On the Queue-Overflow Probability of Wireless Systems:
A New Approach Combining Large Deviations with Lyapunov Functions

V. J. Venkataramanan and Xiaojun Lin
Center for Wireless Systems and Applications
School of Electrical and Computer Engineering,
Purdue University, West Lafayette, IN 47907, U.S.A.
Email: {vvenkat,linx}@ecn.purdue.edu.

Abstract

In this paper we study the problem of characterizing the queue-overflow probability of wireless scheduling algorithms. In wireless networks operated under queue-length-based wireless scheduling algorithms, there often exists a tight coupling between the service-rate process, the system backlog process, the arrival process and the channel variations. Although one can use sample-path large-deviation techniques to form an estimate of the queue-overflow probability under a given offered load, the formulation leads to a multi-dimensional calculus of variations problem that is often very difficult to solve. In this paper, we present a new technique for addressing this complexity issue. Using ideas from the Lyapunov function approach in control theory, this technique maps the complex multi-dimensional calculus of variations problem to a one-dimensional calculus of variations problem, and the latter is often much easier to solve. Further, under appropriate conditions, we
show that when a scheduling algorithm minimizes the drift of the Lyapunov function, the algorithm will be optimal in the sense that it maximizes the asymptotic decay-rate of the probability that the Lyapunov function values exceed a given threshold. We believe that these results can potentially be used to study the queue-overflow probability of a large class of wireless scheduling algorithms.

1 Introduction

A wireless network may be modeled as a system of queues with time-varying service rates. The variability in service rates is due to a number of factors. First, channel fading and mobility can lead to variations in the link capacity even if the transmission power is fixed. Second, the transmission power can vary over time according to the power control policy. Third, due to radio interference, it is usually preferable to schedule only a subset of links to be active at each time, and to alternate the subset of activated links over time. All of these factors lead to a variable service rate at each link.

When we study the performance of any system that involves queues, the first question that we can ask is whether the system is stable or not. Here, stability means that all queue backlog (or equivalently, the delay experienced by the packets) remains finite. Conversely, we can ask the question that, in order to maintain stability, what is the largest offered load that the system can carry. For wireless networks, these questions have led to results on throughput-optimal scheduling and routing algorithms for managing wireless network resources. Here, we use the term scheduling in the broader sense, i.e., it can include various control mechanisms at the MAC/PHY layer, such as link scheduling, power control, and adaptive coding/modulation. In addition, for multi-hop wireless networks, the routing functionality determines the path that each packet traverses, which also plays a key role in determining the capacity of the network. A scheduling and routing algorithm is throughput-optimal if, for any offer load under which any other scheduling and routing algorithm can stabilize the system, this algorithm can stabilize the system as well. For
example, one such throughput-optimal algorithm is the so-called "maximum-weight" and "back-pressure" algorithm proposed in the seminal work by Tassiulas and Ephremides in [1]. This algorithm chooses at each time, among all possible scheduling and routing decisions, the one that maximizes the sum of the link rates weighted by the differential backlog. This algorithm has been shown to be throughput-optimal, and it has been the basis for many other throughput-optimal algorithms for both cellular and multihop wireless networks.

Once we know about stability, we are then tempted to ask further questions regarding the distribution of queue length (or delay). For example, at a given offered load, what is the probability that the queue length at any link exceeds a given threshold (or, that the delay experienced by a packet is greater than a given threshold)? Conversely, what is the largest offered load that the system can support at a given queue-overflow or delay constraint? (In other words, what is the effective capacity region of the system under such constraints?) Clearly, these question are important for applications that require more stringent delay guarantees than just stability.

Such problems for characterizing the queue-overflow probability or delay-violation probability in wireless networks can be difficult to solve. Here we draw a comparison to similar queueing problems in wireline networks. In wireline networks, even though the exact queue distribution can be difficult to obtain, there have been a large body of work, especially those using large-deviation techniques, to obtain sharp estimates of the queue-overflow probability [2–7]. Essentially, we can compute the asymptotic decay-rate with which the queue-overflow probability approaches zero as the overflow threshold approaches infinity. We can then compare the queueing performance of different systems by their corresponding asymptotic decay-rates, and we can ask questions regarding the largest offered load subject to the constraint that the decay-rate must be no smaller than a given threshold value. Most results along this line assume that the service rate of the queue is fixed (i.e., time-invariant), and the packet arrival process is known. These results enable us to define the notion of effective bandwidth of the arrival process based on its (known) statistics [2–7], which can then be used to determine the traffic carrying capability of the system.
at a given queue-overflow constraint. In contrast, in wireless networks, the service rate is time-varying. If the service rate process is again known a priori, large-deviation techniques can be used to compute the effective capacity of the service rate process [8, 9], which is a notion similar to the effective bandwidth of the arrival process. This effective capacity can again be used to determine the traffic carrying capability of the system at a given constraint on the decay-rate of the queue-overflow probability. Unfortunately, under many wireless resource-allocation algorithms of interest, even the service rate process is unknown a priori. For example, for a system operated under the throughput-optimal Tassiulas-Ephremides algorithm of [1], or any queue-length based algorithms, the service rates depend on the queue length, which in turn depend on the past history of the arrival process and the channel state. Hence, the statistics of the service rate process is unknown a priori. In this case, even the computation of the asymptotic decay-rate of the queue-overflow probability becomes a very difficult problem. For these systems, although it is still possible to use sample-path large-deviation techniques to form an estimate of the decay-rate of the queue-overflow probability [10–12], such a formulation leads to a multi-dimensional calculus-of- variations problem. Due to the complex coupling between the service rate, the queue length, the arrival process, and the channel state, this multi-dimensional calculus of variations problem is known to be very difficult [10–13].

Motivated by the Lyapunov function approach for proving stability of complex systems, in this paper we provide a technique that addresses the complexity of the multi-dimensional calculus-of-variations problem that arises in sample-path large-deviation studies. In essence, through the use of a Lyapunov function, we map the multi-dimensional calculus-of-variations problem into a one-dimensional calculus-of-variations problem, and the latter is often much easier to solve. The solution to the one-dimensional calculus-of-variations problem will then provide us with an lower-bound estimate of the decay-rate of the queue-overflow probability, and consequently, a lower-bound estimate of the effective capacity region of the system. For many practical applications, the resulting effective capacity region is useful because the queue-overflow constraint is known
to be satisfied (in the large-deviation sense).

When this idea was first presented in our earlier work [14], we have assumed that a large-deviation principle (LDP) holds for the queue backlog process. Unfortunately, for many wireless systems, such an LDP by itself is often difficult to check. In this paper, we prove that, even without assuming an LDP for the queue backlog process, our Lyapunov function approach will provide a lower bound on the decay-rate of the queue-overflow probability. Further, we provide a useful condition when this bound becomes tight. Specifically, we show that if the scheduling and routing algorithm minimizes the drift of the Lyapunov function at each point of the fluid sample path, then the bound obtained through the Lyapunov function approach is exactly equal to the asymptotic decay-rate of the probability that the Lyapunov function value exceeds a given threshold. When this happens, the algorithm is in fact optimal in maximizing the asymptotic decay-rate of the probability that the Lyapunov function value overflows. Note that many throughput-optimal algorithms are derived by minimizing the drift of a corresponding Lyapunov function. Hence, the above optimality condition is applicable to a large class of scheduling and routing algorithms. For example, according to this result, the back-pressure algorithm [1] maximizes the asymptotic decay-rate of the probability that the sum of the squares of the queues exceeds a given threshold. We also provide another example (which is a more general version than the problems studied in existing work), and demonstrate how to use this approach to derive the exact value of the asymptotic decay-rate of the queue-overflow probability, along with the corresponding effective capacity region.

We believe that this marriage between sample-path large-deviations and Lyapunov functions can develop into a powerful theory to characterize the queueing behavior of wireless systems under sophisticated scheduling and routing algorithms. We can potentially lower the difficulty level of the queueing problem to that of a stability problem. In other words, for any scheduling and routing algorithm that is provably stable, which usually means that there exists a Lyapunov function, we could then apply this theory and recover important properties of the corresponding
queueing performance.

In our prior work [15], we have also used Lyapunov functions to study the large-deviation decay rate of the queue-overflow probability in the setting of a single cell in a cellular system. Such setting is also studied in [13] using a different approach. This paper differs from [15] in several aspects. Firstly, the results of this paper are established for a much larger class of wireless networks (including multi-hop wireless networks), a much larger class of scheduling algorithms (including, e.g., the back-pressure algorithm), and more general overflow metrics. In contrast, the work in [15] focuses on a single cell model, the so-called $\alpha$-algorithms and the max-queue overflow metric. The emphasis there is to develop practical near-optimal scheduling algorithms for that setting. Secondly, this paper provides a more thorough understanding of the Lyapunov function approach than [15]. Specifically, in Proposition 3, we establish that the decay-rate of the stationary over-flow probability is bounded by the minimum-cost-to-overflow. For this result to hold, we only assume that the system has a Lyapunov function that satisfies Assumptions 1 and 2. In contrast, the technique in [15] uses additional structure of the system. Similarly, in Proposition 8, we state the condition under which the scheduling algorithm is optimal. This condition is again only related to the Lyapunov functions and hence is more general than the result in [15]. Lastly, in [15], since the focus is more on the practical aspect, we have not included the detailed proof of the Freidlin-Wentzell construction. In this paper, we provide the proof (See Propositions 3 and 4), again under very general conditions.

The rest of the paper is organized as follows. We present the network model in Section 2. In Section 3, we provide a general lower bound for the asymptotic decay-rate of the queue-overflow probability using sample-path large-deviation theory. However, this bound involves a difficult multi-dimensional calculus-of-variations problem. Then, in Section 4, we provide a Lyapunov function based approach to address the complexity issue, which provides an even simpler lower bound on the asymptotic decay-rate of the queue-overflow probability. In Section 5, we provide a condition under which this lower bound is tight. In Section 6, we provide examples to show
how such an approach can be applied to various network settings. Then we conclude.

Throughout the paper, we use $x$ to denote real numbers and $\vec{x}$ to denote vectors. For convenience, when we refer to a vector-valued stochastic process $\vec{x}(t)$ over a certain time-interval, we often drop the index $t$ and denote it by a bold-face symbol $\mathbf{x}$. In other words, $\vec{x}(t)$ denotes the value of the stochastic process $x$ at time $t$. Unless stated otherwise, we use right derivatives throughout this paper, i.e., $\frac{d}{dt}x(t) = \lim_{\delta \downarrow 0^+} \frac{x(t+\delta) - x(t)}{\delta}$.

2 The System Model

We assume the following model for a wireless system with $N$ nodes and $L$ links.

2.1 The Channel

We assume that time is divided into time-slots of unit length. In order to model channel fading, we assume that at any time-slot $t$ the state of the wireless channel, denoted by $C(t)$, can be in one of $S$ (channel) states $j = 1, 2, ..., S$. We assume that the channel states $C(t), t = 1, 2, ...$ are i.i.d. across time. Let $p_j = \mathbf{P}[C(t) = j], j = 1, 2, ..., S$, and $\vec{p} = [p_1, ..., p_S]$.

For a fixed $B$, define the scaled channel state process on the time interval $[-T, 0]$ as

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

for $t = \frac{m}{B} - T, m = 0, ..., BT$, and by linear interpolation otherwise. The parameter $B$ is a scaling factor that scales in time and magnitude. To see this, note that (1) is of the form $\frac{1}{B}f(B(T + t))$.

The quantity $s_j^B(t)$ can be interpreted as the sum of the (scaled) time in $[-T, t]$ that the system is at channel state $j$. Further, it is easy to check that $\sum_{j=1}^{S} s_j^B(t) = t + T$ for all $t \in [-T, 0]$. Let $\vec{s}^B(t) = [s_1^B(t), ..., s_S^B(t)]$. Further, let $\phi_j^B(t) = \frac{d}{dt}s_j^B(t)$. (Note again that we use right-derivatives, and that the right-derivative of $s_j^B(t)$ is well defined almost everywhere on $[-T, 0]$ except when $t = m/B - T$ for some integer $m$. ) Let $\vec{\phi}^B(t) = [\phi_1^B(t), ..., \phi_S^B(t)]$. Note that $\sum_{j=1}^{S} \phi_j^B(t) = 1$ for almost all $t \in [-T, 0]$. 

7
We will soon see that the scaled channel state process satisfies a large-deviation principle (LDP). For any \( \vec{\phi} = [\phi_j, j = 1, \ldots, S] \geq 0 \) and \( \sum_{j=1}^{S} \phi_j = 1 \), define \( H(\vec{\phi} || \vec{p}) = \sum_{j=1}^{S} \phi_j \log \frac{\phi_j}{p_j} \).

(Here we use the convention that \( 0 \log 0 = 0 \).) Let \( \Phi_s[-T, 0] \) denote the space of functions \( \vec{s}(t) \) on \([-T, 0] \) that satisfy the following condition: \( \vec{s}(t) \) is component-wise non-decreasing on \([-T, 0] \), \( \vec{s}(-T) = 0 \), and \( \sum_{j=1}^{S} s_j(t) = t + T \) for all \( t \). Let this space be equipped with the essential supremum norm \([16, p176,p352]\), denoted here by \( \| \cdot \|_T \). The sequence of scaled channel-state processes \( \vec{s}^B = (\vec{s}^B(t), t \in [-T, 0]) \) are known to satisfy a sample-path large deviation principle \([16, p176]\) with large-deviation rate-function \( I_T^T(\vec{s}) \) given as follows:

\[
I_T^T(\vec{s}) = \int_{-T}^{0} H \left( \frac{d}{dt} \vec{s}(t) || \vec{p} \right) dt,
\]

if \( \vec{s} \in \Phi_s[-T, 0] \) is absolute continuous. (Note that \( \frac{d}{dt} \vec{s}(t) \) is well defined almost everywhere on \([-T, 0] \) when \( \vec{s} = (\vec{s}(t), t \in [-T, 0]) \) is absolute continuous.) Otherwise,

\[
I_T^T(\vec{s}) = +\infty.
\]

Such a large-deviation principle means that, for any set \( \Gamma \) of trajectories in \( \Phi_s[-T, 0] \), the probability that the sequence of scaled channel state processes \( \vec{s}^B \) fall into \( \Gamma \) must satisfy

\[
-\inf_{\vec{s} \in \Gamma^o} I_T^T(\vec{s}) \leq \liminf_{B \to \infty} \frac{1}{B} \log P[\vec{s}^B \in \Gamma] \leq \limsup_{B \to \infty} \frac{1}{B} \log P[\vec{s}^B \in \Gamma] \leq -\inf_{\vec{s} \in \bar{\Gamma}} I_T^T(\vec{s}),
\]

where \( \Gamma^o \) and \( \bar{\Gamma} \) denote the interior and closure, respectively, of the set \( \Gamma \). In addition, if \( \Gamma \) is a continuity set \([16, p5]\), equality is achieved and we then have,

\[
\lim_{B \to \infty} \frac{1}{B} \log P[\vec{s}^B \in \Gamma] = -\inf_{\vec{s} \in \Gamma} I_T^T(\vec{s}).
\]

**Remark:** The large-deviation rate-function \( I_T^T(\cdot) \) characterizes how rare the occurrence of each trajectory is. Note that \( I_T^T(\vec{s}) \geq 0 \) for all trajectory \( \vec{s} \). The larger the value of \( I_T^T(\vec{s}) \) is, the further the “empirical probability distribution” \( \frac{d}{dt} \vec{s}(t) \) deviates from the underlying probability distribution \( \vec{p} \). Hence, it is less likely that trajectory \( \vec{s} \) will occur. Equation (3) reflects the
well-known large-deviation philosophy that “rare events occur in the most-likely way.” Precisely, when $B$ is large, the probability that the scaled channel-state process $s^B$ falls into a set $\Gamma$ is determined by the trajectory in $\Gamma$ that is most likely to occur, i.e., with the smallest $I^*_s(s)$.

2.2 The Arrival

The system serves packets from multiple classes $k = 1, \ldots, K$. Each class $k$ corresponds to a set $D^k$ of destination nodes. In other words, once the class-$k$ packets arrive at any node in $D^k$, they will leave the system. Let $A^k_i(t)$ denote the number of class-$k$ packets arriving at node $i$ at time-slot $t$. We assume that $[A^k_i(t)]$ is i.i.d. across time*, and independent across arriving nodes and across classes. In addition, we assume that $A^k_i(t)$ is bounded by $M$ for all $i, k$ and $t$. Let $\lambda^k_i = E[A^k_i(t)]$.

The above arrival process also satisfies a well-known sample-path LDP as follows. Similar to (1), define the scaled arrival process as,

$$a^B_i(t) = \frac{1}{B} \sum_{\tau=0}^{B(T+t)} A^k_i(\tau) \quad \text{for } t = mB - T, m = 0, \ldots, BT,$$

and by linear interpolation otherwise. Let $\bar{a}^B(t) = [a^B_i(t), i = 1, \ldots, N, k = 1, \ldots, K]$ and $\bar{f}^B(t) = \frac{d}{dt} \bar{a}^B(t)$.

Define

$$L^k_i(f) = \sup_{\theta} \left( \theta f - \log E[\exp(\theta A^k_i(0))] \right).$$

For any $\bar{f} = [f^k_i, i = 1, \ldots, N, k = 1, \ldots, K]$, let $L(\bar{f}) = \sum_{i=1}^N \sum_{k=1}^K L^k_i(f^k_i)$. Further, let $\Phi_a[-T, 0]$ be the space of component-wise non-decreasing functions $\bar{a}(t)$ on $[-T, 0]$ with $\bar{a}(-T) = 0$. Let this space also be equipped with the essential supremum norm [16, p176,p352], denoted here by $|| \cdot ||^T_{\infty}$. Since the arrivals $[A^k_i(t)]$ are i.i.d. in time, the sequence of scaled arrival-processes $a^B = (\bar{a}^B(t), t \in [-T, 0])$ also satisfies a sample-path large-deviation principle [16, p176] with

*The results of this paper can also be generalized to the case with time-correlated arrivals under suitable conditions [4]. The key requirement is that the arrivals satify a large deviations principle.
large-deviation rate-function $I^T_a(a)$ given as follows:

$$I^T_a(a) = \int_{-T}^{0} L \left( \frac{d}{dt} \bar{a}(t) \right) dt,$$

if $a \in \Phi_a[-T, 0]$ is absolute continuous. (Note that $\frac{d}{dt} \bar{a}(t)$ is well-defined almost everywhere on $[-T, 0]$ when $a = (\bar{a}(t), t \in [-T, 0])$ is absolute continuous.) Otherwise,

$$I^T_a(a) = +\infty.$$

A similar interpretation as in Equations (2) and (3) holds for $I^T_a(\cdot)$ as well, where the set $\Gamma$, in this case is chosen from the space $\Phi_a[-T, 0]$.

2.3 The Queue

Let $X^k_i(t)$ denote the backlog of class-$k$ packets at node $i$ at time $t$, and let $\bar{X}(t) = [X^k_i(t), i = 1, ..., N, k = 1, ..., K]$. Let $b(l)$ and $e(l)$ denote the source-node and end-node, respectively, of link $l$. Each link $l$ then corresponds to a server with time-varying service rate, which serves packets at node $b(l)$ and transfers them to node $e(l)$. The service offered by link $l$ is determined by the scheduling and routing algorithm. In general, this service rate may depend on the global system backlog and the global channel state, and hence may correlate with the service at other links/nodes. Let $E^k_l(j, \bar{X})$ denote the service offered by link $l$ to class-$k$ packets at node $b(l)$, when the state of the system is $j$ and the global backlog is $\bar{X}$. We impose the additional constraint that $E^k_l(j, \bar{X}) \leq X^k_{b(l)}$. (In other words, the service offered by link $l$ to class-$k$ packets is no greater than the backlog of class-$k$ packets at the source node $b(l)$.) By definition of $\mathcal{D}^k$ (the set of destination nodes of class-$k$), $X^k_i(t) = 0$ for all nodes $i \in \mathcal{D}^k$. For a node $i$ not in $\mathcal{D}^k$, the evolution of the class-$k$ backlog is given by

$$X^k_i(t + 1) = X^k_i(t) + A^k_i(t)$$

$$- \sum_{j=1}^{S} 1_{\{C(t)=j\}} \sum_{l=1}^{L} R_{il} E^k_l(j, \bar{X}(t)), \quad (5)$$

$$10$$
for all nodes $i \not\in \mathcal{D}^k$, where $R_{il}$ denote the connectivity matrix, i.e.,

$$R_{il} = \begin{cases} 
1, & \text{if } i = b(l) \\
-1, & \text{if } i = e(l) \\
0, & \text{otherwise.}
\end{cases}$$

Let $E_l(j, \tilde{X}) = \sum_{k=1}^{K} E_{lk}(j, \tilde{X})$, which denotes the aggregate service offered by link $l$. We assume that, for each state $j$, the service-rate vector $[E_l(j, \tilde{X}), l = 1, \ldots, L]$ must belong to a set $\mathcal{E}_j$ of feasible service-rate vectors. We assume that for all $j$, the convex hull of $\mathcal{E}_j$, denoted $\text{Conv}(\mathcal{E}_j)$, is closed and bounded, and contains the intersection of a neighborhood of the origin and the positive quadrant.

### 2.4 The Performance Measure

Assume that the offered load $\tilde{\lambda} = [\lambda^k_i, i = 1, \ldots, N, k = 1, \ldots, K]$ is such that the system is stationary and ergodic (in Section 3.2 we will introduce conditions on the Lyapunov function that imply this stability). In this paper, we will focus on the stationary probability that some chosen norm of the system backlog exceeds a certain threshold $B$. In particular, let $\epsilon$ denote our target value of the overflow probability, we would like to ensure that

$$\mathsf{P}[[\|X(0)\| \geq B]] \leq \epsilon,$$

where $\| \cdot \|$ is an appropriately chosen norm, and $B$ is the overflow threshold. Unfortunately, the problem of calculating the exact probability $\mathsf{P}[[\|X(0)\| \geq B]]$ is often mathematically intractable. Instead, we will be interested in the decay rate of the queue-overflow probability, which is defined by

$$I_0(\tilde{\lambda}) \triangleq - \lim_{B \to \infty} \frac{1}{B} \log \mathsf{P}[[\|X(0)\| \geq B]],$$

whenever the limit on the right-hand-side exists. Note that if (7) holds, then when $B$ is large, the overflow probability can be approximated by

$$\mathsf{P}[[\|X(0)\| \geq B]] \approx \exp(-BI_0(\tilde{\lambda})).$$
Hence, we refer to $I_0(\tilde{\lambda})$ as the asymptotic decay-rate of the queue-overflow probability. Using the above approximation, in order to (approximately) satisfy the constraint (6), we only need to ensure that

$$I_0(\tilde{\lambda}) \geq \theta \triangleq -\frac{\log \epsilon}{B}. \quad (8)$$

We can then define the effective capacity region as the set of arrival rates $\tilde{\lambda}$ such that the above inequality holds. In more general settings, the limit in (7) may not exist or may not be easily computed. Still, we may be able to find a lower bound $I'_0(\tilde{\lambda})$ of the decay-rate of the overflow probability such that

$$\limsup_{B \to \infty} \frac{1}{B} \log P[||X(0)|| \geq B] \leq -I'_0(\tilde{\lambda}). \quad (9)$$

Then $\exp(-B I'_0(\tilde{\lambda}))$ provides an (approximate) upper bound on the overflow probability $P[||X(0)|| \geq B]$, and we can find a lower bound on the effective capacity region as the set of $\tilde{\lambda}$ such that $I'_0(\tilde{\lambda}) \geq \theta \triangleq -\frac{\log \epsilon}{B}$. In this paper, our goal is to provide effective techniques to estimate $I_0(\tilde{\lambda})$ and/or $I'_0(\tilde{\lambda})$.

### 2.5 Examples

Before we proceed, we provide a few examples to clarify the above system model.

#### 2.5.1 Example 1: cellular or access-point based networks

Consider a single cell in a cellular network or an access-point based network. Each user communicates directly with the basestation. Further, only one user can be selected for service at a time. Let us focus on the downlink from the basestation to the users (the uplink can be treated analogously). Since this is a single-hop model, we identify each link $l$ with a particular user/class, and hence we can drop the index $k$ for traffic class. In other words, we use $A_l(t)$ to denote the packets generated for the user associated with link $l$ at time slot $t$, and use $X_l(t)$ to denote the backlog of link $l$ at time $t$. Let $E_l(j, X)$ be the service offered to link $l$ when the channel state is $j$ and the global backlog is $\bar{X} = [X_l, l = 1, ..., L]$. Imposing the constraint that $E_l(j, \bar{X}) \leq X_l$, we
the evolution of the queue-length is then given by

\[ X_l(t + 1) = X_l(t) + A_l(t) - \sum_{j=1}^{S} 1_{\{C(t)=j\}} E_l(j, \vec{X}(t)). \]  

(10)

Note that this equation is a simplified version of (5).

When the channel state is \( j \), let \( r_{lj} \) denote the capacity of link \( l \) if it is selected for transmission. Then for each state \( j \) the service-rate vector \( [E_l(j, \vec{X}), l = 1, \ldots, L] \) must belong to the set \( \mathcal{E}_j \) given by

\[ \mathcal{E}_j = \{ [t_l r_{lj}, l = 1, \ldots, L] : t_l \in [0, 1] \} \]

for all \( l \), and only one element of \( [t_l] \) is non-zero.

A Queue-Length-Based Scheduling Policy: One important class of policies for choosing \( E_l(j, \vec{X}) \) (often referred to as the max-weight policy in the literature [1, 17]), is to serve the link \( l \) such that the weighted queue-length \( r_{lj} X_l \) is the largest among all users (ties can be broken arbitrarily). Let \( l^* = \arg\max_l r_{lj} X_l \). The policy can then be written as:

\[ E_l(j, \vec{X}) = \begin{cases} \min\{r_{lj}, X_l\}, & \text{if } l = l^* \\ 0, & \text{otherwise} \end{cases} \]

It is well-known that this class of policies are throughput-optimal, i.e., they can stabilize the system under the largest set of offered loads [17].

Remark: When the arrivals \( A_l(t) \) are constant, i.e., \( A_l(t) = \lambda_l \) for all \( t \), the constraint in (6) becomes equivalent to a constraint on the delay-violation probability. To see this, assume that new packets always arrive at the beginning of a time-slot, and define the delay of a packet as the number of time-slots until it is completely served\(^1\). Let \( D_l(t) \) denote the delay experienced by the last packet that arrives at time slot \( t \). If we use the convention that \( X_l(t) \) denotes the queue-length for link \( l \) at the beginning of time slot \( t \) (before arrivals have occurred), then for any

\(^1\)When the service rates are real numbers, we implicitly assume that the packet can be divided into arbitrarily small elements, and these elements can be served over multiple time-slots.
delay bound $d_t$, we have

$$D_t(0) \geq d_t$$

$$\iff \text{Cumulative service over time } [0, d_t - 1] < X_t(0) + \lambda_t$$

$$\iff \sum_{t=0}^{d_t-1} \sum_{j=1}^{S} 1_{(C(t)=j)} E_t(j, X(t)) < X_t(0) + \lambda_t$$

$$\iff X_t(d_t) > \lambda_t(d_t - 1).$$

Hence, the two types of constraints are related by (see [9, 11])

$$P[D_t(0) \geq d_t] = P[X_t(d_t) > \lambda_t(d_t - 1)]$$

$$= P[X_t(0) > \lambda_t(d_t - 1)],$$

where the last equality follows from the stationary assumption.

### 2.5.2 Example 2: Multihop Ad Hoc Wireless Networks

The full model described earlier in this section is most suitable for multihop wireless networks, where $X_i^k(t)$ represents the backlog of class-$k$ packets at node $i$ at time $t$, and its evolution is given by (5). Unlike in Example 1 where only one user/link can be selected at each time, in multihop wireless networks it is often possible to activate multiple links simultaneously. However, the activated links must satisfy certain interference constraints, and their rates depend on the power and interference levels. As a result, the set of feasible service-rate vectors often becomes more complicated to describe than Example 1. We give two cases.

**Case 1:** Assume that the scheduling policy can decide whether to activate or inactivate a link, but cannot change the transmission power of a link. Let $\pi_l = 1$ if link $l$ is activated, and $\pi_l = 0$, otherwise. Let $\bar{\pi} = [\pi_l, l = 1, ..., L]$, and let $\Pi$ denote the set of feasible activation vectors $\bar{\pi}$. Let $r_{lj}(\bar{\pi})$ denote the rate of link $l$ if the activation vector $\bar{\pi}$ is applied at state $j$. Then at each channel state $j$, the set $E_j$ of feasible service-rate vectors can be written as:

$$E_j = \{ [r_{lj}, l = 1, ..., L] : \text{there exists } \bar{\pi} \in \Pi$$

$$\text{such that } r_{lj} \leq r_{lj}(\bar{\pi}) \text{ for all links } l \}.$$ (11)
Case 2: Assume that the scheduling policy can decide both the activation pattern and the power assignments. Let $\pi_l$ denote the power assignment of link $l$. $\pi_l = 0$ if the link is not activated. Let $\vec{\pi}$ denote the set of feasible power-assignment vectors $\vec{\pi}$. Then at each channel state $j$, each vector $\vec{\pi}$ can again be mapped to a rate-vector $[r_{lj}(\vec{\pi}), l = 1, ..., L]$. The set $\mathcal{E}_j$ of feasible service-rate vectors can also be written as in (11).

A Queue-Length-Based Scheduling and Routing Algorithm: For both cases, the scheduling and routing algorithm proposed in [1] (which is often referred to as the “back-pressure” algorithm) is known to be throughput-optimal. At each time-slot $t$, for each link $l$, first find the class $k^*_l(t)$ with the maximum differential backlog, i.e.,

$$k^*_l(t) = \arg\max_k (X^k_{b(l)}(t) - X^k_{c(l)}(t)).$$

Let the corresponding differential backlog be

$$w_l(t) = \max\{0, X^{k^*_l(t)}_{b(l)}(t) - X^{k^*_l(t)}_{c(l)}(t)\}.$$

Then, when the channel state at time $t$ is $j$, compute the schedule $\vec{\pi}^*_j(t)$ that maximizes the sum of the rates weighted by $w_l(t)$, i.e.,

$$\vec{\pi}^*_j(t) = \arg\max_{\vec{\pi} \in \Pi} w_l(t)r_{lj}(\vec{\pi}).$$

The scheduling and routing decision is then given by the following: if the channel state at time-slot $t$ is $j$,

- **Scheduling**: use the activation vector $\vec{\pi}^*_j(t)$.

- **Routing**: on each link $l$, only serve the packets belonging to class $k^*_l(t)$.

In other words, the service rate vector is given by

$$E^k_l(j, \vec{X}) = 
\begin{cases} 
\min\{X^k_{b(l)}(t), r_{lj}(\vec{\pi}^*_j(t))\}, & \text{if } k = k^*_l(t) \\
0, & \text{otherwise.}
\end{cases}$$
3 A General Lower Bound on the Decay-Rate of the Queue-Overflow Probability

In this section, we first provide a lower bound \( I'_0(\tilde{\lambda}) \) on the decay-rate of the queue-overflow probability as defined in (9). Prior work [10, 11] has derived the decay-rate of the queue-overflow probability as the cost of the most-likely-path to overflow. The approach there requires that the limiting mapping from the scaled channel-state process to the scaled queue-backlog process, to be defined later, is unique and continuous with respect to a suitably chosen topological space, so that the contraction principle [16, p131] can be invoked to establish a sample-path LDP for the queue-backlog process. However, for general wireless systems, the mapping from the channel-state process to the queue-backlog process may not be continuous, and hence the approach in [10, 11] cannot be applied. Nonetheless, in this section we show that, for a large class of wireless systems, the cost of the most-likely-path to overflow is a lower bound of the decay-rate of the queue-overflow probability. Recall that such a lower bound implies that we can then use \( \exp(-BI'_0(\tilde{\lambda})) \) as an (approximate) upper bound on the overflow probability \( P[||X(0)|| \geq B] \).

Similar to the large-deviation scaling used in Section 2.1 and Section 2.2, define the scaled backlog process as,

\[
x_{i}^{k:B}(t) = \frac{1}{B}X_{i}^{k}(B(T + t)),
\]

for \( t = \frac{m}{B} - T, \ m = 0, ..., BT \), and by linear interpolation otherwise. Let \( \bar{x}^B(t) = [x_{i}^{k:B}(t), i = 1, ..., N, k = 1, ..., K] \). The lower bound \( I'_0(\tilde{\lambda}) \) in (9) can then be written as

\[
\limsup_{B \to \infty} \frac{1}{B} \log P[||\bar{x}^B(0)|| \geq 1] \leq -I'_0(\tilde{\lambda}).
\]

Before we proceed, we need to define the notion of a Fluid Sample Path (FSP). Note that according to (5), given an initial condition on \( \bar{x}^B(-T) \), the scaled backlog process \( x^B = (\bar{x}^B(t), t \in [-T,0]) \) is related to the scaled channel-state process \( s^B = (\bar{s}^B(t), t \in [-T,0]) \) and the scaled
arrival process \( a^B = (\bar{a}^B(t), t \in [-T, 0]) \) by

\[
\frac{x^{k,B}_i(t + 1/B) - x^{k,B}_i(t)}{1/B} = \left[ \frac{a^{k,B}_i(t) - a^{k,B}_i(t - 1/B)}{1/B} \right] - \sum_{j=1}^{S} \sum_{l=1}^{L} R_d E^k_l(j, B\bar{x}^B(t)) \left[ \frac{s^{B}_j(t) - s^{B}_j(t - 1/B)}{1/B} \right],
\]

for \( i \notin D^k, t = \frac{mT}{B} - T, m = 0, ..., BT \) and by linear interpolation otherwise. Thus, given any \( T \) and any initial condition \( \bar{x}^B(-T) \), Equation (14) defines a mapping from the scaled channel-state process \( s^B \) and the scaled arrival process \( a^B \) to the scaled backlog process \( x^B \) over the time-interval \([-T, 0]\). As \( B \to \infty \), take any sequence of \( s^B \) and \( a^B \) (which may come from different sample paths). They map to a sequence of scaled backlog processes \( x^B \). Note that \( s^B, a^B \) and \( x^B \) are all Lipschitz-continuous. Hence, for any such sequence \( (s^B, a^B, x^B) \), there must exist a subsequence that converges to a limiting process \( (s, a, x) \) uniformly over compact intervals. We define any such limiting process as a Fluid Sample Path, or an FSP, which we denote as \( (s, a, x)_T \), where the subscript denotes the starting time \(-T\). Such an FSP often satisfies the following differential equation obtained by letting \( B \to \infty \) on Equation (14):

\[
\frac{d}{dt} x^k_i(t) = \frac{d}{dt} a^k_i(t) - \sum_{j=1}^{S} \frac{d}{dt} s^k_j(t) \sum_{l=1}^{L} R_d e^k_l(j, \bar{x}(t)).
\]

where \( e^k_l(j, \bar{x}(t)) \) is some appropriately-chosen limiting value of \( E^k_l(j, B\bar{x}^B(t)) \). In the rest of the paper, we sometimes denote FSPs as \( (s, a, x) \), i.e., without the time subscript \( T \). In doing so, we mean that there is some finite time \( T \) such that \( (s, a, x)_T \) is an FSP.

Let

\[
I'_0(\bar{\lambda}) = \inf_{T \geq 0, (s, a, x)} \int_{-T}^{0} \left[ H \left( \frac{d}{dt} \bar{s}(t) || \bar{\eta} \right) + L \left( \frac{d}{dt} \bar{a}(t) \right) \right] dt
\]

subject to \((s, a, x)_T \) is an FSP

\[
\bar{x}(-T) = 0, ||\bar{x}(0)|| \geq 1.
\]
Our goal in this section is to establish that $I'_0(\lambda)$ defined in (16) satisfies (13), i.e., it is a lower-bound on the decay rate of the queue-overflow probability. The FSP that attains the infimum, on the right-hand-side of (16) (if such an FSP exists) is usually called the “most-likely path to overflow.” Hence, this result (to be proved soon) states that the cost of the most-likely path to overflow is a lower bound on the decay-rate of the queue-overflow probability. We will derive the result in two steps. First, we consider a system that starts at time 0, and derive a lower bound in Section 3.1 on the decay rate of the overflow probability at time $BT$. Then, we let $T \to \infty$ and derive a lower bound on the decay-rate of the stationary overflow probability in Section 3.3.

### 3.1 Bounds for a Finite Time System

As a step towards proving the result on the stationary overflow probability, we first consider the probability of overflow at time $BT$, $P(||\vec{X}(BT)|| \geq B)$, for a system that starts from $\vec{X}(0) = 0$. Note that according to the transformation $\vec{x}^B(t) = \frac{1}{B} \vec{X}(B(T + t))$, we have $\vec{x}^B(-T) = 0$ and $||\vec{x}^B(0)|| \geq 1$. Let $P^{B,T}_0$ denote the probability distribution conditioned on $\vec{x}^B(-T) = 0$. We have the following bound on the probability of overflow.

**Proposition 1** Fix $T > 0$. The following holds,

$$
\limsup_{B \to \infty} \frac{1}{B} \log P^{B,T}_0[||\vec{x}^B(0)|| \geq 1] 
\leq - \inf_{s,a,x} \int_{-T}^0 \left[ H \left( \frac{d}{dt} \vec{s}(t)||\vec{p} \right) + L \left( \frac{d}{dt} \vec{a}(t) \right) \right] dt 
$$

subject to $(s,a,x)_T$ is an FSP

$$
\vec{x}(-T) = 0 \text{ and } ||\vec{x}(0)|| \geq 1. 
$$

(17)

Instead of proving Proposition 1, we will prove a generalized version in Proposition 2. The extra effort will serve useful in proving the stationary overflow case.

Fix $\vec{x}_0$. For the more general version, consider a system that starts with $\vec{X}(0) = B\vec{x}_0$ at time 0 (i.e., $\vec{x}^B(-T) = \vec{x}_0$). Let $P^{B,T}_{\vec{x}_0}$ denote the probability distribution conditioned on $\vec{x}^B(-T) = \vec{x}_0$. Let $\Phi_x[-T,0]$ denote the space of non-negative Lipschitz-continuous functions on the interval
Proposition 2 Fix $T > 0$. Let $\mathcal{X}$ denote a closed set in $\mathbb{R}^{NK}$. Let $\Gamma$ denote a closed set of trajectories $x = (\bar{x}(t), t \in [-T, 0])$ from the topological space $\Phi_x[-T, 0]$ that satisfies $\bar{x}(-T) \in \mathcal{X}$. The following holds,

$$
\limsup_{B \to \infty} \frac{1}{B} \log \sup_{x_0 \in \mathcal{X}} P^{B,T}_{\bar{x}_0}[x^B \in \Gamma] \leq - \inf_{sk, a, x} \int_{-T}^{0} \left[ H \left( \frac{d}{dt} \bar{s}(t) || \bar{p} \right) + L \left( \frac{d}{dt} \bar{a}(t) \right) \right] dt \text{ subject to } (s, a, x)_{-T} \text{ is an FSP.

(18)}$$

Proof: Note that $x^B$ is related to $s^B$ and $a^B$ according to Equation (14). Let $\bar{\Gamma}^B$ denote the set of $(s^B, a^B)$ on the interval $[-T, 0]$ such that there exists $\bar{x}_0 \in \mathcal{X}$ with which $(s^B, a^B)$ maps to a backlog process $x^B$ that starts from $\bar{x}^B(-T) = \bar{x}_0$ and satisfies $x^B \in \Gamma$. Then for every $\bar{x}_0 \in \mathcal{X},$

$$P^{B,T}_{\bar{x}_0}[x^B \in \Gamma] \leq P[(s^B, a^B) \in \bar{\Gamma}^B].$$

Note that the right-hand-side does not depend on $\bar{x}_0$. Further, for any fixed $n$, $\bar{\Gamma}^B \subset \cup_{B' = n}^{\infty} \bar{\Gamma}^{B'}$ when $B \geq n$. Hence, we have, for any fixed $n,$

$$
\limsup_{B \to \infty} \frac{1}{B} \log \sup_{x_0 \in \mathcal{X}} P^{B,T}_{\bar{x}_0}[x^B \in \Gamma] \leq \limsup_{B \to \infty} \frac{1}{B} \log P[(s^B, a^B) \in \cup_{B' = n}^{\infty} \bar{\Gamma}^{B'}].
$$

Using the sample-path LDP of $s^B$ and $a^B$ (see (2)) and the fact that they are independent, we have

$$\limsup_{B \to \infty} \frac{1}{B} \log P[(s^B, a^B) \in \cup_{B' = n}^{\infty} \bar{\Gamma}^{B'}] \leq - \inf_{(s, a)} \int_{-T}^{0} \left[ H \left( \frac{d}{dt} \bar{s}(t) || \bar{p} \right) + L \left( \frac{d}{dt} \bar{a}(t) \right) \right] dt \text{ subject to } (s, a) \in \cup_{B' = n}^{\infty} \bar{\Gamma}^{B'}.$$
where $\bigcup_{B=n}^{\infty} \overline{\Gamma}^B$ denotes the closure of the set $\bigcup_{B=n}^{\infty} \overline{\Gamma}^B$. Note that this inequality holds for all $n$. Further, since the set $\bigcup_{B=n}^{\infty} \overline{\Gamma}^B$ is decreasing in $n$, the right-hand-side is decreasing in $n$ as well. Therefore, we tighten the bound by letting $n \to \infty$. To simplify notation, for any $(s, a)$, define its cost by

$$J_T(s, a) \triangleq \int_{-T}^{0} \left[ H \left( \frac{d}{dt} \tilde{s}(t) || \tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) \right] dt.$$ 

We then have,

$$\limsup_{B \to \infty} \frac{1}{B} \log \sup_{\bar{x}_0 \in \mathcal{X}} \mathbb{P}_{\bar{x}_0}^{B,T} [ x^B \in \Gamma]$$

$$\leq - \lim_{n \to \infty} \inf_{(s, a) \in \bigcup_{B=n}^{\infty} \overline{\Gamma}^B} J_T(s, a).$$

Let

$$Y = \lim_{n \to \infty} \inf_{(s, a) \in \bigcup_{B=n}^{\infty} \overline{\Gamma}^B} J_T(s, a). \quad (19)$$

It remains to show that

$$Y \geq \inf_{s, a, x} J_T(s, a)$$

subject to $(s, a, x)_T$ is an FSP

$$x \in \Gamma.$$ 

To see this, note that by (19), there must exist a sequence $(s_n, a_n), n = 1, 2, \ldots$ such that

$$(s_n, a_n) \in \bigcup_{B=n}^{\infty} \overline{\Gamma}^B \text{ for all } n,$$

and

$$\lim_{n \to \infty} J_T(s_n, a_n) = Y.$$ 

Since both $s_n$ and $a_n$ are non-decreasing and Lipschitz-continuous, there must exist a further subsequence that converges uniformly over compact intervals. Without loss of generality, we can abuse notation and denote this subsequence also as $(s_n, a_n)$, and let $(\bar{s}, \bar{a})$ be the corresponding limit. Then, due to the lower-semicontinuity of the large-deviation rate function $J_T(\cdot, \cdot)$ [16, p4], we must have

$$J_T(\bar{s}, \bar{a}) \leq Y.$$
We now show that we must then be able to find an FSP \((s, a, x)\) with \(x \in \Gamma\). To see this, note that by definition each \((s_n, a_n)\) also corresponds to a sequence \((s_{n,m}, a_{n,m})\) such that \((s_{n,m}, a_{n,m}) \in \bigcup_{B=n}^{\infty} \Gamma^B\) for all \(m = 1, 2, \ldots\), and \((s_{n,m}, a_{n,m})\) converges to \((s_n, a_n)\) uniformly over compact intervals. Assign any sequence \(\epsilon_n > 0, n = 1, 2, \ldots\), such that \(\lim_{n \to \infty} \epsilon_n = 0\). For each \(n\), we can then find an element \((\tilde{s}_n^B, \tilde{a}_n^B)\) from the sequence \((s_{n,m}, a_{n,m})\) such that

\[
||\tilde{s}_n^B(t) - s_n(t)||_{\infty} + ||\tilde{a}_n^B(t) - a_n(t)||_{\infty} \leq \epsilon_n,
\]

and \((\tilde{s}_n^B, \tilde{a}_n^B) \in \Gamma^B\) for some \(B_n \geq n\). Since the sequence \((s_n, a_n)\) converges to \((s, a)\) uniformly over compact intervals, we must have that \((\tilde{s}_n^B, \tilde{a}_n^B)\) also converges to \((s, a)\) uniformly over compact intervals. Further, since each \((\tilde{s}_n^B, \tilde{a}_n^B) \in \Gamma^B\), there must exist a corresponding backlog process \(\tilde{x}_n^B\) such that \(\tilde{x}_n^B \in \Gamma\). Take a further subsequence of \((\tilde{s}_n^B, \tilde{a}_n^B)\) such that the corresponding subsequence of \(\tilde{x}_n^B\) converges uniformly over compact intervals to a limiting backlog process \(\tilde{x}\). Then, since the set \(\Gamma\) is closed, this limiting process \(\tilde{x}\) must also satisfy \(\tilde{x} \in \Gamma\). Hence, \((s, a, \tilde{x})\) is an FSP, and it satisfies the constraints used to define the right hand side of (18). We then have

\[
Y \geq J_T(s, a) \geq \inf_{s, a, x} J_T(s, a)
\]

subject to \((s, a, x)\) is an FSP

\[
x \in \Gamma.
\]

The result then follows. \(Q.E.D.\)

By setting \(X = \{0\}\), and

\[
\Gamma = \{x : \tilde{x}(-T) = 0 \text{ and } ||\tilde{x}(0)|| \geq 1\},
\]

we then recover the result of Proposition 1.

Next, we let \(T \to \infty\). Intuitively, as the starting point of \(\tilde{x}(\cdot)\), i.e., \(-T\), tends to \(-\infty\), one would expect that the distribution \(P_0^{B,T}\) becomes closer and closer to the stationary distribution

21
We can then derive a lower bound similar to (16). In the literature, such a limiting argument is carried out using the so-called Freidlin-Wentzell theory (see [18, Chapter 6] and [13]). Often, to apply the Freidlin-Wentzell theory, one will need to impose additional restrictions on the system model [18, p133]. In the next subsection, we provide a fairly general condition for this limiting argument to hold. As readers will see soon, our condition essentially requires that there exists a Lyapunov function, with respect to which the system is stable.

### 3.2 Lyapunov Functions

In this subsection, we introduce some fairly standard conditions on typical Lyapunov functions, which will then be used in Section 3.3. We first introduce the notion of the fluid limit [19]. Take the scaled queue-evolution equation (14), and take a sequence of \((s^B, a^B, x^B)\) with \(B \to +\infty\). For a large class of dynamic systems, one can show that, with probability 1, there must exist a subsequence \((s^{B_n}, a^{B_n}, x^{B_n})\) that converges uniformly over compact intervals. Let the last component of this limit be denoted as \(x\), which is called the fluid limit of the system [19]. This fluid limit can often be written in the form

\[
\frac{d}{dt} x^k_i(t) = \lambda^k_i - \sum_{j=1}^S p_j \sum_{l=1}^L R_i e^k_l(j, \bar{x}(t)).
\]

(20)

where \(e^k_l(j, \bar{x}(t))\) is some appropriate limit of \(E^k_l(j, \bar{X}(t))\).

**Remark:** Note the similarity between Equation (20) and Equation (15). We briefly comment on the difference between a fluid limit and a fluid sample path (FSP). In essence, a fluid limit is a fluid sample path with \(\frac{d}{dt} \bar{s}(t) = \bar{p}\) and \(\frac{d}{dt} \bar{a}(t) = \bar{\lambda}\) for all \(t\), which is why in Equation (20) \(\lambda^k_i\) and \(p_j\) replace \(\frac{d}{dt} a^k_i(t)\) and \(\frac{d}{dt} s_j(t)\), respectively, from (15). This is the case because the fluid limit is the almost sure limit of the scaled process, and hence a Law of Large Numbers type of argument can be invoked. In other words, the fluid limit dynamics can be viewed as the mean behavior of the system. In contrast, for an FSP we are interested in rare events, and hence the values of \(\frac{d}{dt} \bar{a}(t)\) and \(\frac{d}{dt} \bar{s}(t)\) can deviate from their mean values.
We say that the fluid limit model of the system is stable if there exists a $T > 0$ such that for all fluid limits with $|x(-T)| = 1$, we must have $x(0) = 0$. The main result of [19] shows that, if the fluid limit model of the system is stable, then the original system is also stable (in the sense of Harris positive recurrence).

A standard technique to prove the stability of the fluid limit model of a dynamic system is to use a Lyapunov function $V(\vec{x})$ and to show that it has a negative drift. Specifically, in the rest of the paper, we assume that the system has a Lyapunov-function $V(\vec{x})$ that satisfies the following properties.

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

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

(f) Either of the following holds (for a fixed arrival rate $\bar{\lambda}$ and a fixed channel state distribution $\bar{p}$ assumed in the system model). For all fluid limits $x$,

$$\frac{d}{dt} V(\vec{x}(t)) \triangleq \left( \frac{\partial V}{\partial \vec{x}} \right)^T \frac{d\vec{x}}{dt} \leq -\eta V^{\alpha}(\vec{x}(t)), \quad (21)$$

for almost all $t$, where $0 < \alpha < 1$ and $\eta$ is a positive constant. Or, for all fluid limits $x$,

$$\frac{d}{dt} V(\vec{x}(t)) \triangleq \left( \frac{\partial V}{\partial \vec{x}} \right)^T \frac{d\vec{x}}{dt} \leq -\eta \quad (22)$$

for almost all $t$, where $\eta$ is a positive constant.
Remark: Parts (a)-(c) and (f) of the assumption are typical when one uses Lyapunov functions to establish stability. Although Parts (d) and (e) are not standard, with a proper scaling of the Lyapunov function they will hold for many Lyapunov functions that have been used for wireless systems. Specifically, Part (d) holds (after proper scaling) when $\max_{||\vec{x}|| \leq C} V(\vec{x})$ is upper bounded for some constant $C > 0$, and Part (e) holds when the Lyapunov function has bounded gradients in any finite set. To see how these conditions imply the stability of the fluid limit, note that starting from any initial $\vec{x}(-T)$ with $||\vec{x}(-T)|| = 1$, we must have $||V(\vec{x}(-T))|| \leq \tilde{C}$ from Part (d) of the assumption. Then, using the drift condition (f), we can find a value of $T$ such that for all fluid limits with $||V(\vec{x}(-T))|| \leq \tilde{C}$, we must have $V(\vec{x}(0)) = 0$. Using Part (b), this then implies that $\vec{x}(0) = 0$. Hence, the fluid limit model of the system is stable. By [19], it implies that the original system is positive Harris recurrent. Note that the two drift conditions in (21) and (22) are in fact equivalent. If $V(\cdot)$ satisfies (21), then $U(\vec{x}) = \frac{V^{1-\alpha}(\vec{x})}{1-\alpha}$ satisfies (22).

We now provide examples of Lyapunov functions for the examples discussed in Section 2.5.

**Example 1 (Cellular or access-point-based networks):**

$$V_1(\vec{x}) = \sqrt{\sum_{l=1}^{L} x_l^2}$$

is a Lyapunov function that applies to this system.

**Example 2 (Ad-Hoc Wireless Networks):** A Lyapunov function for the back-pressure algorithm in Section 2.5.2 is

$$V_2(\vec{x}) = \sqrt{\sum_{i=1}^{N} \sum_{k=1}^{K} (x_{ik})^2}.$$ 

It is easy to see that these functions satisfies parts (a) to (e) of Assumption 1.

For part (f) we would like to point out a small but important difference between Lyapunov functions for the original discrete-time systems and Lyapunov functions for fluid-limit models. In part (f), we state the assumption using Lyapunov functions for the fluid-limit model. In the literature, [1,20], the stability of the system is often established through a Lyapunov function for the original discrete-time system, which are of the form in Example 1 and 2. For the scheduling
algorithms in Example 1 and 2, it is not difficult to show that part (f) holds for the Lyapunov functions when applied to the fluid-limit model (see e.g. Proposition 5 of [15]). We also provide another example in Proposition 10. However, we caution that, in general, part (f) may not be satisfied by all Lyapunov functions used for the discrete-time system.

3.3 The Stationary Overflow Probability

We now use the Freidlin-Wentzell theory (see [18, Chapter 6] and [13]) to derive a lower bound on the decay-rate of the stationary queue-overflow probability. We need slightly stronger conditions on the Lyapunov functions than those stated in the previous subsection. Recall that an FSP($s, a, x$)$_T$ is the limit of some sequence of scaled processes ($s^n, a^n, x^n$) over the interval $[-T, 0]$.

**Assumption 2** If the Lyapunov function $V(\cdot)$ satisfies (21), then

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

\[
\frac{d}{dt} V(\bar{x}(t)) \leq -\frac{\eta}{2} V^\alpha(\bar{x}(t)),
\]  

where $0 < \alpha < 1$ and $\eta > 0$ are the same constants as in (21).

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

\[
\frac{d}{dt} V(\bar{x}(t)) \leq M_1 V^\alpha(\bar{x}(t)).
\]

On the other hand, if the Lyapunov function $V(\cdot)$ satisfies (22), then

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

\[
\frac{d}{dt} V(\bar{x}(t)) \leq -\frac{\eta}{2},
\]  

where $\eta > 0$ is the same constant as in (22).
(b) For any $\delta > 0$, there exists $M_1 \geq 0$ such that for all FSP $(s, a, x)_T$ and for all time $t$ with

$$
\| \frac{d}{dt}s(t) - \bar{p}\| \geq \delta \text{ or } \| \frac{d}{dt}a(t) - \bar{\lambda}\| \geq \delta,
$$

the following holds,

$$
\frac{d}{dt} V(\bar{x}(t)) \leq M_1.
$$

(26)

Remark: Although this additional set of assumptions are not used in typical definitions of Lyapunov functions for stability, they do hold for most Lyapunov functions that we work with in wireless systems, including the Lyapunov functions $(V_1(\cdot)$ and $V_2(\cdot)$) in the examples in Section 3.2. We will need these assumptions when we study the large-deviation properties of the stationary queue-overflow probability. Essentially, they imply that the drift behaves nicely not only for the fluid limits but also for FSPs. Specifically, (a) even if we perturbed the channel distribution and the distribution of the arrival process slightly from $(\bar{p}, \bar{\lambda})$, the drift of the Lyapunov function still remains negative; (b) if the perturbation is large, although the drift could become positive, it is upper-bounded by a constant $M_1$ (or this constant multiplied by $V^\alpha(\bar{x}(t))$).

Again, note that the two parts of Assumption 2 are also equivalent. If $V(\cdot)$ satisfies the first part of the assumption, then $U(\bar{x}) = \frac{V^{1-\alpha}(\bar{x})}{1-\alpha}$ satisfies the latter part. We will provide an example in Section 6 how these conditions can be easily verified.

We can now state the main result of the section.

**Proposition 3** Assume that there exists a Lyapunov function $V(\cdot)$ that satisfies Assumptions 1 and 2. Then the following holds,

$$
\limsup_{B \to \infty} \frac{1}{B} \log P[\|x^B(0)\| \geq 1] 
\leq - \inf_{T \geq 0, s, a, x} \int_{-T}^{0} \left[ H \left( \frac{d}{dt}s(t)\|\bar{p}\right) + L \left( \frac{d}{dt}a(t)\right) \right] dt
$$

subject to $(s, a, x)_T$ is an FSP

$$
\bar{x}(-T) = 0, \|\bar{x}(0)\| \geq 1.
$$

(27)

Note that the differences between (17) and (27) is that, the infimum in (17) is for a fixed $T$, while the infimum in (27) is taken over all $T > 0$. 

26
The proof of Proposition 3 is very similar to the proof of Proposition 4 that follows. Hence, in order to avoid repetition we omit its proof.

**Proposition 4** Assume that there exists a Lyapunov function $V(\cdot)$ that satisfies both Assumption 1 and Assumption 2. Then the following holds,

$$
\limsup_{B \to \infty} \frac{1}{B} \log P[V(\bar{x}^B(0)) \geq 1] \\
\leq - \inf_{T \geq 0, s, a, x} \int_{-T}^{0} \left[ H \left( \frac{d}{dt} \bar{s}(t) \right) \right] dt
$$

subject to $(s, a, x)_T$ is an FSP

$$
\bar{x}(-T) = 0, V(\bar{x}(0)) \geq 1. \tag{28}
$$

Note that the statements of the two propositions are very similar. The difference is that Proposition 3 considers the overflow event $||\bar{x}^B(0)|| \geq 1$, whereas Proposition 4 considers the overflow event $V(\bar{x}^B(0)) \geq 1$. The importance of Proposition 4 will become clear in the later sections. Specifically, it is needed in the proof of Proposition 8.

The proof of Proposition 4 uses the Freidlin-Wentzell Theory. It is fairly technical and is provided in the Appendix. We emphasize that Propositions 3 and 4 provide a lower bound on the decay-rate of the stationary queue-overflow probability under very general assumptions.

## 4 Combining Large Deviations with Lyapunov Functions:

**A Simpler Lower Bound on the Decay-Rate**

In the previous section, we have derived a lower bound (Propositions 3 and 4) on the decay-rate of the stationary queue-overflow probability for a wireless system under fairly general assumptions. The infimum on the right-hand-side of (27) and (28) is often referred to as the “minimum-cost-to-overflow,” and the fluid sample path (FSP) that attains the infimum (if such an FSP exists) is referred to as the “most-likely path to overflow.” Unfortunately, searching for the most-likely
path to overflow is a multi-dimensional calculus-of-variations problem, which is usually very difficult to solve. To view this difficulty in another way, suppose now we want to verify that $\theta$ lower-bounds the minimum-cost-to-overflow (or, equivalently, the probability of overflow is approximately upper bounded by $\exp(-B\theta)$ when $B$ is large). We then need to ensure that

$$\int_{-T}^{0} \left[ H \left( \frac{d}{dt} \bar{z}(t)||\bar{p} \right) + L \left( \frac{d}{dt} \bar{a}(t) \right) \right] dt \geq \theta$$

(29)

for all FSPs $(s, a, x)_{T}$ that go from $\bar{x}(-T) = 0$ at some past time $-T$ to $||\bar{x}(0)|| \geq 1$. For advanced wireless resource-allocation algorithms like those in Examples 1 and 2 of Section 2.5, the complexity of enumerating all such paths soon becomes prohibitive except for some restrictive cases [10–12].

In this section, we develop a new technique to address this difficulty. Our new technique combines the large-deviation lower bound in Proposition 4 with Lyapunov functions to derive another simpler lower-bound on the decay-rate of the queue-overflow probability. The reason that we seek help from a Lyapunov function approach is actually very simple and intuitive. Note that this difficulty of evaluating all FSPs is in fact not unique. A similar scenario also arises when we want to prove stability of a dynamic system. For example, recall that in the fluid limit approach [19], in order to show that the fluid limit model of a system is stable, we need to show that there exists a $T > 0$, such that for all fluid limits with $||\bar{x}(-T)|| = 1$, we must have $\bar{x}(0) = 0$. Again, it would have been very difficult if one attempts to evaluate all possible multi-dimensional fluid limits. The Lyapunov function approach is indeed developed to address this complexity issue. The basic idea of a Lyapunov function approach is to map each multi-dimensional path $\bar{x}(t)$ to a one-dimensional path $V(\bar{x}(t))$. Recall from part (d) of Assumption 1 that such a function maps $||\bar{x}(-T)|| = 1$ to $V(\bar{x}(-T)) \leq \bar{C}$. By establishing that $V(\bar{x}(t))$ has a negative drift, we can show that $V(\bar{x}(t))$ must go from $V(\bar{x}(-T)) \leq \bar{C}$ at time $-T$ to $V(\bar{x}(0)) = 0$ at time 0, which then implies that $\bar{x}(0) = 0$. In other words, the key idea of the Lyapunov function approach is this mapping from a multi-dimensional space to a one-dimensional space, which greatly reduces the complexity for proving stability.
Can we use a similar Lyapunov function approach to characterize the queue-overflow probability of wireless resource-allocation algorithms? Indeed, Lyapunov functions have been used to solve other calculus-of-variations problems in the control literature. We next demonstrate how such an approach can be used to derive an even simpler lower bound on the minimum-cost-to-overflow (i.e., the infimum given in Proposition 4). Recall that in Assumption 1, part (d), the Lyapunov function \( V(\cdot) \) is chosen such that \( ||\vec{x}|| \geq 1 \) implies \( V(\vec{x}) \geq 1 \). For any \( v \geq 0 \) and \( w \), define

\[
l_V(v, w) \triangleq \inf_{s, a, x} H(\vec{\phi}||\vec{p}) + L(\vec{f})
\]

subject to \( (s, a, x) \) is an FSP

such that for some \( t \)

\[
\frac{d}{dt} \vec{s}(t) = \vec{\phi}
\]

\[
\frac{d}{dt} \vec{a}(t) = \vec{f}
\]

\[
V(\vec{x}(t)) = v
\]

\[
\frac{d}{dt} V(\vec{x}(t)) = w.
\]

According to (21) (or (22), correspondingly), for any \( w > -\eta V(\vec{x}(t)) \) (or \( w > -\eta \), correspondingly), the trajectory with \( \frac{d}{dt} V(\vec{x}(t)) = w \) becomes a “rare” event. The function \( l_V(v, w) \) provides a bound on the local rate-function [18, p71] for \( V(\vec{x}(t)) \), i.e., it bounds how rarely that, given \( V(\vec{x}(t)) = v \) at some time \( t \), the trajectory \( V(\vec{x}(t)) \) will move in the direction \( \frac{d}{dt} V(t) = w \) immediately after \( t \). Note that the infimum in (30) is taken over all possible FSPs such that the corresponding trajectory \( V(\vec{x}(t)) \) passes through \( v \) with slope \( w \).

For any FSP \((s, a, x)_T\), since both the arrival rate and the service rate are bounded, the function \( \vec{x}(t) \) must be Lipschitz-continuous. Further, \( \vec{x}(t) \) must be bounded over any finite interval. Hence, due to Part (e) of Assumption 1, the function \( V(\vec{x}(t)) \) must also be Lipschitz-continuous over any finite interval, and thus it must be differentiable almost everywhere. Using the definition of
\( l_V(\cdot, \cdot) \), we then have the following inequality for any FSP\((s, a, x)_T\):

\[
\int_{-T}^{0} H \left( \frac{d}{dt} \vec{s}(t) || \vec{p} \right) + L \left( \frac{d}{dt} \vec{a}(t) \right) dt \geq \int_{-T}^{0} l_V(V(\vec{x}(t)), \frac{d}{dt} V(\vec{x}(t))) dt.
\]

Let

\[
\theta_0 = \inf_{T > 0} \int_{-T}^{0} l_V(V(t), \frac{d}{dt} V(t)) dt
\]

subject to \( V(t) \) is continuous and almost-everywhere differentiable,

\[
V(-T) = 0 \text{ and } V(0) \geq 1.
\]

We then obtain a lower bound for the calculus-of-variations problem in (28), as stated in the following proposition.

**Proposition 5** Assume that there exists a Lyapunov function \( V(\cdot) \) that satisfies Assumption 1 and Assumption 2. Then \( \theta_0 \) in (32) is a lower bound on the decay-rate of the queue-overflow probability. In other words,

\[
\limsup_{B \to \infty} \frac{1}{B} \log \mathbb{P}[||\vec{x}^B(0)|| \geq 1] \leq \limsup_{B \to \infty} \frac{1}{B} \log \mathbb{P}[V(\vec{x}^B(0)) \geq 1] \leq -\theta_0
\]

**Proof:**

First, by Assumption 1, the following is true

\[
\mathbb{P}[||\vec{x}^B(0)|| \geq 1] \leq \mathbb{P}[V(\vec{x}^B(0)) \geq 1].
\]

This proves the first inequality. To show the second inequality, by Proposition 4, we only need to show that, for all FSP\((s, a, x)_T\) that goes from \( \vec{x}(-T) = 0 \) to \( V(\vec{x}(0)) \geq 1 \), the following must hold,

\[
\int_{-T}^{0} H \left( \frac{d}{dt} \vec{s}(t) || \vec{p} \right) + L \left( \frac{d}{dt} \vec{a}(t) \right) dt \geq \theta_0.
\]
Using (31), it suffices to show that, for all such FSPs,

\[ \int_{-T}^{0} l_{V}(V(x(t)), \frac{d}{dt} V(x(t))) \geq \theta_{0}. \]  

(34)

Note that, for all FSP\((s, a, x)_{T}\) that goes from \(x(-T) = 0\) to \(V(x(0)) \geq 1\), inequality (34) must hold due to the definition of \(\theta_{0}\) in (32). The result of the proposition then follows. \(Q.E.D.\)

It is also easy to see that a sufficient condition for all fluid samples paths to satisfy the constraint (29) is to ensure that

\[ \int_{-T}^{0} l_{V}(V(t), \frac{d}{dt} V(t))dt \geq \theta \]  

(35)

for all one-dimensional paths \(V(t)\) that go from \(V(-T) = 0\) to \(V(0) = 1\). Again, we have successfully reduced the original multi-dimensional calculus-of-variations problem to a one-dimensional calculus-of-variations problem. The one-dimensional calculus-of-variations problem in (32) and (35) is usually much easier to solve (Fig. 1).

Remark: Lyapunov functions have been used in the control literature to solve other calculus-of-variations problems. Often, the key to success of such an approach is to find the right Lyapunov function. The unique feature of the scheduling problem studied in this paper is that the Lyapunov function for stability automatically becomes the suitable Lyapunov function for the calculus-of-variations problem. Note that for any scheduling algorithm that is provably stable, which usually means that there exists a Lyapunov function for stability, we may then apply the above techniques to characterize the delay performance. In other words, the difficulty level of the delay-characterization problem is reduced to that of a stability problem. Since (35) is a sufficient condition to (29), we can obtain an upper bound on the overflow probability, and correspondingly, if a constraint on the overflow probability is imposed, we obtain a lower bound on the effective capacity region. The hope of this approach is that, if the function \(V(\cdot)\) is appropriately chosen, we may recover a large fraction of, or even the entire effective capacity region.
4.1 Scale-Linear Lyapunov Functions

In this section, we consider the special case when the Lyapunov function is linear in scale as defined in the following assumption:

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

*Remark:* Assumption 3 holds for any Lyapunov function that is a norm. For example, it holds for the Lyapunov functions for Examples 1 and 2 ($V_1(\cdot)$ and $V_2(\cdot)$) mentioned in Section 2.5.

In this case, we can show that the solution to $\theta_0$ in (32) can be further simplified. This is possible because the function $l_V(v, w)$ turns out to be independent of $v$. We need the following simple Lemma.

**Lemma 6** If $(s, a, x)_T$ is an FSP, then for any given $\hat{t} \in [-T, 0]$ and for any $c \geq 0$, there exists
another FSP \( (s, a, x)_{cT} \) such that

\[
\frac{d}{dt}\overline{s}(t) \bigg|_{cT} = \frac{d}{dt}\overline{a}(t) \bigg|_{cT} \quad (36)
\]

\[
\frac{d}{dt}\overline{a}(t) \bigg|_{cT} = \frac{d}{dt}\overline{a}(t) \bigg|_{cT} \quad (37)
\]

\[
\tilde{x}(ct) = c\tilde{x}(t), \quad \frac{d}{dt}\tilde{x}(t) \bigg|_{cT} = \frac{d}{dt}\tilde{x}(t) \bigg|_{cT} \quad (38)
\]

**Proof:** Let \( \frac{1}{B}S^B(B(T+t)), \frac{1}{B}A^B(B(T+t)), \frac{1}{B}X^B(B(T+t)) \) be the sequence of processes that converge to the FSP\((s, a, x)_T\). Consider the new sequence \( \frac{1}{B}S^{Bc}(B(cT+t)), \frac{1}{B}A^{Bc}(B(cT+t)), \frac{1}{B}X^{Bc}(B(cT+t)) \) for \( t \in [-cT, 0] \). In other words, we are choosing a sub-sequence from the original sequence, and shift for a different amount of time. This new sequence can be rewritten as \( c\frac{1}{B}S^{Bc}(Bc(T+cT)), c\frac{1}{B}A^{Bc}(Bc(T+cT)), c\frac{1}{B}X^{Bc}(Bc(T+cT)) \). Taking the limit as \( B \to \infty \), we get the FSP\((s, a, x)_{cT}\) = \((c\overline{s}(t), c\overline{a}(t), c\tilde{x}(t))\). It is easy to verify that this satisfies the conditions in (36),(37) and (38).

\[Q.E.D.\]

Next we prove that under Assumption 3, \( l_V(v, w) \) is independent of \( v \).

**Proposition 7** When Assumption 3 holds, the function \( l_V(v, w) \) is independent of \( v \), i.e.,

\[ l_V(v, w) = l_V(cv, w) \]

for all \( c > 0 \).
**Proof:** Consider a fixed $c > 0$. According to definition (30),

$$l_V(cv, w) = \inf_{s, a, x} H(\vec{\phi} || \vec{\rho}) + L(\vec{f})$$

subject to $(s, a, x)$ is an FSP such that for some $t$

$$\frac{d}{dt} \vec{s}(t) = \vec{\phi}$$
$$\frac{d}{dt} \vec{a}(t) = \vec{f}$$
$$V(\vec{x}(t)) = cv$$
$$\frac{d}{dt} V(\vec{x}(t)) = w.$$

For any FSP$(s, a, x)_T$ and $t \in [-T, 0]$, such that $\frac{d}{dt}\vec{s}(t)|_i = \vec{\phi}$, $\frac{d}{dt}\vec{a}(t)|_i = \vec{f}$, $V(\vec{x}(t)) = v$, $\frac{d}{dt} V(\vec{x}(t))|_i = \vec{w}$, according to Lemma 6, there must exist another FSP$(s, a, x)_{cT}$ such that $\frac{d}{dt}\vec{s}(t)|_{ci} = \vec{\phi}$, $\frac{d}{dt}\vec{a}(t)|_{ci} = \vec{f}$, $\vec{x}(ct) = c\vec{x}(t)$, and $\frac{d}{dt}\vec{x}(t)|_{ci} = \frac{d}{dt}\vec{x}(t)|_i$. Using Assumption 3, we then have

$$V(\vec{x}(ct)) = cv.$$

Further,

$$\left.\frac{d}{dt} V(\vec{x}(t))\right|_{ci} = \lim_{\tau \to 0} \frac{V(\vec{x}(ct + \tau)) - V(\vec{x}(ct))}{\tau}$$

$$= \lim_{\tau \to 0} \left[ V\left(\vec{x}(ct) + \frac{d\vec{x}(t)}{dt}|_{ci} \tau\right) - V(\vec{x}(ct)) \right]$$

$$+ \frac{V(\vec{x}(ct + \tau)) - V\left(\vec{x}(ct) + \frac{d\vec{x}(t)}{dt}|_{ci} \tau\right)}{\tau}.$$
According to Assumption 3, the first term is equal to
\[
\lim_{\tau \to 0} V \left( \ddot{x}(c\hat{t}) + \frac{d\ddot{x}(t)}{dt} \right|_{c\hat{t}} \tau \right) - V(\ddot{x}(c\hat{t}))
\]
\[
= \lim_{\tau \to 0} \frac{V(\ddot{x}(i)) + \frac{d\ddot{x}(i)}{dt} \frac{\tau}{c}}{\tau} - V(\ddot{x}(i))
\]
\[
= \frac{d}{dt} V(\ddot{x}(t)) \bigg|_{\hat{t}}.
\]

According to Assumption 1, the second term satisfies,
\[
\lim_{\tau \to 0} \left| V \left( \ddot{x}(c\hat{t}) + \tau \right) - V \left( \ddot{x}(c\hat{t}) + \frac{d\ddot{x}(t)}{dt} \right|_{c\hat{t}} \tau \right) \right|
\]
\[
\leq \lim_{\tau \to 0} L_0 \left| \ddot{x}(c\hat{t} + \tau) - \ddot{x}(c\hat{t}) - \frac{d\ddot{x}(t)}{dt} \right|_{c\hat{t}} \tau \right|
\]
\[
= 0.
\]

Hence, we have,
\[
\frac{d}{dt} V(\ddot{x}(t)) \bigg|_{c\hat{t}} = \frac{d}{dt} V(\ddot{x}(t)) \bigg|_{\hat{t}} = w.
\]

This implies that the FSP(s,a,x)_{cT} satisfies the constraint in the definition of l_V(cv, w). Hence,
\[
l_V(cv, w) \leq l_V(v, w)
\]

A similar argument proves the opposite direction that l_V(cv, w) ≥ l_V(v, w). Since c > 0 is arbitrary, the result then follows. 

Q.E.D.
When the function $l_V(v, w)$ is independent of $v$, the trajectory $V(\cdot)$ that attains the infimum in (32) is in fact very easy to solve [18, p520], and the infimum is equal to $\inf_{w>0} \frac{l_V(1, w)}{w}$, i.e.,

$$
\theta_0 = \inf_{w>0, s, a, x} \frac{1}{w} \left[ H(\vec{\phi}||\vec{p}) + L(\vec{f}) \right]
$$

subject to 

$$(s, a, x) \text{ is an FSP such that for some } t$$

$$\frac{d}{dt} \bar{s}(t) = \vec{\phi}$$

$$\frac{d}{dt} \bar{a}(t) = \vec{f}$$

$$V(\bar{x}(t)) = 1$$

$$\frac{dV(\bar{x}(t))}{dt} = w.$$  

The value of $\theta_0$ has an intuitive interpretation. If we interpret $w$ as the rate of increase of the value of the Lyapunov function, then the objective function in (39) can be viewed as the minimum per-unit cost to increase the Lyapunov function, where the minimization is taken over all backlog levels $\bar{x}(t)$, channel states $\bar{s}(t)$, and arrivals $\bar{a}(t)$. In order to overflow, we must lift the value of $V(\bar{x}(t))$ from zero to one. Hence, $\theta_0$ becomes a lower bound on the minimum cost to overflow. According to Proposition 5, $\theta_0$ then corresponds to a lower bound on the decay rate of the overflow probability.

## 5 A Condition For The Minimum-Cost-To-Overflow To Be Exact

In the previous section, we have shown that $\theta_0$ is a lower bound on the decay rate of the queue overflow probability (see Proposition 5). In this section, we provide a condition when the value of $\theta_0$ becomes the exact decay-rate of the overflow probability.

Note that many scheduling policies are designed to minimize the drift of the respective Lyapunov function, as stated in the following assumption. For any $\vec{f}, \vec{\phi}$, let $\delta^k = f^k - \sum_{j=1}^S \phi_j \sum_{l=1}^L R_{il} e^k_{lj}$. 

36
Define

\[ \tilde{V}(\tau, \vec{e}|\vec{x}, \vec{\phi}, \vec{f}) \triangleq V([\vec{x} + \delta \tau]^+). \quad (40) \]

Then \( \frac{d}{d\tau} \tilde{V}(\tau, \vec{e}|\vec{x}, \vec{\phi}, \vec{f}) \) can be viewed as the drift of the Lyapunov function from \( \vec{x}(t) = \vec{x} \) if the service-rate vector is chosen as \( \vec{e} \), conditioned on the channel state being \( \vec{\phi} \) and \( \frac{d}{dt} \vec{a}(t) = \vec{f} \).

Recall that throughout this paper, we use right-derivatives unless otherwise stated.

**Assumption 4** For any FSP(\( s, a, x \)), the following holds for all \( t \):

\[ \frac{d}{dt} V(\vec{x}(t)) = \min_{e_j \in \text{Conv}(\xi_j)} \frac{\partial}{\partial \tau} \tilde{V}(\tau, e_j|\vec{x}(t), \vec{\phi}(t), \vec{f}(t)). \quad (41) \]

where \( \vec{\phi}(t) = \frac{d}{dt} \vec{s}(t), \vec{f}(t) = \frac{d}{dt} \vec{a}(t) \).

This assumption states that at any point of the FSP(\( s, a, x \)), the scheduling algorithm minimizes the drift of the Lyapunov function over all possible scheduling decisions.

Readers may refer to the example in Section 6 where it is shown that under a simplified version of the cellular model of Section 2.5.1, the QLB policy minimizes the drift of the Lyapunov function \( V(\vec{x}) = \max_l x_l \).

In addition, we assume the following for the Lyapunov function.

**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 \( \vec{x}_2 \geq 0 \).

Note that Assumptions 3 and 6 combined implies that the Lyapunov function \( V(\vec{x}) \) almost behaves as a norm except that it may not be defined for negative values of the variable \( \vec{x} \). We are ready for the following proposition.

**Proposition 8** Under Assumptions 1, 2, 3, 4, 5 and 6, the value of \( \theta_0 \) is the exact decay-rate of the probability of overflow according to the Lyapunov function metric, i.e.,

\[ \lim_{B \to \infty} \frac{1}{B} \log \mathbb{P}[V(\vec{x}^B(0)) \geq 1] = -\theta_0. \quad (42) \]
Further, the policy is optimal in minimizing this decay-rate. In other words, for any policy $\pi$ we must have

$$\liminf_{B \to \infty} \frac{1}{B} \log P^\pi[V(x^B(0)) \geq 1] \geq -\theta_0,$$

where $P^\pi$ denote the stationary distribution under the policy $\pi$.

**Remark:** We will soon show that $\theta_0$ is the same as $\tilde{\theta}_0$ which is defined in Equation (44).

The proof of Proposition 8 contains two parts. First, we show that the decay rate of the probability of overflow, in terms of the Lyapunov metric, is bounded from above for all scheduling policies. Then we show that under the assumptions on the Lyapunov function, this bound matches with the lower bound $\theta_0$.

Consider the following optimization problem:

$$\tilde{w}(\vec{\phi}, \vec{f}) = \min_{\vec{x}} V(\vec{x})$$

subject to

$$x_i^k = [f_i^k - \sum_{j=1}^S \phi_j \sum_{l=1}^L R_d e_{lj}]^+$$

$$[e_{lj}] \in \text{Conv}(E_j) \text{ for all } j.$$ 

The function $\tilde{w}(\vec{\phi}, \vec{f})$ can be viewed as the minimum rate of increase of the Lyapunov function if the channel state is $\vec{\phi}$ and the arrivals are $\vec{f}$. Let

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

(44)

We first show the following.

**Proposition 9** For any policy $\pi$ we must have

$$\liminf_{B \to \infty} \frac{1}{B} \log P^\pi[V(x^B(0)) \geq 1] \geq -\tilde{\theta}_0,$$

where $P^\pi$ denotes the stationary distribution under the policy $\pi$.

**Proof:** By the definition of $\tilde{\theta}_0$, for any $\delta \in (0, 1)$ there exists $\tilde{\phi}_0$ and $\tilde{f}_0$ such that

$$\frac{1}{\tilde{w}(\tilde{\phi}_0, \tilde{f}_0)} \left[ H(\tilde{\phi}_0||\vec{p}) + L(\tilde{f}_0) \right] \leq \tilde{\theta}_0 + \delta.$$
Further, it is easy to show that the function $\tilde{w}(\cdot, \cdot)$ is continuous with respect to $\tilde{\phi}$ and $\tilde{f}$. Hence, there exists $\epsilon$ such that for any $|\tilde{\phi} - \tilde{\phi}_0| \leq \epsilon$ and $|\tilde{f} - \tilde{f}_0| \leq \epsilon$, the following holds

$$\tilde{w}(\tilde{\phi}, \tilde{f}) \geq \tilde{w}(\tilde{\phi}_0, \tilde{f}_0)(1 - \delta). \quad (46)$$

Let $\gamma > 0$ be a small number and let $T = \frac{1 + \gamma}{\tilde{w}(\tilde{\phi}_0, \tilde{f}_0)(1 - \delta)}$. Define a channel-state process $\tilde{s}_0(\cdot)$ and an arrival process $\tilde{a}_0(\cdot)$ on the interval $[-T, 0]$ as follows:

$$\tilde{s}_0(t) = (t + T)\tilde{\phi}_0 \quad \text{and} \quad \tilde{a}_0(t) = (t + T)\tilde{f}_0.$$  

Let $B_T(\tilde{s}_0(\cdot))$ denote an $\epsilon$-ball around $\tilde{s}_0(\cdot)$, i.e., it contains all $\tilde{s}(\cdot)$ such that $\tilde{s}(-T) = 0$ and

$$||\tilde{s}(t) - \tilde{s}_0(t)||_\infty < \epsilon.$$  

Similarly, define an $\epsilon$-ball $B_T(\tilde{a}_0(\cdot))$ around $\tilde{a}_0(\cdot)$. We will now show that $V(\tilde{x}^B(0)) \geq 1$ if $\tilde{s}^B(\cdot) \in B_T(\tilde{s}_0(\cdot))$ and $\tilde{a}^B(\cdot) \in B_T(\tilde{a}_0(\cdot))$.

From the mapping in (14), we have,

$$x_i^{k,B}(0) - x_i^{k,B}(-T + \frac{1}{B}) = a_i^{k,B}(-\frac{1}{B}) - \sum_{j=1}^{s} \sum_{l=1}^{L} R_{il} A_{ij}^k$$

where we have used $A_{ij}^k$ to denote

$$E_l^k(j, B\tilde{x}^B(-\frac{1}{B})) \left[ s_j^B(-\frac{1}{B}) - s_j^B(-\frac{2}{B}) \right] + \ldots + E_l^k(j, B\tilde{x}^B(-T + \frac{1}{B})) \left[ s_j^B(-T + \frac{1}{B}) - s_j^B(-T) \right].$$

We make the following observations to simplify (47).

$$x_i^{k,B}(-T + \frac{1}{B}) - x_i^{k,B}(-T) = O\left(\frac{1}{B}\right) \quad a_i^{k,B}(-\frac{1}{B}) - a_i^{k,B}(0) = O\left(\frac{1}{B}\right) \quad s_j^B(-\frac{1}{B}) - s_j^B(0) = O\left(\frac{1}{B}\right).$$
Further, we can write $A^k_{ij} = \left[ s^B_j \left( -\frac{1}{B} \right) - s^B_j (-T) \right] e^k_{ij}$ where $e^k_{ij} \in \text{Conv}(E_j)$. This follows because $E^k_i(j, B\bar{x}^B(.)$ is in the set $E_j$ and the terms $\frac{s^B_j \left( -\frac{1}{B} \right) - s^B_j (-\frac{2}{B})}{s^B_j \left( -\frac{1}{B} \right) - s^B_j (-T)} \ldots \frac{s^B_j \left( -T + \frac{1}{B} \right) - s^B_j (-T)}{s^B_j \left( -\frac{1}{B} \right) - s^B_j (-T)}$ can be thought of as weights that sum to 1. Since $s^B(-T) = 0$, we have $A^k_{ij} = s^B_j \left( -\frac{1}{B} \right) e^k_{ij}$. (47) simplifies to

$x^{k,B}_i(0) - x^{k,B}_i(-T) = a^{k,B}_i(0) - \sum_{j=1}^{L} \sum_{l=1}^{R} R_{il} s^B_j(0) e^k_{ij}$

$+ O\left( \frac{1}{B} \right)$.

Since $x^{k,B}_i(-T) \geq 0$ and $x^{k,B}_i(0) \geq 0$, we can show that

$x^{k,B}_i(0) \geq \left[ a^{k,B}_i(0) - \sum_{j=1}^{L} \sum_{l=1}^{R} R_{il} s^B_j(0) e^k_{ij} \right] + O\left( \frac{1}{B} \right)$,

Rewriting the equation using $\bar{\phi} \triangleq \frac{\bar{a}^B(0)}{T}$ and $\bar{f} \triangleq \frac{\bar{a}^B(0)}{T}$, we have

$x^{k,B}_i(0) + O\left( \frac{1}{B} \right) \geq T \left[ f^k_i - \sum_{j=1}^{S} \bar{\phi}_j \sum_{l=1}^{L} R_{il} e^k_{ij} \right] + O\left( \frac{1}{B} \right)$

where $e^k_{ij} \in \text{Conv}(E_j)$.

Using Assumption 5 and Assumption 6 on this inequality, we obtain

$V(\bar{x}^B(0)) + V\left( O\left( \frac{1}{B} \right) \right) \geq V(T \bar{x})$

where $x^k_i = [ f^k_i - \sum_{j=1}^{S} \bar{\phi}_j \sum_{l=1}^{L} R_{il} e^k_{ij} ]^+$

This provides a bound on $V(\bar{x}^B(0))$. However, this bound depends on the particular value of $e^k_{ij}$. To avoid this difficulty, we loosen the bound by allowing $e^k_{ij}$ to be any element of $\text{Conv}(E_j)$.

Therefore,

$V(\bar{x}^B(0)) + V\left( O\left( \frac{1}{B} \right) \right) \geq T \min V(\bar{x})$

subject to $x^k_i = [ f^k_i - \sum_{j=1}^{S} \bar{\phi}_j \sum_{l=1}^{L} R_{il} e^k_{ij} ]^+$

$[ e^k_{ij} ] \in \text{Conv}(E_j)$ for all $j$. 

40
This implies that
\[ V(\bar{x}^B(0)) + V(O(\frac{1}{B})) \geq T\bar{w}(\bar{\phi}, \bar{f}) \geq 1 + \gamma, \]
where in the last step we have used the definition of \( T \) and the fact that (46) holds.

Therefore, there exists \( B_\gamma \) such that for all \( B > B_\gamma \), if \( \bar{s}^B(\cdot) \in B_T(\bar{s}_0(\cdot)) \) and \( \bar{a}^B(\cdot) \in B_T(\bar{a}_0(\cdot)) \) then
\[ V(\bar{x}^B(0)) \geq 1. \]

Now, using the LDP for \( \bar{s}^B(\cdot) \) and \( \bar{a}^B(\cdot) \), we complete the proof as follows,

\[ \liminf_{B \to \infty} \frac{1}{B} \log P^\pi\left[ V(\bar{x}^B(0)) \geq 1 \right] \geq -\bar{\theta}_0. \]

(48)

Q.E.D.

We are now ready to prove Proposition 8.

**Proof of Proposition 8**: By Proposition 5,

\[ \limsup_{B \to \infty} \frac{1}{B} \log P[V(\bar{x}^B(0)) \geq 1] \leq -\theta_0, \]
where \( \theta_0 \) is given by (39). By Proposition 9, we have

\[
\liminf_{B \to \infty} \frac{1}{B} \log \mathbb{P}^\pi [V(\bar{x}^B(0)) \geq 1] \geq -\tilde{\theta}_0, \tag{49}
\]

for any policy \( \pi \). Hence, to show Proposition 8, it only remains to show that \( \theta_0 \geq \tilde{\theta}_0 \). Consider any FSP\((s, a, x)\) that satisfies the constraint in the definition of \( \theta_0 \) (see Equation (39)). Define \( \tilde{\phi} \triangleq \frac{d}{dt} \tilde{x}(t), \tilde{f} \triangleq \frac{d}{dt} \tilde{a}(t) \) and \( w \triangleq \frac{dV(\bar{x}(t))}{dt} \). By Assumption 4, we must have

\[
w \leq \frac{\partial \tilde{V}(\tau, \tilde{e}|\tilde{x}(t), \tilde{\phi}, \tilde{f})}{\partial \tau} \tag{50}
\]

for any feasible \( \tilde{e} \).

Define \( \bar{\delta} = [\delta^k_i, i = 1, ..., N, k = 1, ..., K] \), where \( \delta^k_i = f^k_i - \sum_{j=1}^{S} \phi^i_j \sum_{l=1}^{L} R_{il} e_{kl} \). We have

\[
\tilde{V}(\tau, \tilde{e}|\tilde{x}(t), \tilde{\phi}, \tilde{f}) - \tilde{V}(0, \tilde{e}|\tilde{x}(t), \tilde{\phi}, \tilde{f}) = V([\bar{x}(t) + \bar{\delta}\tau]^+) - V(\bar{x}(t))
\]

Further, using Assumptions 3, 5 and 6, we have

\[
V([\bar{x}(t) + \bar{\delta}\tau]^+) - V(\bar{x}(t)) \leq V(\bar{x}(t) + [\bar{\delta}]^+\tau) - V(\bar{x}(t)) \leq V([\bar{\delta}]^+\tau) = \tau V([\bar{\delta}]^+).
\]

Hence, for any feasible \( \tilde{e} \), by (50), we must have

\[
w \leq V([\bar{\delta}]^+).
\]

Minimizing the right-hand-side as in the definition of \( \bar{w}(\tilde{\phi}, \tilde{f}) \), we have

\[
w \leq \bar{w}(\tilde{\phi}, \tilde{f}).
\]

This inequality and the definition of \( \theta_0 \) imply

\[
\theta_0 \geq \frac{1}{\bar{w}} \left[ H(\tilde{\phi}||\tilde{p}) + L(\tilde{f}) \right] \geq \frac{1}{\bar{w}(\tilde{\phi}, \tilde{f})} \left[ H(\tilde{\phi}||\tilde{p}) + L(\tilde{f}) \right] \geq \tilde{\theta}_0.
\]
The result of the proposition then follows. \( Q.E.D. \)

Recall the examples discussed in Section 2.5. The Lyapunov functions \( V_1(\vec{x}) = \sqrt{\sum_{l=1}^{L} x_l^2} \) and \( V_2(\vec{x}) = \sqrt{\sum_{i=1}^{N} \sum_{k=1}^{K} (x_k^i)^2} \) mentioned in Section 3.2 can be shown to satisfy Assumptions 1, 2, 3, 4, 5 and 6. Hence, as a result of Proposition 8, we conclude that the max-weight policy is optimal in minimizing the decay rate of \( P[V_1(\vec{x}^B(0)) \geq 1] \) and that the back-pressure algorithm is optimal in minimizing the decay rate of \( P[V_2(\vec{x}^B(0)) \geq 1] \). In the following section, we provide another example (see Proposition 10).

6 Example

Now, we use an example to illustrate the application of the results from the previous sections.

Consider the model of a base-station serving \( N \) users described in Section 2.5.1 (See Fig. 2). We recall here the assumptions made on the arrival process and the channel model from Section 2, and list some additional simplifying assumptions needed for our current purpose. \( A_l(t) \) denotes the number of packets generated by user \( l \). We assume that \( A_l(t) \) is \( i.i.d. \) across time and across users. Let \( \lambda_l = E[A_l(t)] \). Only one user can be scheduled for transmission at any time. We assume an ON-OFF channel between the base-station and the users. \( C(t) \) denotes the channel state at time slot \( t \) and \( \mathcal{S} \) denotes the set of all possible channel states. The probability that the channel state \( C(t) \) at time \( t \) is \( j \) is \( p_j \). For any subset \( \mathcal{A} \subset \{1,2,...N\} \), \( \mathcal{S}(\mathcal{A}) \) denotes the set of states \( j \) such that some user \( l \in \mathcal{A} \) is ON. Let \( F \) denote the bandwidth of the system. Hence, if a user’s channel is ON and it is scheduled for transmission, its service rate is \( F \). (Remark: The model is similar to the one in [11] although we do not assume identical arrival rates and \( i.i.d. \) channel state distribution for the users.) The throughput-optimal Tassiulas-Ephremides algorithm [1] in this case is the QLB (Queue-Length Based) algorithm as follows.

QLB-scheduling policy: At each time-slot \( t \), the base-station schedules the ON user with the
largest backlog. If there are multiple ON-users that all have the largest backlog, the basestation can schedule any one of these users.

We assume that the system is stable, which requires that there exists \( e_{lj} \geq 0 \), \( l = 1, \ldots, N \); \( j = 1, \ldots, S \) such that \( \sum_{l=1}^{N} e_{lj} = F \) for all \( j \), and for some \( \hat{\epsilon} > 0 \),

\[
\lambda_l(1 + \hat{\epsilon}) < \sum_{j \in S(l)} p_j e_{lj} F \text{ for all users } l = 1, \ldots, N
\]

(51)

The interpretation of \( e_{lj} \) is the long-term fraction of system bandwidth given to user \( l \) in channel state \( j \).

In this section, we would like to characterize the decay-rate of the tail probability of any user’s backlog exceeding a given threshold \( B \), i.e., the decay-rate of the probability

\[
P\left[ \max_{l=1,\ldots,N} X_l \geq B \right]
\]

(52)

when \( B \to \infty \).

Define \( \tilde{x}^B(t), \tilde{a}^B(t), \tilde{x}^B(t) \) as in (1), (4) and (12), and define the fluid sample path (FSP) accordingly. For any FSP(\( s, a, x \)), let \( \mathcal{I}_1(\tilde{x}(t)) = \{i| x_i(t) = \max_k x_k(t) \} \) be the set of users with the (identically) largest queue at time \( t \). Further, let \( \mathcal{I}_2(\tilde{x}(t), \tilde{x}'(t)) = \{ i \in \mathcal{I}_1(\tilde{x}(t)) | \frac{d}{dt} x_i(t) = \max_{k \in \mathcal{I}_1(\tilde{x}(t))} \frac{d}{dt} x_k(t) \} \). That is, \( \mathcal{I}_2(\tilde{x}(t), \tilde{x}'(t)) \) is the set of users that, among those users with the largest queue at time \( t \), also have the largest queue growth rate. In other words, these set of users will have the largest queue immediately after time \( t \). Then, immediately after time \( t \), as long as one user in \( \mathcal{I}_2(\tilde{x}(t), \tilde{x}'(t)) \) is ON, according to the QLB-policy this group of users collectively
must receive the full service rate \( F \). Therefore, we must have

\[
\sum_{i \in \mathcal{I}_2(\bar{x}(t), \bar{x}'(t))} \frac{d}{dt} x_i(t) = \sum_{i \in \mathcal{I}_2(\bar{x}(t), \bar{x}'(t))} \frac{d}{dt} a_i(t) - F \sum_{j \in \mathcal{S}(\mathcal{I}_2(\bar{x}(t), \bar{x}'(t)))} \frac{d}{dt} s_j(t). \tag{53}
\]

(Remark: Note that this is an example of Equation (15).)

Let \( V(\bar{x}) = \max_{l=1, \ldots, N} x_l \). Note that we have chosen the Lyapunov function to be the same as the norm for the overflow metric (52). We now show the following properties of \( V(\bar{x}) \).

**Proposition 10** The function \( V(\bar{x}) \) satisfies Assumptions 1, 2, 3, 4, 5 and 6.

**Proof:** Most of the conditions in the assumptions are easy to verify. Hence, we only provide proofs of Assumption 1 part (f), Assumption 2 and Assumption 4.

We first show Assumption 2 part (a). Consider an FSP(\( \mathbf{s}, \mathbf{a}, \mathbf{x} \)). Let \( \bar{\phi}(t) = \frac{d}{dt} \bar{s}(t) \) and let \( \bar{f}(t) = \frac{d}{dt} \bar{a}(t) \). According to the definition of \( \mathcal{I}_2(\bar{x}(t), \bar{x}'(t)) \), the Lyapunov drift for the QLB policy is given by

\[
\frac{d}{dt} V(\bar{x}(t)) = \frac{1}{|\mathcal{I}_2(\bar{x}(t), \bar{x}'(t))|} \sum_{l \in \mathcal{I}_2(\bar{x}(t), \bar{x}'(t))} \frac{d}{dt} x_i(t) - \frac{1}{|\mathcal{I}_2(\bar{x}(t), \bar{x}'(t))|} \left[ \sum_{l \in \mathcal{I}_2(\bar{x}(t), \bar{x}'(t))} f_i(t) \right. \\
\left. - F \sum_{j \in \mathcal{S}(\mathcal{I}_2(\bar{x}(t), \bar{x}'(t)))} \phi_j(t) \right].
\tag{54}
\]

Let \( \epsilon \leq \frac{\lambda_{\min}}{2(N+FS)} \). Assume \( || \frac{d}{dt} \bar{s}(t) - \bar{p} || < \epsilon \) and \( || \frac{d}{dt} \bar{a}(t) - \bar{\lambda} || < \epsilon \),

\[
\frac{d}{dt} V(\bar{x}(t)) \leq \frac{1}{|\mathcal{I}_2(\bar{x}(t), \bar{x}'(t))|} \left[ \sum_{l \in \mathcal{I}_2(\bar{x}(t), \bar{x}'(t))} \lambda_l \right. \\
\left. - F \sum_{j \in \mathcal{S}(\mathcal{I}_2(\bar{x}(t), \bar{x}'(t)))} p_j \right] + \epsilon(N + FS). \tag{56}
\]
By the stability condition (51), we have

\[
\sum_{l \in I_2(x(t), x'(t))} \lambda_l (1 + \hat{\epsilon}) \\
\leq \sum_{l \in I_2(x(t), x'(t))} \sum_{j \in S(l)} p_j \epsilon_{lj} F \\
\leq \sum_{j \in S(I_2(x(t), x'(t)))} p_j F.
\]

Therefore, Equation (56) becomes

\[
\frac{d}{dt} V(x(t)) \leq -\hat{\epsilon} \left| \sum_{l \in I_2(x(t), x'(t))} \lambda_l \right| + \epsilon (N + FS) \\
\leq -\hat{\epsilon} \min_{l=1, \ldots, N} \lambda_l.
\]

This shows Assumption 2 part (a). Note that Assumption 1 part (f) can be shown with a similar proof.

Now we show Assumption 2 part (b): By (55), and using the fact that the arrival process is bounded by \( M \), we have

\[
\frac{d}{dt} V(x(t)) \leq \frac{1}{|I_2(x(t), x'(t))|} \left| \sum_{l \in I_2(x(t), x'(t))} M \right| = M.
\]

To show Assumption 4, we will first bound the Lyapunov drift for any scheduling policy and show that the bound is in fact the drift for the QLB policy (55).

For any \( \bar{e} \), define \( \delta_l = f_l - \sum_{j=1}^S \phi_{j} e_{lj} \). By definition in Equation (40),

\[
\bar{V} = V([x + \bar{e} \tau] +) = \max_{l=1, \ldots, N} [x_l + \delta_l \tau] +,
\]

we must have,

\[
\frac{\partial}{\partial \tau} \bar{V}(\tau, \bar{e} | x(t), \phi(t), f(t)) \\
\geq \max_{l \in I_1(x(t))} (f_l(t) - \sum_{j=1}^s \phi_j(t) e_{lj})
\]

46
Further, since $\mathcal{I}_2(\vec{x}(t), \vec{x}′(t)) \subset \mathcal{I}_1(\vec{x}(t))$, we have

$$\frac{\partial}{\partial \tau} \tilde{V}(\tau, \vec{e}|\vec{x}(t), \vec{\phi}(t), \vec{f}(t)) \geq \max_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} (f_l(t) - \sum_{j=1}^{S} \phi_j(t) e_{lj})$$

$$\geq \frac{1}{|\mathcal{I}_2(\vec{x}(t), \vec{x}′(t))|} \left[ \sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} (f_l(t) - \sum_{j=1}^{S} \phi_j(t) e_{lj}) \right].$$

(57)

Now, for any $\vec{e}$, we must have

$$\sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} e_{lj} \leq F, \text{ if } j \in \mathcal{S}(\mathcal{I}_2(\vec{x}(t), \vec{x}′(t))),$$

and

$$\sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} e_{lj} = 0, \text{ if } j \notin \mathcal{S}(\mathcal{I}_2(\vec{x}(t), \vec{x}′(t))).$$

Hence

$$\sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} \sum_{j=1}^{S} \phi_j(t) e_{lj}$$

$$= \sum_{j=1}^{S} \phi_j(t) \sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} e_{lj}$$

$$\leq F \sum_{j \in \mathcal{S}(\mathcal{I}_2(\vec{x}(t), \vec{x}′(t)))} \phi_j(t).$$

Therefore, inequality (57) reduces to

$$\frac{\partial}{\partial \tau} \tilde{V}(\tau, \vec{e}|\vec{x}(t), \vec{\phi}(t), \vec{f}(t)) \geq \frac{1}{|\mathcal{I}_2(\vec{x}(t), \vec{x}′(t))|} \left[ \sum_{l \in \mathcal{I}_2(\vec{x}(t), \vec{x}′(t))} f_l(t) - F \sum_{j \in \mathcal{S}(\mathcal{I}_2(\vec{x}(t), \vec{x}′(t)))} \phi_j(t) \right]$$

where the right-hand-side is the drift of the QLB scheduler (55). The inequality (41) then follows.
The other assumptions are easily verified. \( Q.E.D. \)

By Proposition 8 and Proposition 10, we then have

\[
\lim_{B \to \infty} \frac{1}{B} \log P[V(x^B(0)) \geq 1] = -\theta_0.
\]

where \( \theta_0 \) is given by

\[
\theta_0 = \inf_{\{\bar{\phi}, \bar{f}, \tilde{w}(\bar{\phi}, \bar{f}) > 0\}} \frac{1}{\tilde{w}(\bar{\phi}, \bar{f})} \left[ H(\bar{\phi} || \bar{p}) + L(\bar{f}) \right], \quad (58)
\]

where

\[
\tilde{w}(\bar{\phi}, \bar{f}) = \min_{[e_{ij}]} \max_{I=1,\ldots,N} x_l
\]

subject to

\[
x_l = [f_l - \sum_{j=1}^{S} \phi_j e_{ij}]^+
\]

\[
[e_{ij}] \in \text{Conv}(E_j) \text{ for all } j.
\]

Unfortunately, (58) is not a convex program and thus not easy to solve. To derive a simpler characterization of \( \theta_0 \), we first introduce a decomposition. For any subset \( \mathcal{M} \subset \{1,2,\ldots,N\} \), define

\[
\theta_0(\mathcal{M}) = \inf_{w > 0, \bar{\phi}, \bar{f}} \frac{1}{w} \left[ H(\bar{\phi} || \bar{p}) + L(\bar{f}) \right]
\]

subject to

\[
w = \frac{1}{|\mathcal{M}|} \left[ \sum_{l \in \mathcal{M}} f_l - F \sum_{j \in S(\mathcal{M})} \phi_j \right]
\]

We then have the following results.

Lemma 11

\[
\theta_0 = \min_{\mathcal{M} \subset \{1,2,\ldots,N\}} \theta_0(\mathcal{M})
\]

**Proof:** Fix a set \( \mathcal{M} \). For any \( \bar{\phi}, \bar{f}, \) and \( w > 0 \) that satisfy the constraints in (59), we automatically have

\[
\frac{1}{|\mathcal{M}|} \left[ \sum_{l \in \mathcal{M}} \left( f_l - F \sum_{j \in S(\mathcal{M})} \phi_j \right) \right] = w.
\]
Now, for any feasible $\vec{e}$, we must have
\[
\max_{l=1,\ldots,N} \left[ f_l - \sum_{j=1}^S \phi_j e_{lj} \right]^+ \geq \frac{1}{|\mathcal{M}|} \left[ \sum_{l \in \mathcal{M}} \left[ f_l - \sum_{j=1}^S \phi_j e_{lj} \right]^+ \right] \geq \frac{1}{|\mathcal{M}|} \left[ \sum_{l \in \mathcal{M}} f_l - F \sum_{j \in \mathcal{S}(\mathcal{M})} \phi_j \right] = w.
\]

Hence, $\bar{w}(\vec{\phi}, \vec{f}) \geq w$, and
\[
\theta_0 \leq \frac{1}{w} \left[ H(\vec{\phi}||\vec{p}) + L(\vec{f}) \right].
\]

Since this is true for all $\vec{\phi}, \vec{f}$ and $w > 0$, we then have $\theta_0 \leq \theta_0(\mathcal{M})$. To show the other direction, it is sufficient to show that for every $\vec{\phi}$ and $\vec{f}$, there is some $\mathcal{M}$ for which $w = \bar{w}(\vec{\phi}, \vec{f})$ satisfies the constraint (59).

Let $w = \bar{w}(\vec{\phi}, \vec{f})$. By the definition of $\bar{w}(\cdot, \cdot)$, there must exist $\vec{e} \in \text{Conv}(\mathcal{E}_j)$ such that
\[
\max_{l=1,\ldots,N} \left[ f_l - \sum_{j=1}^S \phi_j e_{lj} \right]^+ = w.
\]

Let $\mathcal{M}$ be the set of $l$ such that $f_l - \sum_{j=1}^S \phi_j e_{lj} = w$. Note that we must have $\sum_{j \in \mathcal{S}(\mathcal{M})} e_{lj} = F$ for any state $j \in \mathcal{S}(\mathcal{M})$ because otherwise we should be able to further reduce $\max_{l=1,\ldots,N} \left[ f_l - \sum_{j=1}^S \phi_j e_{lj} \right]^+$. Hence, we have
\[
\sum_{l \in \mathcal{M}} f_l - F \sum_{j \in \mathcal{S}(\mathcal{M})} \phi_j = w|\mathcal{M}|
\]
This equation implies that $\vec{\phi}, \vec{f}$ and $w$ satisfies the constraint in (59). Hence, we must have $\theta_0 \geq \theta_0(\mathcal{M})$. The result of the lemma then follows.

Next, consider the case when the arrivals are at a constant rate $\lambda_l$. In other words, $L_l(f_l) = +\infty$ except when $f_l = \lambda_l$. In this case, we can use Lemma 11 to obtain the following characterization of $\theta_0$. For any subset $\mathcal{C}$ of the possible channel states, let $p(\mathcal{C}) = \sum_{j \in \mathcal{C}} p_j$.  

49
Proposition 12  When the arrivals are at a constant rate $\lambda_i$,
\[
\theta_0 = \min_{M \subseteq \{1, \ldots, N\}} \inf_{0 \leq u \leq \sum_{i \in M} \lambda_i} \frac{|M| D_M(u||p)}{\left(\sum_{i \in M} \lambda_i\right) - u F}
\]
where
\[
D_M(u||p) = u \log \frac{u}{p(S(M))} + (1 - u) \log \frac{1 - u}{1 - p(S(M))}.
\]

Proof: According to Lemma 11, it suffices to show that
\[
\theta_0(M) = \inf_{0 \leq u \leq \sum_{i \in M} \lambda_i} \frac{|M| D_M(u||p)}{\left(\sum_{i \in M} \lambda_i\right) - u F}.
\]
To see this, note that for any fixed $w$, $\theta_0(M)$ defined in (59) corresponds to a convex program. Associate a Lagrange multiplier $\eta$ for the constraint of (59), and a Lagrange multiplier $\gamma$ for the constraint $\sum_{j=1}^{S} \phi_j = 1$. Ignoring the term $L(\vec{f})$ and letting $f_i = \lambda_i$, we can then construct the Lagrangian as
\[
\mathcal{L}(\vec{\phi}, \vec{f}, w, \eta, \gamma) = \sum_{j=1}^{S} \phi_j \log \frac{\phi_j}{p_j} - \eta \left[ \sum_{i \in M} \lambda_i \right] - F \sum_{j \in S(M)} \phi_j - w |M| + \gamma \left[ \sum_{j=1}^{S} \phi_j - 1 \right]
\]
\[
= \left[ \sum_{j=1}^{S} \phi_j \log \frac{\phi_j}{p_j} + \eta F \sum_{j \in S(M)} \phi_j + \gamma \phi_j \right]
- \eta \left[ \sum_{i \in M} \lambda_i - w |M| \right].
\]
It is easy to verify that, in order to minimize the Lagrangian over all $\vec{\phi}$, we must have
\[
\phi_j = p_j \exp[-(1 + \eta F + \gamma)] \text{ if } j \in S(M),
\]
\[
\phi_j = p_j \exp[-(1 + \gamma)] \text{ if } j \not\in S(M).
\]
The optimal $\eta$ and $\gamma$ are such that the constraint of (59) and $\sum_{j=1}^{S} \phi_j = 1$ are both satisfied. Hence, we must have

$$\sum_{j \in S(M)} \phi_j = \exp[-(1 + \eta F + \gamma)] \sum_{j \in S(M)} p_j = \frac{\sum_{l \in M} \lambda_l - w|M|}{F}.$$ 

Let $u = \frac{\sum_{l \in M} \lambda_l - w|M|}{F}$. We then have,

$$\exp[-(1 + \eta F + \gamma)] = \frac{u}{p(S(M))}.$$

and

$$\phi_j = p_j \frac{u}{p(S(M))} \text{ if } j \in S(M).$$

Similarly, we must have

$$\sum_{j \notin S(M)} \phi_j = \exp[-(1 + \gamma)] \sum_{j \notin S(M)} p_j = 1 - \frac{\sum_{l \in M} \lambda_l - w|M|}{F} = 1 - u.$$ 

Hence,

$$\phi_j = p_j \frac{(1 - u)}{(1 - p(S(M)))} \text{ if } j \notin S(M).$$

We thus have, for any fixed $w$, the minimum value of the objective function in (59) is equal to

$$\frac{1}{w} \sum_{j=1}^{S} \phi_j \log \frac{\phi_j}{p_j}$$

$$= \frac{1}{w} \left[ u \log \frac{u}{p(S(M))} + (1 - u) \log \frac{(1 - u)}{(1 - p(S(M)))} \right]$$

$$= \frac{\sum_{l \in M} \lambda_l - uF}{|M|} D_M(u||p),$$

where in the last step we have used the definition of $D_M(u||p)$ and the assignment that $u = \frac{\sum_{l \in M} \lambda_l - wM}{F}$. Now, taking a minimum over $w$, or equivalently over $u$, the result then follows.

Q.E.D.

**Remark:** Readers can verify that when the channel states are i.i.d and when the arrivals from all users are at a constant rate $\lambda$, Proposition 12 reduces to Theorem 5 in [11].
6.1 Effective Capacity

When the arrivals are time-varying, it is no longer possible to derive a simpler characterization of $\theta_0$ as in Proposition 12. Instead, we can concentrate on the effective capacity region. Note that for any $\theta > 0$, if we would like to ensure that

$$\lim_{B \to \infty} \frac{1}{B} \log P[\max_i X_i(0) \geq B] \leq -\theta$$

it is equivalent to require that $\theta \leq \theta_0$. We then have the following result. Let $M_i(\theta) = \log E[\exp(\theta A_i(t))]$, and let $c_\theta(\theta) = \frac{M_i(\theta)}{\theta}$. The quantity $c_\theta(\theta)$ is typically referred to as the effective bandwidth of the system.

**Proposition 13** $\theta \leq \theta_0$ is equivalent to the following condition:

$$\sum_{i \in M} \frac{|M|}{\theta} M_i\left(\frac{\theta}{|M|}\right) \leq -\frac{|M|}{\theta} \log[p(S(M)) \exp(-\frac{\theta F}{M})] + 1 - p(S(M)).$$

for all $M \subset \{1, 2, ..., N\}$.

**Remark:** Given $\theta$, Proposition 13 provides a necessary and sufficient condition on the arrival process $A_i(t)$ for it to belong to the effective capacity region.

**Proof:** According to Lemma 11, we only need to show that (60) is equivalent to

$$\theta \leq \inf_{w > 0} \frac{1}{w} \left[H(\phi||\bar{p}) + L(\bar{f})\right]$$

(61)
where \( w = \frac{\sum_{l \in M} f_l - F \sum_{j \in S(M)} \phi_j}{|M|} \). To see this, note that

Inequality (61)

\[
\Leftrightarrow \theta w \leq H(\bar{\phi}\|\bar{p}) + L(\bar{f}) \quad \text{for all } \bar{\phi}, \bar{f} \text{ and}
\]

\[
w = \frac{\sum_{l \in M} f_l - F \sum_{j \in S(M)} \phi_j}{|M|}
\]

\[
\Leftrightarrow H(\bar{\phi}\|\bar{p}) + L(\bar{f}) \geq \frac{\theta}{|M|} \left[ \sum_{l \in M} f_l - F \sum_{j \in S(M)} \phi_j \right]
\]

for all \( \bar{\phi}, \bar{f} \)

\[
\Leftrightarrow \frac{\theta}{|M|} \sum_{l \in M} f_l - L(\bar{f}) \leq \frac{\theta F}{|M|} \sum_{j \in S(M)} \phi_j + H(\bar{\phi}\|\bar{p})
\]

for all \( \bar{\phi}, \bar{f} \)

\[
\Leftrightarrow \sup_{\bar{f}} \frac{\theta}{|M|} \sum_{l \in M} f_l - L(\bar{f}) \leq -\sup_{\bar{\phi}} \left[ -\frac{\theta F}{|M|} \sum_{j \in S(M)} \phi_j - H(\bar{\phi}\|\bar{p}) \right].
\]

Now, by definition \( L_l(f_l) = \sup_{\theta} \theta f_l - M_l(\theta) \). Using properties of Legendre transforms (Lemma 4.5.8 in [21, p.152]), we have

\[
\sup_{\bar{f}} \frac{\theta}{|M|} \sum_{l \in M} f_l - L(\bar{f}) = \sum_{l \in M} M_l(\frac{\theta}{|M|}).
\]

Similarly, by definition

\[
H(\bar{\phi}\|\bar{p}) = \sup_{\theta} \left\{ \sum_{j=1}^S \theta_j \phi_j - \log E[\exp(\sum_{j=1}^S \theta_j \Phi_j(t))] \right\}
\]

Hence, using the properties of Legendre transforms again, we have,

\[
\sup_{\bar{\phi}} \left[ -\frac{\theta F}{|M|} \sum_{j \in S(M)} \phi_j - H(\bar{\phi}\|\bar{p}) \right] = \log E[\exp(-\sum_{j \in S(M)} \frac{\theta F}{|M|} \Phi_j(t))] = \log[p(S(M)) \exp(-\frac{\theta F}{|M|}) + (1 - p(S(M))]\].
Hence, (61) is equivalent to
\[
\sum_{\ell \in \mathcal{M}} \frac{|\mathcal{M}|}{\theta} M_\ell \left( \frac{\theta}{|\mathcal{M}|} \right) \leq \frac{-|\mathcal{M}|}{\theta} \log[p(S(\mathcal{M})) \exp(-\frac{\theta F}{\mathcal{M}})] \\
+ (1 - p(S(\mathcal{M})))].
\]

Combining these conditions over all \( \mathcal{M} \), the result then follows. \( Q.E.D. \)

The quantity \( \frac{M_i(\theta)}{\theta} \) is typically refered to as the effective bandwidth of the system. The quantity on the right-hand-side of (60) can be interpreted as the effective capacity available to the users in \( \mathcal{M} \). Proposition 13 then carries the intuitive explanation that the total effective bandwidth of users in \( \mathcal{M} \) must be no greater than the effective capacity available to them. Note that both the effective bandwidth and effective capacity can be computed independent from each other for any given set \( \mathcal{M} \).

Remark: Readers can verify that, when the channel states are \( i.i.d. \) and when arrivals from all users are constant rate \( \lambda \), Proposition 13 reduces to Corollary 6 in [11].

7 Conclusions

In this paper we study the problem of characterizing the queue-overflow probability of complex wireless scheduling algorithms. We present a new technique for addressing the complexity issue of the calculus-of-variations problem involved in the sample-path large deviation approach. Our new technique combines sample-path large deviations with Lyapunov stability, which may develop into a powerful approach to study a large class of scheduling algorithms. We also show that when a scheduling algorithm minimizes the drift of the Lyapunov function, it is optimal in maximizing the asymptotic decay-rate of the probability that the Lyapunov function values exceed a threshold. We illustrate the potential of this approach through examples.
For ease of exposition, in this Appendix we will use the following scaling that does not involve a shift in time (readers can compare with (1), (4) and (12)). Specifically, for $t \geq 0$, let

$$s_j^B(t) = \frac{1}{B} \sum_{\tau=0}^{Bt} 1_{\{C(\tau)=j\}},$$
$$a_i^{k,B}(t) = \frac{1}{B} \sum_{\tau=0}^{Bt} A_i^k(\tau),$$
$$x_i^{k,B}(t) = \frac{1}{B} X_i^k(Bt),$$

if $t = 0, \frac{1}{B}, \frac{2}{B}, \ldots$, and by linear interpolation otherwise. Similarly, we define the fluid sample path $(s, a, x)_T$ as the limit of some sequence of $(s^B, a^B, x^B)$ over the interval $[0, T]$, as $B \to \infty$.

Pick any $0 < \rho < 1$. Pick another two positive constants $0 < \delta < \epsilon < \rho$. Let $P$ denote the stationary probability distribution of the system. Assume that each $x^B$ corresponds to a system that has reached the stationary regime at time 0. We are interested in the following quantity

$$\limsup_{B \to \infty} \frac{1}{B} \log P[V(\bar{x}^B(0)) \geq 1].$$

Let $\lceil t \rceil$ denote the smallest integer that is greater than or equal to $t$. For each $B$, consider the following sequence of stopping times defined on a sample path $x^B$. (Here we use the notations from [13].)

$$\beta_1^B \triangleq \left\lceil \inf_{t \geq 0} \left\{ t : V(\bar{x}^B(t)) \leq \delta \right\} B \right\rceil,$$

and

$$\eta_n^B \triangleq \frac{\left\lceil \inf_{t \geq \beta_n^B} \left\{ t : V(\bar{x}^B(t)) \geq \epsilon \right\} B \right\rceil}{B}, \quad n = 1, 2, \ldots,$$
$$\beta_n^B \triangleq \frac{\left\lceil \inf_{t \geq \eta_n^B} \left\{ t : V(\bar{x}^B(t)) \leq \delta \right\} B \right\rceil}{B}, \quad n = 2, 3, \ldots.$$

Fig. 3 provides an example of these stopping times.

Consider the Markov chain $\tilde{x}^B(n)$ obtained by sampling $\bar{x}^B(t)$ at the stopping times $\eta_n^B$, i.e.,

$$\tilde{x}^B(n) = \bar{x}^B(\eta_n^B), \quad n = 1, 2, \ldots, +\infty.$$ Since $x^B$ is stationary, there must also exist a stationary
distribution for the Markov chain \( \hat{x}^B(n) \). Denote this stationary distribution (of the Markov chain \( \hat{x}^B(n) \)) by \( \hat{P}^B \). Further, let \( \Theta^B \) denote the state space of the Markov chain \( \hat{x}^B(n) \). We can then express the stationary distribution of \( \bar{x}^B(0) \) as (see [22, Lemma 10.1]):

\[
P[\hat{V}(\bar{x}^B(0)) \geq 1] = \frac{\int_{\Theta^B} \hat{P}^B(d\bar{x}) E_{\bar{x}}(\int_0^{\eta_1^B} 1_{\{\hat{V}(\bar{x}^B(t)) \geq 1\}} dt)}{\int_{\Theta^B} P^B(d\bar{x}) E_{\bar{x}}(\eta_1^B)}
\]

where \( E_{\bar{x}}(\cdot) \) denotes the expectation conditioned on the event that \( \bar{x}^B(0) = \bar{x} \).

### A.1 Bounding the Denominator of (62)

Consider first the denominator in (62). From (5) and the boundedness of both the arrival-rate \( A_i^k(t) \) and the service-rate \( E_i^k(j, X(t)) \), there exists an upper bound \( M_1 \) such that \( ||X(t + 1) - X(t)|| \leq M_1 \) for all \( t \). From Assumption 1, we have that \( V(\bar{x}^B(\eta_1^B)) - V(\bar{x}^B(\beta_1^B)) \leq \mathcal{L}||\bar{x}^B(\eta_1^B) - \bar{x}^B(\beta_1^B)|| \). Denote \( M_0 \triangleq \mathcal{L}M_1 \). Hence, we must have,

\[
V(\bar{x}^B(\eta_1^B)) - V(\bar{x}^B(\beta_1^B)) \leq \mathcal{L}||\bar{x}^B(\eta_1^B) - \bar{x}^B(\beta_1^B)|| \\
\leq (\eta_1^B - \beta_1^B)M_0.
\]

Further, since \( \eta_1^B \) is the first time \( \hat{V}(\bar{x}^B(t)) \) exceeds \( \epsilon \), we must have

\[
V(\bar{x}^B(\eta_1^B)) \geq \epsilon - \frac{M_0}{B}.
\]
Similarly, we have \( V(\bar{x}^B(\beta_1)) \leq \delta + \frac{M_0}{B} \). Therefore,

\[
E_{\bar{x}}(\eta_1^B) \geq E(\eta_1^B - \beta_1^B) \\
\geq \frac{1}{M_0} \left( \epsilon - \delta - \frac{2M_0}{B} \right).
\]

Thus, there exists \( B_1 > 0 \) such that for all \( B \geq B_1 \), the denominator of (62) can be bounded from below by

\[
E_{\bar{x}}(\eta_1^B) \geq \frac{\epsilon - \delta}{2M_0}.
\] (63)

A.2 Bounding the Numerator of (62)

We next estimate the asymptotics of the numerator of (62). Recall that, by definition, each \( \eta_n^B \) is at most one \( \frac{1}{B} \) time-unit after \( V(\bar{x}^B(t)) \) just exceeds \( \epsilon \). Since \( \epsilon < \rho \), using the boundedness of the arrival rates and the service-rates, we can conclude that there exists \( B_2 > 0 \) (which depends on \( \rho - \epsilon \)), such that \( V(\bar{x}^B(\eta_n^B)) \leq \rho \) for all \( B \geq B_2 \).

We next define the following additional stopping time (see Fig. 3 for an example):

\[
\eta_{B,1}^B \triangleq \left\lceil \inf \left\{ t \geq 0 : V(\bar{x}^B(t)) \geq 1 \right\} B \right\rceil.
\]

Then, for any \( \bar{x} \in \Theta^B \), we must have,

\[
E_{\bar{x}} \left[ \int_0^{\eta_{B,1}^B} 1_{\{V(\bar{x}^B(t)) \geq 1\}} dt \right] \\
\leq E_{\bar{x}} \left[ 1_{\{\eta_{B,1}^B \leq \beta_1^B\}} (\beta_1^B - \eta_{B,1}^B) \right].
\]

The above inequality holds because: (a) if \( \beta_1^B \) occurs before \( \eta_{B,1}^B \), then both sides will be zero; and (b) if \( \beta_1^B \) occurs after \( \eta_{B,1}^B \), then the amount of time \( V(\bar{x}^B(t)) \geq 1 \) must be no greater than \( \beta_1^B - \eta_{B,1}^B \). Let \( P_{\bar{x}} \) denote the probability distribution conditioned on \( \bar{x}^B(0) = \bar{x} \). We then have

\[
E_{\bar{x}} \left[ \int_0^{\eta_{B,1}^B} 1_{\{V(\bar{x}^B(t)) \geq 1\}} dt \right] \\
\leq E_{\bar{x}} \left[ \beta_1^B - \eta_{B,1}^B | \eta_{B,1}^B \leq \beta_1^B \right] P_{\bar{x}}(\eta_{B,1}^B \leq \beta_1^B)
\]
Now, $\mathbb{E}_x [\beta^B_1 - \eta^{B,1} | \eta^{B,1} \leq \beta^B_1]$ is equal to $\mathbb{E}_x \{ \mathbb{E}_x [\beta^B_1 - \eta^{B,1} | \bar{x}^B(\eta^{B,1})], \eta^{B,1} \leq \beta^B_1 | \eta^{B,1} \leq \beta^B_1 \}$ and, due to the Markovian property of $x^B$, we must have $\mathbb{E}_x [\beta^B_1 - \eta^{B,1} | \bar{x}^B(\eta^{B,1})], \eta^{B,1} \leq \beta^B_1] = \mathbb{E}_{\bar{x}^B(\eta^{B,1})}(\beta^B_1)$. Hence,

$$
\mathbb{E}_x \left[ \int_0^{\eta^{B,1}} 1_{\{V(\bar{x}^B(t)) \geq 1\}} dt \right]
\leq \mathbb{E}_x \left[ \mathbb{E}_{\bar{x}^B(\eta^{B,1})}(\beta^B_1) | \eta^{B,1} \leq \beta^B_1 \right] \mathbb{P}_x(\eta^{B,1} \leq \beta^B_1).
$$

Let $C$ be a number that is slightly larger than 1. Using the boundedness of the arrival rates and the service-rates again, we can find $B_3 > 0$ such that for all $B \geq B_3$, $V(\bar{x}^B(\eta^{B,1})) \leq C$. Hence, we can bound the above quantity by

$$
\mathbb{E}_x \left[ \int_0^{\eta^{B,1}} 1_{\{V(\bar{x}^B(t)) \geq 1\}} dt \right]
\leq \left[ \sup_{\{\bar{y} : V(\bar{y}) \leq C\}} \mathbb{E}_\bar{y}(\beta^B_1) \right] \mathbb{P}_x(\eta^{B,1} \leq \beta^B_1).
$$

Let $T$ be a positive number (which will be chosen later). Recall that $V(\bar{x}) \leq \rho$ for all $\bar{x} \in \Theta^B$ when $B \geq B_2$. Hence, for any such $\bar{x} \in \Theta^B$, we have,

$$
\mathbb{E}_x \left[ \int_0^{\eta^{B,1}} 1_{\{V(\bar{x}^B(t)) \geq 1\}} dt \right]
\leq \left[ \sup_{\{\bar{y} : V(\bar{y}) \leq C\}} \mathbb{E}_\bar{y}(\beta^B_1) \right] \mathbb{P}_x(\eta^{B,1} \leq T) + \mathbb{P}_x(\beta^B_1 \geq T)
$$

$$
\leq \left[ \sup_{\{\bar{y} : V(\bar{y}) \leq C\}} \mathbb{E}_\bar{y}(\beta^B_1) \right] \left[ \sup_{\{\bar{x} : V(\bar{x}) \leq \rho\}} \mathbb{P}_x(\eta^{B,1} \leq T) + \sup_{\{\bar{x} : V(\bar{x}) \leq \rho\}} \mathbb{P}_x(\beta^B_1 \geq T) \right].
$$

(64)
Substituting (63) and (64) into (62), we then have, for all \( B \geq \max\{B_1, B_2, B_3\} \),

\[
P[V(\bar{x}^B(0)) \geq 1] \leq \frac{2M_0}{\epsilon - \delta} \left[ \sup_{\{\bar{y}: V(\bar{y}) \leq C\}} E_{\bar{y}}(\beta_1^B) \times \right.
\]
\[
\left. \left( \sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_x(\eta^{B,1} \leq T) + \sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_x(\beta_1^B \geq T) \right) \right].
\] (65)

We next study the asymptotics for each term in the above inequality.

**A.2.1 Bound for \( \sup_{\{\bar{y}: V(\bar{y}) \leq C\}} E_{\bar{y}}(\beta_1^B) \)**

We will show that \( \sup_{\{\bar{y}: V(\bar{y}) \leq C\}} E_{\bar{y}}(\beta_1^B) \) is bounded from above and hence does not affect the asymptotics of (65). Due to the continuity of the Lyapunov function and the assumption that \( V(\bar{x}) = 0 \) only if \( ||\bar{x}|| = 0 \), there exists a \( \gamma > 0 \) such that \( ||\bar{x}|| < \gamma \) implies \( V(\bar{x}) < \delta \). Further, we can find a \( K > 0 \) such that \( V(\bar{x}) < C \) implies \( ||\bar{x}|| < K \).

Consider the following additional stopping time

\[
\hat{\beta}_1^B \triangleq \left\lceil \inf \{t \geq 0 : ||\bar{x}^B(t)|| \leq \gamma \} \right\rceil B.
\]

It is easy to see that \( \sup_{\{\bar{y}: ||\bar{y}|| \leq K\}} E_{\bar{y}}(\hat{\beta}_1^B) \) is an upper bound on \( \sup_{\{\bar{y}: V(\bar{y}) \leq C\}} E_{\bar{y}}(\beta_1^B) \).

We now proceed to show that \( \sup_{\{\bar{y}: ||\bar{y}|| \leq K\}} E_{\bar{y}}(\hat{\beta}_1^B) \) is bounded from above. From Assumption 1, the fluid limit of the system satisfies either (21) or (22). For both cases, it follows that there exists a constant \( t_0 \) such that for all fluid limits \( \bar{x} \) with \( ||\bar{x}(0)|| \leq 1 \), we must have \( ||\bar{x}(t_0)|| = 0 \). This not only implies that the original system is stable (see [19, Theorem 4.2]), but also leads to the following limit:

\[
\lim_{||\bar{x}|| \to \infty} \frac{1}{||\bar{x}||} E \left[ \bar{X}(t_0||\bar{x}||) \bigg| \bar{X}(0) = \bar{x} \right] = 0.
\]

(See the proof of Theorem 4.2 in [19]). Then using the techniques in the proof of Theorem 3.1 in [19], there must exist numbers \( \bar{c} > 0, \kappa > 0, \bar{b} \geq 0 \), and a bounded set \( \mathcal{B} \triangleq \{\bar{X} : ||\bar{X}|| \leq \kappa\} \)
such that for all $\bar{x}$, conditioned on $\bar{X}(0) = \bar{x}$, the following holds,

$$
E[\tau_B(t_0)|\bar{X}(0) = \bar{x}] \leq \frac{t_0}{\epsilon} (||\bar{x}|| + \bar{b}),
$$

where $\tau_B(t_0) \triangleq \inf\{t \geq t_0 : \bar{X}(t) \in B\}$ is the first time after $t_0$ when $\bar{X}(t)$ returns to the set $B$.

Recall the transformation $\bar{x}^B(t) = \frac{1}{B} \bar{X}(Bt)$, $t = 0, \frac{1}{B}, \frac{2}{B}, \ldots$. For all $B \geq \frac{\epsilon}{\gamma}$, as long as $||\bar{X}(Bt)|| \leq \kappa$, it implies that $||\bar{x}(t)|| \leq \gamma$ and thus $\hat{\beta}_1^B \leq t$. Hence, for any $\bar{y}$ such that $||\bar{y}|| \leq K$ and for any $B \geq \frac{\epsilon}{\gamma}$, the following holds

$$
E_{\bar{y}}(\hat{\beta}_1^B) \leq \frac{1}{B} E[\tau_B(t_0)|\bar{X}(0) = B\bar{y}]
\leq \frac{t_0}{\epsilon B} (B||\bar{y}|| + \bar{b}).
$$

Let $B_5 = \max\{\frac{\epsilon}{\gamma}, \frac{\bar{b}}{K}\}$. Then, for all $B \geq B_5$, we have,

$$
\sup_{\{\bar{y}: V(\bar{y}) \leq C\}} E_{\bar{y}}(\hat{\beta}_1^B) \leq \sup_{\{\bar{y}: ||\bar{y}|| \leq K\}} E_{\bar{y}}(\hat{\beta}_1^B) \leq 2 \frac{Kt_0}{\epsilon}. \quad (66)
$$

### A.2.2 Asymptotics for $\sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_{\bar{x}}(\eta^{B,\dagger} \leq T)$

Let

$$
\Gamma_{\leq \rho} \triangleq \{x : V(x(0)) \leq \rho \text{ and } V(x(t)) \geq 1 \text{ for some } t \in (0, T]\}
$$

Then, by Proposition 2, we have

$$
\limsup_{B \to \infty} \frac{1}{B} \log \sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_{\bar{x}}(\eta^{B,\dagger} \leq T)
= \limsup_{B \to \infty} \frac{1}{B} \log \sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_{\bar{x}}(x^B \in \Gamma_{\leq \rho})
\leq - \inf_{\text{all FSP}(s,a,x): x \in \Gamma_{\leq \rho}} \int_0^T \left[H \left(\frac{d}{dt} \bar{s}(t)||\bar{p}\right) + L \left(\frac{d}{dt} \bar{a}(t)\right)\right] dt. \quad (67)
$$
A.2.3 Asymptotics for $\sup_{\{\bar{x}: V(\bar{x}) \leq \rho\}} P_{\bar{x}}[\beta_1^B \geq T]$ 

Let

$$\Upsilon_{\leq \rho} \triangleq \{x : V(\bar{x}(0)) \leq \rho \text{ and } V(\bar{x}(t)) > \delta \}
\quad \text{for all } t \in [0, T - 1]\}.$$ 

Then, by Proposition 2, we have

$$\limsup_{B \to \infty} \frac{1}{B} \log \sup_{\{\bar{x}, V(\bar{x}) \leq \rho\}} P_{\bar{x}}[\beta_1^B \geq T]$$

$$\leq \limsup_{B \to \infty} \frac{1}{B} \log \sup_{\{\bar{x}, V(\bar{x}) \leq \rho\}} P_{\bar{x}}[\bar{x} \in \Upsilon_{\leq \rho}]$$

$$\leq - \inf_{\text{all FSP}(s, a, x)_T : x \in \Upsilon_{\leq \rho}} \int_0^{T-1} \left[ H \left( \frac{d}{dt} \bar{s}(t) || \bar{p} \right)
+ L \left( \frac{d}{dt} \bar{a}(t) \right) \right] dt. \quad (68)$$

For any FSP $(s, a, x)_T$ such that $x \in \Upsilon_{\leq \rho}$, we have

$$\delta \leq V(\bar{x}(0)) + \int_0^{T-1} \frac{d}{dt} V(\bar{x}(t)) dt$$

$$\leq \rho + \int_0^{T-1} \frac{d}{dt} V(\bar{x}(t)) dt.$$ 

We now need to use Assumption 2. Since the two parts in Assumption 2 are equivalent, in the following we will assume the latter part holds\(^\dagger\). Let $\eta$ be defined as in Assumptions 1 and 2. Then, according to Assumption 2, there exists $\epsilon' > 0$ such that for all FSPs $(s, a, x)_T$, if at any time $t$ we have $||\frac{d}{dt} \bar{s}(t) - \bar{p}|| \leq \epsilon'$ and $||\frac{d}{dt} \bar{a}(t) - \bar{\lambda}|| \leq \epsilon'$, then the following holds

$$\frac{d}{dt} V(\bar{x}(t)) \leq -\frac{\eta}{2}.$$ 

Further, there exists $M_1 \geq 0$ such that if at any time $t$ we have $||\frac{d}{dt} \bar{s}(t) - \bar{p}|| \geq \epsilon'$ or $||\frac{d}{dt} \bar{a}(t) - \bar{\lambda}|| \geq \epsilon'$, then the following holds

$$\frac{d}{dt} V(\bar{x}(t)) \leq M_1.$$

\(^\dagger\)If the first part of Assumption 2 holds, the following proof can be easily modified by using the Lyapunov function $U(\bar{x}) = V^{1-\alpha}(\bar{x})$. 

61
Let $\mathcal{M}$ denote the set of $(\vec{\phi}, \vec{f})$ such that $||\vec{\phi} - \vec{p}|| \leq \epsilon'$ and $||\vec{f} - \vec{\lambda}|| \leq \epsilon'$. We then have, 

$$
\delta \leq \rho + \int_0^{T-1} \left[ -\frac{\eta}{2} 1_{\{(\frac{d}{dt}\vec{s}(t), \frac{d}{dt}\vec{a}(t)) \in \mathcal{M}\}} + M_1 1_{\{(\frac{d}{dt}\vec{s}(t), \frac{d}{dt}\vec{a}(t)) \notin \mathcal{M}\}} \right] dt.
$$

Hence, 

$$
\left( M_1 + \frac{\eta}{2} \right) \int_0^{T-1} 1_{\{(\frac{d}{dt}\vec{s}(t), \frac{d}{dt}\vec{a}(t)) \notin \mathcal{M}\}} dt 
\geq (T - 1) \frac{\eta}{2} + \delta - \rho. \quad (69)
$$

Let

$$
J_{\min} = \min_{\vec{\phi}, \vec{f}} J(\vec{\phi}, \vec{f})
$$

subject to $||\vec{\phi} - \vec{p}|| \geq \epsilon'$ or $||\vec{f} - \vec{\lambda}|| \geq \epsilon'$.

It is easy to see that $J_{\min}$ is positive. Thus, for any FSP $(s, a, x)_T$ such that $x \in \Upsilon_{\leq \rho}$, we have

$$
\int_0^T \left[ H \left( \frac{d}{dt} \vec{s}(t) || \vec{p} \right) + L \left( \frac{d}{dt} \vec{a}(t) \right) \right] dt
\geq \int_0^{T-1} J_{\min} 1_{\{(\frac{d}{dt}\vec{s}(t), \frac{d}{dt}\vec{a}(t)) \notin \mathcal{M}\}} dt
\geq \frac{(T - 1) \frac{\eta}{2} + \delta - \rho}{M_1 + \eta/2} J_{\min}.
$$

where the last inequality follows from (69). Substituting into (68), we then have

$$
\limsup_{B \to \infty} \frac{1}{B} \log \sup_{\{x : V(\vec{x}) \leq \rho\}} \mathbb{P}_x[\beta^B \geq T] \leq - \frac{(T - 1) \frac{\eta}{2} + \delta - \rho}{M_1 + \eta/2} J_{\min}. \quad (71)
$$

Clearly, for fixed $\delta$ and $\rho$, by choosing large $T$, we can make the right-hand-side arbitrarily small.
A.3 Completing the Proof of Proposition 4

We are now ready to prove the statement of Proposition 4. Pick any FSP \((s, a, x)_{T_0}\) such that \(x(0) = 0, V(x(T_0)) \geq 1\). Suppose the cost of this FSP is \(\tilde{J}\), i.e.

\[
\int_0^{T_0} \left[ H \left( \frac{d}{dt} \tilde{s}(t)||\tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) \right] dt = \tilde{J}.
\]

Clearly, for any \(T \geq T_0\), the right-hand-side of (67) must be no smaller than \(-\tilde{J}\). According to (71), for fixed \(\delta\) and \(\rho\) there must exist \(T_1 > T_0\) (which is independent of \(\rho\)), such that for all \(T \geq T_1\), the right-hand-side of (71) is smaller than \(-\tilde{J}\). Fix such a \(T \geq T_1\). Substituting (66), (67) and (71) into (65), and taking the appropriate limits, we then have,

\[
\limsup_{B \to \infty} \frac{1}{B} \log P[V(x^B(0)) \geq 1] \leq -\inf_{\text{all FSP}(s, a, x)_{T_0}: x \in \Gamma} \int_0^{T_0} H \left( \frac{d}{dt} \tilde{s}(t)||\tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) dt.
\]

Note that the above inequality holds for all \(\rho > 0\). Let \(J_\rho\) denote the infimum on the right-hand-side. As \(\rho \to 0\), let \(J^*\) denote the limit, i.e., \(J^* = \lim_{\rho \to 0} J_\rho\). We then have

\[
\limsup_{B \to \infty} \frac{1}{B} \log P[V(x^B(0)) \geq 1] \leq -J^*.
\]

Let

\[
J_0 = \inf_{\text{all FSP}(s, a, x)_{T_0}: x(0) = 0, V(x(T)) \geq 1} \int_0^{T_0} H \left( \frac{d}{dt} \tilde{s}(t)||\tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) dt.
\]

It only remains to show that \(J^* \geq J_0\). To see this, take a sequence \(\rho_n \to 0\). There must exist a sequence of FSPs \((s_n, a_n, x_n)_{T}\) such that \(V(x_n(0)) \leq \rho_n\) and \(V(x_n(T)) \geq 1\) for each \(n\), and

\[
\lim_{n \to \infty} \int_0^{T_0} \left[ H \left( \frac{d}{dt} \tilde{s}_n(t)||\tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}_n(t) \right) \right] dt = J^*.
\]
Take a further subsequence that converges uniformly over compact intervals. Without loss of generality, we can denote this subsequence also by \((s_n, a_n, x_n)_T\), and let \((s, a, x)_T\) be the corresponding limit. Then, using the lower-semicontinuity of the cost function, we must have

\[
\int_0^T H \left( \frac{d}{dt} \tilde{x}(t) || \tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) dt \leq J^*.
\]

Using a similar argument as in the proof of Proposition 2, we can show that \((s, a, x)_T\) is also an FSP, and it satisfies the condition that \(x(0) = 0\) and \(V(x(T)) \geq 1\). Hence, it belongs to the set of FSP in the constraint set in (72). We thus have \(J_0 \leq J^*\). In other words, we have shown that for all \(T \geq T_1\),

\[
\limsup_{B \to \infty} \frac{1}{B} \log \mathbb{P}[V(\bar{x}_B(0)) \geq 1] \\
\leq - \inf_{\text{all FSP } (s, a, x)_T, \ x(0)=0 \text{ and } V(x(T)) \geq 1} \int_0^T H \left( \frac{d}{dt} \tilde{x}(t) || \tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) dt.
\]

Note that the above inequality is for a fixed \(T \geq T_1\). Note that the infimum on the right-hand-side decreases as \(T\) increases. Hence, taking another infimum over all \(T > 0\), we must then have

\[
\limsup_{B \to \infty} \frac{1}{B} \log \mathbb{P}[V(\bar{x}_B(0)) \geq 1] \\
\leq - \inf_{\text{all FSP } (s, a, x)_T, T > 0, \ x(0)=0 \text{ and } V(x(T)) \geq 1} \int_0^T H \left( \frac{d}{dt} \tilde{x}(t) || \tilde{p} \right) + L \left( \frac{d}{dt} \tilde{a}(t) \right) dt.
\]

The result of the proposition then follows with a trivial change of the origin of time.

References


