt}V(t) \leq (V(\vec{x}(t)))^{\beta}\frac{-\tilde{\epsilon}/2}{N^{\frac{\alpha}{\alpha+1}}K^{\frac{\beta}{\beta+1}}\max_{\{k=1,\ldots,K\}}\lambda^k}.$$

Now that we have verified that $V(\cdot)$ satisfies all Assumptions 1) to 6), we can invoke Proposition 8 of [20] and conclude that the quantity

$$\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B])$$

exists and for any scheduling policy π,

$$\begin{aligned}&\lim_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\alpha\beta}[V(\vec{X}(0)) > B])\\ &\qquad\leq \liminf_{B\to\infty}\frac{1}{B}\log(\mathbf{P}^{\pi}[V(\vec{X}(0)) > B]).\end{aligned}$$

References

[1] L. Tassiulas and A. Ephremides, “Dynamic Server Allocation to Parallel Queues with Randomly Varying Connectivity,” *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, March 1993.

[2] ——, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Trans. Automat. Contr.*, vol. 4, pp. 1936–1948, December 1992.

[3] ——, “Dynamic server allocation to parallel queues with randomly varying connectivity,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 466–478, March 1993.

[4] X. Lin and N. Shroff, “Joint rate control and scheduling in multihop wireless networks,” in *Proc. Conf. on Decision and Control*, Paradise Island, Bahamas, December 2004.

[5] A. Eryilmaz and R. Srikant, “Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control,” in *Proc. IEEE Infocom.*, 2005.

[6] A. Stolyar, “Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm,” *Queueing Systems*, vol. 50, no. 4, pp. 401–457, August 2005.

[7] M. Neely, E. Modiano, and C. Li, “Fairness and optimal stochastic control for heterogeneous networks,” in *Proc. IEEE Infocom.*, vol. 3, Miami, FL, March 2005, pp. 1723–1734.

[8] L. Ying, S. Shakkottai, and A. Reddy, “On combining shortest-path and back-pressure routing over multihop wireless networks,” in *Proc. IEEE Infocom.*, Rio de Janeiro, Brazil, 2009.

[9] L. Bui, R. Srikant, and A. L. Stolyar, "Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing," *Proc. IEEE Infocom. Mini Conf.*, 2009.

[10] M. J. Neely, "Order Optimal Delay for Opportunistic Scheduling in Multi-User Wireless Uplinks and Downlinks," *IEEE/ACM Transactions on Networking*, 2008.

[11] L. Ying, R. Srikant, A. Eryilmaz, and G. E. Dullerud, "A Large Deviations Analysis of Scheduling in Wireless Networks," *IEEE Transactions on Information Theory*, vol. 52, no. 11, November 2006.

[12] A. L. Stolyar, "MaxWeight Scheduling in a Generalized Switch: State Space Collapse and Workload Minimization in Heavy Traffic," *Annals of Applied Probability*, vol. 14, no. 1, pp. 1–53, 2004.

[13] S. Shakkottai, R. Srikant, and A. Stolyar, "Pathwise Optimality of the Exponential Scheduling Rule for Wireless Channels," *Advances in Applied Probability*, pp. 1021–1045, December 2004.

[14] D. Shah and D. Wischik, "Optimal Scheduling Algorithms for Input-Queued Switches," in *Proceedings of IEEE INFOCOM*, Barcelona, Spain, April 2006.

[15] S. P. Meyn, "Stability and Asymptotic Optimality of Generalized MaxWeight Policies," *SIAM J. Control and Optimization*, to appear.

[16] S. Shakkottai, "Effective Capacity and QoS for Wireless Scheduling," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, April 2008.

[17] A. L. Stolyar, "Large Deviations of Queues Sharing a Randomly Time-varying Server," *Queueing Systems*, vol. 59, pp. 1–35, 2008.

[18] V. J. Venkataramanan and X. Lin, "On Wireless Scheduling Algorithms for Minimizing the Queue-Overflow Probability," *submitted to IEEE/ACM Trans. on Networking*, 2008. *Preprint available at http://min.ecn.purdue.edu/∼linx/publications.html.*

[19] A. Eryilmaz and R. Srikant, "Scheduling with quality of service constraints over Rayleigh fading channels," in *Proc. Conf. on Decision and Control*, vol. 4, December 2004, pp. 3447–3452.

[20] V. J. Venkataramanan and X. Lin, "On the Queue-Overflow Probability of Wireless Systems: A New Approach Combining Large Deviations with Lyapunov Functions," *submitted to IEEE Trans. on Information Theory*, 2009. *Preprint available at http://min.ecn.purdue.edu/∼linx/publications.html.*

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