Wireless Scheduling Design for Optimizing Both Service Regularity and Mean Delay in Heavy-Traffic Regimes

Bin Li, *Member, IEEE*, Ruogu Li, *Student Member, IEEE*, and Atilla Eryilmaz, *Member, IEEE*

***Abstract*—We consider the design of throughput-optimal scheduling policies in multihop wireless networks that also possess good mean delay performance and provide regular service for all links—critical metrics for real-time applications. To that end, we study a parametric class of maximum-weight-type scheduling policies, called Regular Service Guarantee (RSG) Algorithm, where each link weight consists of its own queue length and a counter that tracks the time since the last service, namely Time-Since-Last-Service (TSLS). The RSG Algorithm not only is throughput-optimal, but also achieves a tradeoff between the service regularity performance and the mean delay, i.e., the service regularity performance of the RSG Algorithm improves at the cost of increasing mean delay. This motivates us to investigate whether satisfactory service regularity and low mean-delay can be simultaneously achieved by the RSG Algorithm by carefully selecting its design parameter. To that end, we perform a novel Lyapunov-drift-based analysis of the steady-state behavior of the stochastic network. Our analysis reveals that the RSG Algorithm can minimize the total mean queue length to establish mean delay optimality under heavily loaded conditions as long as the design parameter weighting for the TSLS scales no faster than the order of $\frac{1}{\sqrt[5]{\epsilon}}$, where ϵ measures the closeness of the network load to the boundary of the capacity region. To the best of our knowledge, this is the first work that provides regular service to all links while also achieving heavy-traffic optimality in mean queue lengths.**

***Index Terms*—Heavy-traffic analysis, mean delay, service regularity, throughput, wireless scheduling.**

I. INTRODUCTION

REAL-TIME applications, such as voice over IP or live multimedia streaming, are becoming increasingly popular as smartphones proliferate in wireless networks. To support real-time applications, network algorithm design should not only efficiently manage the interference among simultaneous transmissions, but also meet the requirements of quality of service (QoS) including delay, packet delivery ratio, and jitter. Such QoS requirements, in turn, depend on the higher-order statistics of the arrival and service process, which poses significant challenges for efficient network algorithm design.

In recent years, there has been an increasing understanding on the algorithm design that targets various aspects of QoS, especially packet delivery ratio requirement (e.g., [8], [9], and [11]) and low end-to-end delay (e.g., [2], [6], [25], and [26]). However, these QoS metrics do not fully characterize the quality of experience (QoE) of users in real-time applications in wireless networks. For example, in the network where each individual user wants to watch its video that is delivered by the base station, each mobile user would like to receive the data from the base station regularly. Yet, both the time-varying nature of wireless channels and the scheduling policy significantly affect the regularity of the received data of each mobile user. The traditional scheduling policies aiming to maximize the system throughput (e.g., [10], [17], and [22]) or provide various fairness guarantees (e.g., [3], [16], [18], and [21]) at the base station side do not take users' experience into account and thus lead to the high irregularity of the received data of mobile users.

Our work is motivated by the recent advances made in [13] and [15] that provide a promising approach for managing this critical QoS metric. In particular, [13] provides a throughput-optimal algorithm that prioritizes service of links with the largest link weight in general network topologies, where each link weight is the weighted sum of its own queue length and a counter, namely the Time-Since-Last-Service (TSLS), that tracks the time since the last service. This algorithm improves service regularity as its design parameter γ weighting for the TSLS increases (see Section III for more details). Yet, increasing γ also has an averse effect on the mean delay performance, which is also vital for most applications.

With this motivation, this paper focuses on the tradeoff between the service regularity and the mean delay performance that this class of policies achieves. In particular, we are interested in identifying the range of values for γ in which the mean delay performance guarantees can be provided, while the regularity characteristics are preserved. To that end, we build on the recently developed approach of using Lyapunov drifts for the steady-state analysis of queueing networks [4]. *The main result emanating from this analysis is the scaling law of* γ *as*

Manuscript received May 05, 2014; revised December 23, 2014 and April 20, 2015; accepted May 08, 2015; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor D. Manjunath. Date of publication June 01, 2015; date of current version June 14, 2016. This work was supported by the NSF under Grants CAREER-CNS-0953515, CNS-WiFiUS-1456806, and CCSS-EARS-1444026; the QNRF under Grants NPRP 09-1168-2-455 and NPRP 7-923-2-344; and the DTRA under Grant HDTRA1-15-1-0003. An earlier version of this paper appeared in the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Bangalore, India, July 29–August 1, 2013.

B. Li is with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: lib@illinois.edu).

R. Li is with ASSIA, Inc., Redwood City, CA 94065 USA (e-mail: lirg03@gmail.com).

A. Eryilmaz is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: eryilmaz.2@osu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2015.2432119

the system gets more and more heavily loaded so that the algorithm is mean-delay-optimal among all feasible scheduling policies and provides the best service regularity among this class of policies. Specifically, we show that the heavy-traffic optimality is preserved as long as γ scales[1] as $O\left(\frac{1}{\sqrt[5]{\epsilon}}\right)$, where ϵ is the heavy-traffic parameter characterizing the closeness of the arrival rate vector to the boundary of the capacity region.

Our analysis relates to the vast literature on heavy-traffic analysis of queueing networks (for example, [1], [5], [19], [20], [23], and [24]), and in particular extends the Lyapunov drift-based approach in [4]. A critical step in most of these results is to establish a *state-space collapse* along a single dimension, and thus relate the multidimensional system operation to a *resource-pooled* single-dimensional system. Our construction also follows such line of argument in broad strokes. However, the new dynamics of the considered class of algorithms require new Lyapunov functions and techniques in establishing their heavy-traffic optimality.

Note on Notation: We use bold and script font of a variable to denote a vector and a set. Also, let $|\mathcal{A}|$ denote the cardinality of the set $\mathcal{A}$. We use $\text{Int}(\mathcal{A})$ to denote the set of interior points of the set $\mathcal{A}$. We use $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\mathbf{x} \cdot \mathbf{y}$ to denote the inner product and componentwise product of the vector $\mathbf{x}$ and $\mathbf{y}$, respectively. We use $\mathbf{x}^2$ and $\sqrt{\mathbf{x}}$ to denote the componentwise square and square root of the vector $\mathbf{x}$, respectively. We also use $\preceq, \succeq, \prec, \succ$ to denote componentwise comparison of two vectors, respectively. Let $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|$ denote the l_1 and l_2 norm of the vector $\mathbf{x}$, respectively.

II. System Model

We consider a wireless network represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links. A node represents a wireless transmitter or receiver, while a link represents a pair of transmitter and receiver that are within the transmission range of each other. We use $L \triangleq |\mathcal{L}|$ for convenience. We consider the *link-based conflict model*, where links conflicting with each other cannot be active at the same time. We call a set of links that can be active simultaneously as a *feasible schedule* and denote it as $\mathbf{S}[t] = (S_l[t])_{l \in \mathcal{L}}$, where $S_l[t] = 1$ if the link l is scheduled in time-slot t, and $S_l[t] = 0$ otherwise. Let $\mathcal{S}$ be the set of all feasible schedules.

We capture the channel fading over link l in time-slot t via a *non-negative-integer-valued* random variable $C_l[t]$, with strictly positive mean and $C_l[t] \leq C_{\max}, \forall l, t$, for some $C_{\max} < \infty$, which measures the maximum amount of service available in slot t, if scheduled. We assume that $\{\mathbf{C}[t] = (C_l[t])_{l=1}^{L}\}_{t \geq 0}$ is an independently and identically distributed (i.i.d.) sequence of random vectors with $\psi_{\mathbf{c}} \triangleq \Pr\{\mathbf{C}[t] = \mathbf{c}\}$. We use $\mathcal{C}$ to denote the set of all possible channel states. Note that $|\mathcal{C}|$ is finite. Let $\mathcal{S}^{(\mathbf{c})} \triangleq \{\mathbf{S} \cdot \mathbf{c} : \mathbf{S} \in \mathcal{S}\}$ denote the set of feasible rate vectors in the channel state $\mathbf{c} \in \mathcal{C}$. Then, the *capacity region* is defined as $\mathcal{R} \triangleq \sum_{\mathbf{c}} \psi_{\mathbf{c}} \cdot \text{Conv}\{\mathcal{S}^{(\mathbf{c})}\}$, where $\text{Conv}\{\mathcal{A}\}$ denotes the convex hull of the set $\mathcal{A}$.

[1]We say $a_n = O(b_n)$ if there exists a $c > 0$ such that $|a_n| \leq c|b_n|$ for two real-valued sequences $\{a_n\}$ and $\{b_n\}$.

We assume a per-link traffic model, where $A_l[t]$ denotes the number of packets arriving at link l in slot t that are independently distributed over links and i.i.d. over time with finite mean $\lambda_l > 0$, and $A_l[t] \leq A_{\max}, \forall l, t$, for some $A_{\max} < \infty$. Accordingly, a queue is maintained for each link l with $Q_l[t]$ denoting its queue length at the beginning of time-slot t. Let $U_l[t] = \max\{0, C_l[t]S_l[t] - Q_l[t] - A_l[t]\}$ be the unused service for queue l in slot t. Then, the evolution of queue l is described as follows:

$$Q_l[t+1] = Q_l[t] + A_l[t] - C_l[t]S_l[t] + U_l[t] \quad \forall l. \tag{1}$$

We say that the queue l is *strongly stable* if it satisfies

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Q_l[t]] < \infty.$$

We call an algorithm *throughput-optimal* if it makes all queues strongly stable for any arrival rate vector $\boldsymbol{\lambda} = (\lambda_l)_{l=1}^{L}$ that lies strictly within the capacity region.

Our goal is to design a *throughput-optimal* scheduling algorithm that also possesses the following desirable properties for satisfying the QoS requirements: 1) provides *regular services* in the sense that the *normalized second moment of the interservice times* of the links is small; and 2) achieves *low mean delay* in the sense that the total mean queue lengths is small, especially in the regime where the system is *heavily loaded*—when delay effects are most pronounced.

Next, we provide a *regular service scheduler* that possesses throughput optimality and regular service guarantees, and then investigate its mean-delay performance under the heavy-traffic regime.

III. Regular Service Scheduler

One of our goals is to provide regular services for each link, which is related to the second moment of the interservice times. To characterize the interservice time, we introduce a counter T_l for each link l, namely TSLS, to keep track of the time since link l was last served. In particular, each T_l increases by 1 in each time-slot when link l has zero transmission rate, either because it is not scheduled or because its channel is unavailable, i.e., $C_l[t] = 0$, and drops to 0 otherwise. More precisely, the evolution of T_l is described as follows:

$$T_l[t+1] = \begin{cases} 0, & \text{if } S_l[t]C_l[t] > 0 \\ T_l[t] + 1, & \text{if } S_l[t]C_l[t] = 0. \end{cases} \tag{2}$$

Thus, the TSLS records the link "age" since the last time it received service and is closely related to the interservice time. Indeed, in [13], we showed that the normalized second moment of the interservice times of each link is proportional to the mean value of its TSLS for any stabilizing policy. Thus, the TSLS has a direct impact on service regularity: The smaller the mean TSLS value, the more regular the service.

This connection motivates the following maximum-weight type algorithm that uses a combination of queue lengths and TSLS values as its weights.

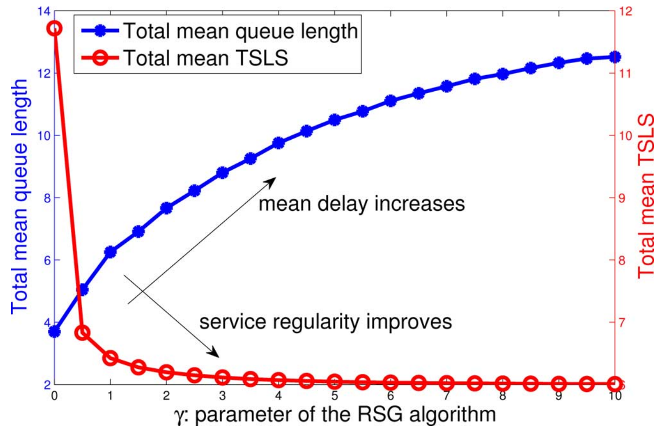


Fig. 1. Performance of the RSG Algorithm.

Regular Service Guarantee (RSG) Algorithm[2]:

In each time-slot t, select a schedule $\mathbf{S}^*[t]$ such that

$$\mathbf{S}^*[t] \in \arg\max_{\mathbf{S}\in\mathcal{S}} \sum_{l=1}^{L} (\alpha_l Q_l[t] + \gamma\beta_l T_l[t]) C_l[t] S_l \qquad (3)$$

where $\alpha_l > 0$, $\beta_l \geq 0$ and $\gamma \geq 0$ are design parameters.

The parameters $(\alpha_l)_{l=1}^L$ are weighting factors for the queue lengths, where a larger α_l will result in a smaller average queue length. The parameters $(\beta_l)_{l=1}^L$ weigh $T_l[t]$ differently for each link l, with γ being a common scaling factor for all links. Note that the RSG Algorithm coincides with the MWS Algorithm when $\gamma = 0$. Yet, the true significance of the RSG algorithm is observed for large γ since as γ increases, the RSG Algorithm prioritizes the schedule with the larger TSLS, hence providing more regular services for each link. In [13], we have showed that the RSG Algorithm not only achieves throughput optimality, but also provides regular service guarantees.

Yet, large values of γ may also deteriorate the mean delay performance. We demonstrate this tradeoff in a single-hop non-fading network with 4 links, where the number of packets arriving at each link follows a Bernoulli distribution with the arrival rate of 0.225. Let $\alpha_l = \beta_l = 1$ for each link l. Fig. 1 shows the mean delay and service regularity performance of the RSG Algorithm with varying γ.

Fig. 1 reveals that the improved service regularity of the RSG Algorithm with increasing γ comes at the cost of larger mean delays. We can show that the mean of the total TSLS value is minimized as γ goes to ∞ (see [13]). On the other hand, it is known (e.g., [4] and [20]) that the mean queue lengths are minimized under heavily loaded conditions (cf. Section IV for more detail) when $\gamma = 0$. In view of the tradeoff observed in the figure, our objective is to understand whether *both the regularity and the mean-delay optimality characteristics of the RSG Algorithm can be preserved, especially under heavily loaded conditions, by carefully selecting* γ.

In Section IV, we answer this question in the affirmative by explicitly characterizing how γ should scale with respect to the traffic load in order to achieve the heavy-traffic optimality while also optimizing the service regularity performance of the RSG Algorithm.

[2]For wireless networks with multihop traffic, where packets may traverse multiple links before their departure, we maintain queue length and TSLS counter for each flow at each link. Then, we propose a backpressure-type algorithm with the link weight consisting of queue length and TSLS. We can still show that the proposed backpressure-type algorithm achieves maximum throughput. However, the service regularity of each flow is hard to characterize in this setup, which is left for future research.

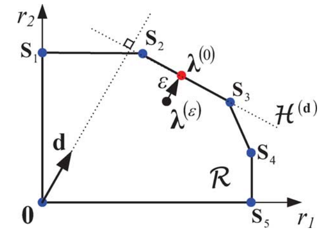


Fig. 2. Geometric structure of capacity region.

IV. Heavy-Traffic Optimality Result

In this section, we present our main result for the RSG Algorithm in terms of its mean delay optimality under the heavy-traffic limit, where the arrival rate vector approaches the boundary of the capacity region.

In the rest of the paper, we consider the RSG Algorithm[3] with $\alpha_l = 1, \forall l$. We first note that the capacity region $\mathcal{R}$ is a polyhedron due to the discreteness and finiteness of the service rate choices, and thus has a finite number of faces. We consider the exogenous arrival process $\{\mathbf{A}^{(\epsilon)}[t]\}_{t\geq 0}$ with mean rate vector $\boldsymbol{\lambda}^{(\epsilon)} \in \text{Int}(\mathcal{R})$, where ϵ measures the Euclidean distance of $\boldsymbol{\lambda}^{(\epsilon)}$ to the boundary of $\mathcal{R}$ (see Fig. 2).

In heavy-traffic analysis, we study the system performance as ϵ decreases to zero, i.e., as the arrival rate vector approaches $\boldsymbol{\lambda}^{(0)}$ belonging to the *relative interior* of a face, referred to as the *dominant hyperplane* $\mathcal{H}^{(\mathbf{d})}$. We denote $\mathcal{H}^{(\mathbf{d})} \triangleq \{\mathbf{r} \in \mathbb{R}^L : \langle \mathbf{r}, \mathbf{d}\rangle = b\}$, where $b \in \mathbb{R}$, and $\mathbf{d} \in \mathbb{R}^L$ is the normal vector of the hyperplane $\mathcal{H}^{(\mathbf{d})}$ satisfying $\|\mathbf{d}\| = 1$ and $\mathbf{d} \succeq \mathbf{0}$.

We are interested in understanding the steady-state queue length values with vanishing ϵ. To that end, we first provide a generic lower bound for all feasible schedulers by constructing

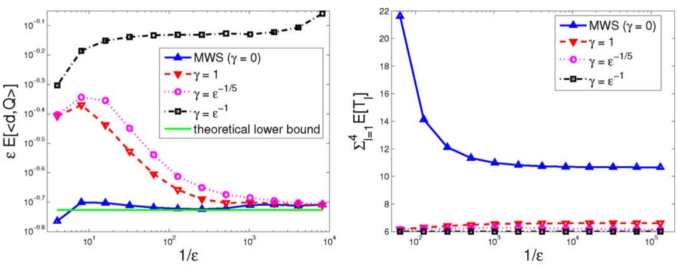
[3]Let $A_l'[t] \triangleq \sqrt{\alpha_l} A_l[t]$ and $C_l'[t] \triangleq \sqrt{\alpha_l} C_l[t]$, $\forall t \geq 0, \forall l$. Also let $\mathcal{S}'^{(\mathbf{c})} \triangleq \{\sqrt{\boldsymbol{\alpha}} \cdot \mathbf{c} \cdot \mathbf{S} : \mathbf{S} \in \mathcal{S}\}$ and $\mathcal{R}' \triangleq \sum_{\mathbf{c}} \psi_{\mathbf{c}} \cdot \text{CH}\{\mathcal{S}'^{(\mathbf{c})}\}$. Construct a hypothetical system with the arrival process $\{\mathbf{A}'[t] = (A_l'[t])_{l=1}^L\}_{t\geq 0}$ and the channel fading process $\{\mathbf{C}'[t] = (C_l'[t])_{l=1}^L\}_{t\geq 0}$ under the following RSG Algorithm in the hypothetical system:

$$\mathbf{S}'^*[t] \in \arg\max_{\mathbf{S}\in\mathcal{S}} \sum_{l=1}^{L} \left(Q_l'[t] + \gamma \frac{\beta_l}{\sqrt{\alpha_l}} T_l'[t] \right) C_l'[t] S_l$$

where T_l' is the TSLS counter for link l in the hypothetical system and evolves as follows:

$$T_l'[t+1] = \begin{cases} 0, & \text{if } S_l'[t]C_l'[t] > 0 \\ T_l'[t] + 1, & \text{if } S_l\text{'}[t]C_l\text{'}[t] = 0. \end{cases}$$

Let $\{\mathbf{Q}[t] = (Q_l[t])_{l=1}^L\}_{t\geq 0}$ and $\{\mathbf{Q}'[t] = (Q_l'[t])_{l=1}^L\}_{t\geq 0}$ be the queue length process under the RSG Algorithm in the original system and the RSG Algorithm with $\alpha_l = 1$ for each link l in the hypothetical system, respectively. Then, it is easy to show that $\{\sqrt{\boldsymbol{\alpha}} \cdot \mathbf{Q}[t]\}_{t\geq 0}$ is stochastically equal to $\{\mathbf{Q}'[t]\}_{t\geq 0}$. Thus, we can study the queue length behavior under the RSG Algorithm with $\alpha_l = 1$ for each link l in the hypothetical system.



(a) (b)

Fig. 3. Symmetric arrivals in a single-hop network. (a) Mean queue length. (b) Service regularity.

a hypothetical single-server queue with the arrival process $\langle \mathbf{d}, \mathbf{A}^{(\epsilon)}[t]\rangle$, and the i.i.d. service process $\Psi[t]$ with the probability distribution

$$\Pr\{\Psi[t] = b_{\mathbf{c}}\} = \psi_{\mathbf{c}}, \quad \text{for each channel state } \mathbf{c} \in \mathcal{C} \tag{4}$$

where $b_{\mathbf{c}} \triangleq \max_{\mathbf{s}\in\mathcal{S}^{(\mathbf{c})}} \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{s}\rangle$ is the maximum $\mathbf{d}$-weighted service rate achievable in channel state $\mathbf{c} \in \mathcal{C}$. By the construction of capacity region $\mathcal{R}$, we have $\mathbb{E}[\Psi[t]] = b$. Also, it is easy to show that the constructed single-server queue length $\{\Phi[t]\}_{t\geq 0}$ is stochastically smaller than the queue length process $\{\langle \mathbf{d}, \mathbf{Q}^{(\epsilon)}[t]\rangle\}_{t\geq 0}$ under any feasible scheduling policy. Hence, by using [4, Lemma 4], we have the following lower bound on the expected steady-state queue length under any feasible scheduling policy.

Proposition 1: Let $\overline{\mathbf{Q}}^{(\epsilon)}$ be a random vector with the same distribution as the steady-state distribution of the queue length processes under any feasible scheduling policy. Consider the heavy-traffic limit $\epsilon \downarrow 0$, suppose that the variance vector $\left(\boldsymbol{\sigma}^{(\epsilon)}\right)^2$ of the arrival process $\{\mathbf{A}^{(\epsilon)}[t]\}_{t\geq 0}$ converges to a constant vector $\boldsymbol{\sigma}^2$. Then

$$\lim_{\epsilon\downarrow 0} \epsilon\mathbb{E}\left[\langle \mathbf{d}, \overline{\mathbf{Q}}^{(\epsilon)}\rangle\right] \geq \frac{\zeta}{2} \tag{5}$$

where $\zeta \triangleq \langle \mathbf{d}^2, \boldsymbol{\sigma}^2\rangle + \mathrm{Var}(\Psi)$.

This fundamental lower bound of all feasible scheduling policies motivates the following definition of *heavy-traffic optimality* of a scheduler.

Definition 1 (Heavy-Traffic Optimality): A scheduler is called heavy-traffic optimal if its steady-state queue length vector $\overline{\mathbf{Q}}^{(\epsilon)}$ satisfies

$$\lim_{\epsilon\downarrow 0} \epsilon\mathbb{E}\left[\langle \mathbf{d}, \overline{\mathbf{Q}}^{(\epsilon)}\rangle\right] \leq \frac{\zeta}{2} \tag{6}$$

where ζ is defined in Proposition 1.

It is well known that the MWS Algorithm, which corresponds to the RSG Algorithm with $\gamma = 0$, is heavy-traffic optimal (e.g., [4] and [20]). This is shown by first establishing a *state-space collapse*, i.e., the deviations of queue lengths from the direction $\mathbf{d}$ are bounded, independent of heavy-traffic parameter ϵ. Since the lower bound of mean queue length is of order of $\frac{1}{\epsilon}$, the deviations from the direction $\mathbf{d}$ are negligible compared to the large queue length for a sufficiently small ϵ, and thus the queue lengths concentrate along the normal vector $\mathbf{d}$. Because of this, we also call the normal vector $\mathbf{d}$ the *line of attraction*.

However, as discussed in Section III, we are interested in large values of γ to provide satisfactory service regularity. Yet, it is unknown whether the RSG Algorithm can remain heavy-traffic optimal when γ is nonzero since larger values of γ lead to higher mean queue lengths (cf. Fig. 1). Also, the state-space collapse result is not applicable since the deviations from the line of attraction depend on γ. This raises the question of how $\gamma(\epsilon)$ should scale with ϵ in order to achieve heavy-traffic optimality while allowing $\gamma(\epsilon)$ to take large values (providing more regular services). We answer this interesting and challenging question by providing the following main result, proved in Section VI.

Proposition 2: Let $\overline{\mathbf{Q}}^{(\epsilon)}$ be a random vector with the same distribution as the steady-state distribution of the queue length processes under the RSG Algorithm. Consider the heavy-traffic limit $\epsilon \downarrow 0$, and suppose that the variance vector $\left(\boldsymbol{\sigma}^{(\epsilon)}\right)^2$ of the arrival process $\{\mathbf{A}^{(\epsilon)}[t]\}_{t\geq 0}$ converges to a constant vector $\boldsymbol{\sigma}^2$. Suppose the channel fading satisfies the mild assumption[4] $\Pr\{C_l[t] = 0\} > 0$, for all $l \in \mathcal{L}$. Then

$$\epsilon\mathbb{E}\left[\langle \mathbf{d}, \overline{\mathbf{Q}}^{(\epsilon)}\rangle\right] \leq \frac{\zeta^{(\epsilon)}}{2} + \overline{B}^{(\epsilon)} \tag{7}$$

where $\zeta^{(\epsilon)} \triangleq \langle \mathbf{d}^2, \left(\boldsymbol{\sigma}^{(\epsilon)}\right)^2\rangle + \mathrm{Var}(\Psi) + \epsilon^2$ and $\overline{B}^{(\epsilon)}$ is defined in (23).

Furthermore, if $\gamma(\epsilon) = O(\frac{1}{\sqrt[5]{\epsilon}})$, then $\lim_{\epsilon\downarrow 0} \overline{B}^{(\epsilon)} = 0$ and thus the RSG Algorithm is heavy-traffic optimal.

This result is interesting in that it provides an explicit scaling regime in which the design parameter $\gamma(\epsilon)$ can be increased to utilize the service regulating nature of the RSG Algorithm without sacrificing the heavy-traffic optimality. Intuitively, if $\gamma(\epsilon)$ scales slowly as ϵ vanishes, each link weight is dominated by its own queue length in the heavy-traffic regime, and thus the heavy-traffic optimality may be maintained; otherwise, the heavy-traffic optimality result may not hold, as will be demonstrated in Section V.

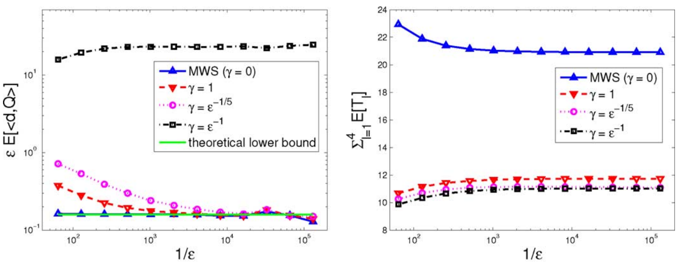
V. Simulation Results

In this section, we provide simulation results to compare the mean delay and service regularity performance of the RSG Algorithm with the MWS Algorithm. In the simulation, we consider a single-hop nonfading network with 4 links. Its capacity region is $\mathcal{R} = \{\boldsymbol{\lambda} = (\lambda_l)_{l=1}^4 \succeq \mathbf{0} : \sum_{l=1}^4 \lambda_l < 1\}$. We use arrival process where the number of arrivals in each slot follows a Bernoulli distribution. We consider the symmetric case $\boldsymbol{\lambda}^{(\epsilon)} = \left(1 - \frac{\epsilon}{2}\right) \times \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]$, and the asymmetric case $\boldsymbol{\lambda}^{(\epsilon)} = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\right] + \left(1 - \frac{\epsilon}{32}\right) \times \left[\frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right]$.

From Figs. 3(a) and 4(a), we can observe that the RSG Algorithm with both $\gamma = 1$ and $\gamma = \frac{1}{\sqrt[5]{\epsilon}}$ and the MWS Algorithm converge to the theoretical lower bound and thus is heavy-traffic optimal, which confirms our theoretical results. Yet, the RSG Algorithm with $\gamma = \frac{1}{\epsilon}$ has large mean queue length, which does not match with the theoretical lower bound and thus is not heavy-traffic optimal. Hence, γ should scale as slowly as $O\left(\frac{1}{\sqrt[5]{\epsilon}}\right)$ to preserve heavy-traffic optimality.

From Figs. 3(b) and 4(b), we can see that the RSG Algorithm with even $\gamma = 1$ significantly outperforms the MWS Algorithm

[4]We note that our result holds in single-hop network topologies without this assumption, and its extension to more general settings is part of our future work.



(a) (b)

Fig. 4. Asymmetric arrivals in a single-hop network. (a) Mean queue length. (b) Service regularity.

in terms of service regularity. More remarkably, the RSG Algorithm with $\gamma = \frac{1}{\sqrt[5]{\epsilon}}$ can achieve the lower bound (see [13]) achieved by the round-robin policy under symmetric arrivals.

VI. Detailed Heavy-Traffic Analysis

In this section, we prove Proposition 2 by using the analytical approach in [4], which includes two parts: 1) showing state-space collapse; 2) using the state-space collapse result to obtain an upper bound on the mean queue lengths. Yet, it is worth noting that the abrupt dynamics of TSLS counters pose significant challenges in heavy-traffic analysis. In particular, it requires new Lyapunov functions and a novel technique to establish heavy-traffic optimality of the RSG Algorithm.

A. State-Space Collapse

In this section, we establish a *state-space collapse* result under the RSG Algorithm. That is, we develop upper bounds for the deviation of steady-state queue lengths from the *line of attraction* $\mathbf{d}$ and TSLS counters. These upper bounds are crucial to establish our main result.

We have mentioned in Section III that the RSG Algorithm is throughput-optimal, i.e., it stabilizes all queues for any arrival rate vector that are strictly within the capacity region. Let $\{\mathbf{Q}^{(\epsilon)}[t]\}_{t\geq 0}$ and $\{\mathbf{T}^{(\epsilon)}[t]\}_{t\geq 0}$ be queue length processes and TSLS counters under the RSG Algorithm, respectively. Also, we use $\overline{\mathbf{Q}}^{(\epsilon)}$ and $\overline{\mathbf{T}}^{(\epsilon)}$ to denote their steady-state queue length random vector and TSLS random vector, respectively. Then, by the continuous mapping theorem, we have

$$\mathbf{Q}_{\|}^{(\epsilon)}[t] \Rightarrow \overline{\mathbf{Q}}_{\|}^{(\epsilon)} \quad \mathbf{Q}_{\perp}^{(\epsilon)}[t] \Rightarrow \overline{\mathbf{Q}}_{\perp}^{(\epsilon)} \tag{8}$$

$$\mathbf{T}_{\|}^{(\epsilon)}[t] \Rightarrow \overline{\mathbf{T}}_{\|}^{(\epsilon)} \quad \mathbf{T}_{\perp}^{(\epsilon)}[t] \Rightarrow \overline{\mathbf{T}}_{\perp}^{(\epsilon)} \tag{9}$$

where $\Rightarrow$ denotes convergence in distribution, and we define the projection and the perpendicular vector of any given L-dimensional vector $\mathbf{I}$ with respect to the normal vector $\mathbf{d}$ as

$$\mathbf{I}_{\|} \triangleq \langle \mathbf{d}, \mathbf{I}\rangle \mathbf{d} \quad \mathbf{I}_{\perp} \triangleq \mathbf{I} - \mathbf{I}_{\|}.$$

Next, we will show that under the RSG Algorithm, the second moment of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$ is bounded, dependent on $\gamma(\epsilon)$, while the second moment of $\|\overline{\mathbf{T}}^{(\epsilon)}\|$ is bounded by some constant independent of ϵ. Noting that the mean queue length under any policy scales at least the order of $1/\epsilon$ by Proposition 1, the state-space collapse happens in the following sense: By carefully selecting the scaling law of $\gamma(\epsilon)$ with respect to ϵ, both $\mathbb{E}\left[\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|^2\right]$ and $\mathbb{E}\left[\|\overline{\mathbf{T}}^{(\epsilon)}\|^2\right]$ are negligible compared to the large mean queue length for a sufficiently small ϵ.

Proposition 3: If $\Pr\{C_l[t] = 0\} > 0, \forall l \in \mathcal{L}$, then, under the RSG Algorithm, there exists a constant $N_{T,2}$, independent of ϵ, such that

$$\mathbb{E}[\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|^2] = O\left((\gamma(\epsilon))^4 (\log\gamma(\epsilon))^2\right) \tag{10}$$

$$\mathbb{E}[\|\boldsymbol{\beta}\cdot\overline{\mathbf{T}}^{(\epsilon)}\|^2] \leq N_{T,2} \tag{11}$$

where we recall that $\boldsymbol{\beta} \triangleq (\beta_l)_{l\in\mathcal{L}}$.

It is quite challenging to directly give an upper bound on $\mathbb{E}[\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|^2]$. Instead, we will first upper-bound the moment generation function of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$, and then use the relationship between the moments of a random variable and its moment generation function to upper-bound $\mathbb{E}[\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|^2]$.

In order to obtain an upper bound on the moment generation function of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$, we first study the drift of the Lyapunov function

$$V_{\perp}(\mathbf{Q}^{(\epsilon)}, \mathbf{T}^{(\epsilon)}) \triangleq \left\|\left(\mathbf{Q}_{\perp}^{(\epsilon)}, \sqrt{2\gamma(\epsilon)C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}^{(\epsilon)}}\right)\right\|$$

and show that when $V_{\perp}(\mathbf{Q}^{(\epsilon)}, \mathbf{T}^{(\epsilon)})$ is sufficiently large, it has a strictly negative drift independent of ϵ, which is characterized in the following key lemma.

Lemma 1: Under the RSG Algorithm, there exist positive constants κ and ς, independent of ϵ, such that whenever $V_{\perp}(\mathbf{Q}^{(\epsilon)}[t], \mathbf{T}^{(\epsilon)}[t]) > \kappa$, we have

$$\mathbb{E}[\Delta V_{\perp}(\mathbf{Q}^{(\epsilon)}[t], \mathbf{T}^{(\epsilon)}[t]) | \mathbf{Q}^{(\epsilon)}[t], \boldsymbol{T}^{(\epsilon)}[t]] < -\varsigma \tag{12}$$

where

$$\Delta V_{\perp}(\mathbf{Q}^{(\epsilon)}[t], \mathbf{T}^{(\epsilon)}[t]) \triangleq V_{\perp}(\mathbf{Q}^{(\epsilon)}[t+1], \mathbf{T}^{(\epsilon)}[t+1]) - V_{\perp}(\mathbf{Q}^{(\epsilon)}[t], \mathbf{T}^{(\epsilon)}[t]).$$

The proof of Lemma 1 is available in Appendix A.

In [4], after showing that $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$ has a negative drift under the MWS Algorithm when $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$ is large enough and observing that the drift of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$ is bounded, the authors utilize [7, Theorem 2.3] to develop an upper bound on the moment generation function of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$, which implies the existence of all moments of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$. However, the absolute value of the drift $\Delta V_{\perp}(\mathbf{Q}^{(\epsilon)}, \mathbf{T}^{(\epsilon)})$ has neither an upper bound nor an exponential tail given the current system state $(\mathbf{Q}^{(\epsilon)}, \mathbf{T}^{(\epsilon)})$ since the TSLS counters have bounded increment but unbounded decrement, i.e., they can at almost increase by 1 and drop to 0 once their corresponding links are scheduled. Therefore, we cannot directly apply [7, Theorem 2.3], which requires either boundedness or the exponential tail of the Lyapunov drift to establish the existence of the moment generation function of the system-state variables in addition to Lemma 1. Indeed, for a Markov chain with a strictly negative drift of Lyapunov function, if its Lyapunov drift has bounded increment but unbounded decrement, its moment generation function may not exist.

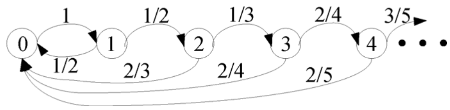


Fig. 5. Markov chain $\{X[t]\}_{t\geq 0}$.

Counterexample: Consider a Markov chain $\{X[t]\}_{t\geq 0}$ with the following transition probability:

$$P_{j,j+1}=\begin{cases}1, & \text{if } j=0\\ \frac{1}{2}, & \text{if } j=1\\ \frac{j-1}{j+1}, & \text{if } j\geq 2\end{cases}\quad P_{j,0}=\begin{cases}\frac{1}{2}, & j=1\\ \frac{2}{j+1}, & \text{if } j\geq 1.\end{cases}$$

The state transition diagram of Markov chain $\{X[t]\}_{t\geq 0}$ is shown in Fig. 5. Consider a linear Lyapunov function X. For any $X\geq 2$, we have

$$\mathbb{E}[X[t+1]-X[t]|X[t]=X]=\frac{X-1}{X+1}-\frac{2X}{X+1}=-1.$$

Thus, the Lyapunov function X has a strictly negative drift when $X\geq 2$ and hence the steady-state distribution of the Markov chain exists. Recall that its drift increases at almost by 1, but has unbounded decrement, which has similar dynamics with the system under the RSG Algorithm.

Next, we will show that even the first moment of this Markov chain does not exist, let alone its moment generation function. Let $\overline{X}$ be the steady-state random variable of the Markov chain and $\pi_j\triangleq\Pr\{\overline{X}=j\}$. According to the global balance equations, we can easily calculate

$$\pi_1=\pi_0=\frac{1}{3},\quad \pi_j=\frac{1}{3j(j-1)}\quad \forall j\geq 2. \tag{13}$$

Thus, we have

$$\mathbb{E}[\overline{X}]=\sum_{j=1}^{\infty}j\pi_j=\frac{1}{3}+\sum_{j=2}^{\infty}\frac{1}{3(j-1)}=\infty.$$

Fortunately, we can establish an upper bound on the moment generation function of $\|\overline{\mathbf{Q}}_{\perp}^{(\epsilon)}\|$ under the RSG Algorithm by exploiting the coupling between queue length processes and TSLS counters under the RSG Algorithm. Here, we need a mild assumption that $\Pr\{C_l[t]=0\}>0,\forall l\in\mathcal{L}$, which leads to the following lemma that all TSLS counters have an exponential tail independent of ϵ.

Lemma 2: If $p_l\triangleq\Pr\{C_l[t]=0\}>0,\forall l\in\mathcal{L}$, then, under the RSG Algorithm, there exists a $\vartheta\in(0,1)$, independent of ϵ, such that

$$\Pr\{T_l[t]\geq m\}\leq\vartheta^m\quad \forall t\geq m\geq 0,\ \forall l\in\mathcal{L}. \tag{14}$$

The proof of Lemma 2 is available in Appendix B.

Remark: We can also show that all TSLS counters still have an exponential tail independent of ϵ in nonfading single-hop network topologies. The extension to the more general setup is left for future research.

Lemma 2 directly implies (11). The rest of proof mainly builds on the analytical technique in [7], while it requires carefully partitioning the space $(\mathbf{Q}_{\perp}^{(\epsilon)},\mathbf{T}^{(\epsilon)})$ and exploiting the coupling between the queue length processes and TSLS counters. The detailed proof can be found in Appendix C.

B. Proof of Main Result

Having established the state-space collapse result, we are ready to show the heavy-traffic optimality of the RSG Algorithm. In this section, we first give an upper bound on $\mathbb{E}[\langle\mathbf{d},\overline{\mathbf{Q}}^{(\epsilon)}\rangle]$ under the RSG Algorithm and then establish its heavy-traffic optimality by selecting $\gamma(\epsilon)=O\left(\frac{1}{\sqrt[5]{\epsilon}}\right)$. In [13], we have shown that all moments of steady-state system variables, such as queue lengths and TSLS, are bounded under the RSG Algorithm, which enables us to analyze its heavy-traffic performance by using the methodology of "setting the drift of a Lyapunov function equal to zero."

We will omit the superscript ϵ associated with the queue lengths and TSLS counters for brevity in the rest of proof. In order to derive an upper bound on $\mathbb{E}[\langle\mathbf{d},\overline{\mathbf{Q}}^{(\epsilon)}\rangle]$, we need the following fundamental identity (see [4, Lemma 8]):

$$\frac{\mathbb{E}\left[\langle\mathbf{d},\mathbf{U}(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle^2\right]}{2}+\frac{\mathbb{E}\left[\langle\mathbf{d},\mathbf{A}-\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle^2\right]}{2}$$
$$+\mathbb{E}\left[\langle\mathbf{d},\overline{\mathbf{Q}}+\mathbf{A}-\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\langle\mathbf{d},\mathbf{U}(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right]$$
$$=\mathbb{E}\left[\langle\mathbf{d},\overline{\mathbf{Q}}\rangle\langle\mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right] \tag{15}$$

which is derived by setting the expected drift of $\langle\mathbf{d},\mathbf{Q}\rangle^2$ to 0.

Next, we give upper bounds for each individual term in the left-hand side (LHS) of (15) and a lower bound for the right-hand side (RHS) of (15). By simply setting the expected drift of $\langle\mathbf{d},\mathbf{Q}\rangle$ equal to zero, we have

$$\mathbb{E}\left[\langle\mathbf{d},\mathbf{U}(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right]=\langle\mathbf{d},\mathbb{E}[\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})]\rangle-\langle\mathbf{d},\boldsymbol{\lambda}\rangle$$
$$\overset{(a)}{=}\langle\mathbf{d},\mathbb{E}[\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})]\rangle-(b-\epsilon)\overset{(b)}{\leq}\epsilon \tag{16}$$

where the step (a) follows from the definition of $\boldsymbol{\lambda}^{(0)}\triangleq\boldsymbol{\lambda}+\epsilon\mathbf{d}$ and $\langle\mathbf{d},\boldsymbol{\lambda}^{(0)}\rangle=b$ (see Fig. 2); (b) follows from the facts that $\mathbb{E}[\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})]$ must be within capacity region $\mathcal{R}$ and that $\langle\mathbf{d},\mathbf{r}\rangle\leq b$ for any vector $\mathbf{r}\in\mathcal{R}$.

We are ready to provide an upper bound on the first term in the LHS of (15). Noting the fact that the amount of unused service in one time-slot at each link l cannot be greater than the maximum channel rate $C_{\max}$, i.e., $U_l\leq C_{\max}$, we have

$$\frac{1}{2}\mathbb{E}\left[\langle\mathbf{d},\mathbf{U}(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle^2\right]$$
$$\leq\frac{1}{2}\langle\mathbf{d},C_{\max}\mathbf{1}\rangle\mathbb{E}\left[\langle\mathbf{d},\mathbf{U}(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right]$$
$$\leq\frac{\epsilon}{2}\langle\mathbf{d},C_{\max}\mathbf{1}\rangle \tag{17}$$

where the last step utilizes inequality (16).

Next, we focus on the second term in the LHS of (15). The system stability under the RSG Algorithm implies that it selects the schedule $\mathbf{S}$, which maximizes $\langle\mathbf{d},\mathbf{S}\rangle$ with high probability when the arrival rate vector is very close to the face $\mathcal{H}^{(\mathbf{d})}$. See Fig. 6(a) for an example. Based on this observation, we can show

$$\mathbb{E}\left[\langle\mathbf{d},\mathbf{A}-\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle^2\right]$$
$$\leq\zeta^{(\epsilon)}+\sum_{\mathbf{c}\in\mathcal{C}}\frac{\epsilon}{\chi_{\mathbf{c}}}\left(2bb_{\mathbf{c}}+(b_{\mathbf{c}})^2+\langle\mathbf{d},C_{\max}\mathbf{1}\rangle^2\right) \tag{18}$$

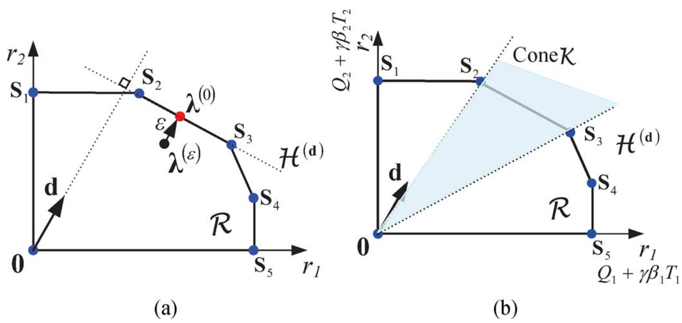


Fig. 6. Key properties of the RSG Algorithm (illustrated in nonfading case for an easier exposition): (a) when the arrival rate vector $\boldsymbol{\lambda}^{(\epsilon)}$ is very close to the face $\mathcal{H}^{(\mathbf{d})}$, the RSG Algorithm should select the feasible schedules $\mathbf{S}_2$ and $\mathbf{S}_3$ with high probability due to the system stability; (b) when the vector $\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T}$ is within the cone $\mathcal{K}$, then the RSG Algorithm selects the schedules $\mathbf{S}_2$ and $\mathbf{S}_3$.

where we recall that $\zeta^{(\epsilon)} \triangleq \langle \mathbf{d}^2, (\boldsymbol{\sigma}^{(\epsilon)})^2\rangle + \mathrm{Var}(\Psi) + \epsilon^2$ is defined in Proposition 2, and

$$\chi_{\mathbf{c}} \triangleq \min\left\{b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{r}\rangle : \text{for all } \mathbf{r}\in\mathcal{S}^{(\mathbf{c})} \setminus \{\mathbf{w} : b_{\mathbf{c}} = \langle \mathbf{d}, \mathbf{w}\rangle\}\right\}.$$

The detailed proof is provided in Appendix F. Inequality (18) indicates that the second moment of $\mathbf{d}$-weighted difference between arrivals and services is dominated by the $\mathbf{d}^2$-weighted variance of the arrival process and the variance of the channel fading process in the heavy-traffic limit.

In order to analyze the third term in the LHS of (15), we restrict vectors to $|\mathcal{L}_+|$-dimensional space, where $\mathcal{L}_+ \triangleq \{l \in \mathcal{L} : d_l > 0\}$. We use $\widetilde{I} \triangleq (I_l)_{l\in\mathcal{L}_+}$ to denote the vector $\mathbf{I}$ restricted in the $|\mathcal{L}_+|$-dimensional space. Therefore, we have

$$\begin{aligned}
&\mathbb{E}\left[\langle \mathbf{d}, \mathbf{Q}[t] + A[t] - \mathbf{C}[t]\cdot\mathbf{S}^*[t]\rangle\langle \mathbf{d}, \mathbf{U}[t]\rangle\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\langle \mathbf{d}, \mathbf{Q}[t+1]\rangle\langle \mathbf{d}, \mathbf{U}[t]\rangle\right] - \mathbb{E}\left[\langle \mathbf{d}, \mathbf{U}[t]\rangle^2\right] \\
&\overset{(b)}{\le} \mathbb{E}\left[\langle \mathbf{d}, \mathbf{Q}[t+1]\rangle\langle \mathbf{d}, \mathbf{U}[t]\rangle\right] \\
&= \mathbb{E}\left[\langle \widetilde{\mathbf{d}}, \widetilde{\mathbf{Q}}[t+1]\rangle\langle \widetilde{\mathbf{d}}, \widetilde{\mathbf{U}}[t]\rangle\right] \\
&\overset{(c)}{=} \mathbb{E}\left[\langle -\widetilde{\mathbf{Q}}_\perp[t+1], \widetilde{\mathbf{U}}[t]\rangle\right] \\
&\overset{(d)}{\le} \sqrt{\mathbb{E}\left[\|\widetilde{\mathbf{Q}}_\perp[t+1]\|^2\right]\mathbb{E}\left[\|\widetilde{\mathbf{U}}[t]\|^2\right]} \\
&\overset{(e)}{\le} \sqrt{\mathbb{E}\left[\|\mathbf{Q}_\perp[t+1]\|^2\right]\frac{C_{\max}}{d_{\min}}\mathbb{E}\left[\langle \mathbf{d}, \mathbf{U}[t]\rangle\right]}
\end{aligned} \tag{19}$$

where the step (a) utilizes (1); (b) is true since $\mathbb{E}\left[\langle \mathbf{d}, \mathbf{U}[t]\rangle^2\right] \ge 0$; (c) utilizes [4, Lemma 9]; (d) uses Cauchy–Schwarz inequality; (e) follows from the fact that $\|\widetilde{\mathbf{I}}\| \le \|\mathbf{I}\|$ for any vector $\mathbf{I}$ and the fact that $\widetilde{U}_l \le C_{\max}, \forall l \in \mathcal{L}_+$ and $d_{\min} \triangleq \min_{l\in\mathcal{L}_+} d_l$. By taking the limit $t\to\infty$ on both sides of (19), we have

$$\begin{aligned}
&\mathbb{E}\left[\langle \mathbf{d}, \overline{\mathbf{Q}} + \mathbf{A} - \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle\langle \mathbf{d}, \mathbf{U}(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle\right] \\
&\le \sqrt{\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2]\frac{C_{\max}}{d_{\min}}\mathbb{E}\left[\langle \mathbf{d}, \mathbf{U}(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle\right]} \\
&\le \sqrt{\epsilon\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2]\frac{C_{\max}}{d_{\min}}}
\end{aligned} \tag{20}$$

where the last step uses inequality (16).

Finally, we consider the RHS of (15). By using the definition of the RSG Algorithm, we can show

$$\begin{aligned}
&\mathbb{E}\left[\langle \mathbf{d}, \overline{\mathbf{Q}}\rangle\langle \mathbf{d}, \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C}) - \mathbf{A}\rangle\right] \\
&\ge \epsilon\mathbb{E}\left[\|\overline{\mathbf{Q}}_\parallel\|\right] - \cot(\theta)\sqrt{\left(\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2] + (\gamma(\epsilon))^2 N_{\mathbf{T},2}\right)\epsilon} \\
&\quad\times\sqrt{\sum_{\mathbf{c}\in\mathcal{C}}\frac{1}{\chi_{\mathbf{c}}}\left((b_{\mathbf{c}})^2 + \langle \mathbf{d}, C_{\max}\mathbf{1}\rangle^2\right)}
\end{aligned} \tag{21}$$

where $\theta \in \left(0, \frac{\pi}{2}\right]$ is an angle such that $\langle \mathbf{d}, \mathbf{R}^*(\mathbf{Q}, \mathbf{T})\rangle = b$, for all $\mathbf{Q}$ and $\mathbf{T}$ satisfying $\frac{\|(\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T})_\parallel\|}{\|\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T}\|} \ge \cos(\theta)$, $\mathbf{R}^*(\mathbf{Q}, \mathbf{T}) \triangleq \mathbb{E}[\mathbf{C}\cdot\mathbf{S}^*(\mathbf{Q}, \mathbf{T}, \mathbf{C})|\mathbf{Q}, \mathbf{T}]$ (Fig. 6(b)) provides an example illustrating this fact), and $N_{\mathbf{T},2}$ (cf. Proposition 3) is independent of ϵ. The proof is provided in Appendix G.

By substituting bounds (17), (18), (20), and (21) into identity (15), we have

$$\epsilon\mathbb{E}\left[\|\overline{\mathbf{Q}}_\parallel\|\right] \le \frac{\zeta^{(\epsilon)}}{2} + \overline{B}^{(\epsilon)} \tag{22}$$

where

$$\begin{aligned}
\overline{B}^{(\epsilon)} &\triangleq \frac{\epsilon}{2}\langle \mathbf{d}, C_{\max}\mathbf{1}\rangle + \sqrt{\epsilon\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2]\frac{C_{\max}}{c_{\min}}} \\
&\quad + \frac{1}{2}\sum_{\mathbf{c}\in\mathcal{C}}\frac{\epsilon}{\chi_{\mathbf{c}}}\left(2bb_{\mathbf{c}} + (b_{\mathbf{c}})^2 + \langle \mathbf{d}, C_{\max}\mathbf{1}\rangle^2\right) \\
&\quad + \cot(\theta)\sqrt{2\left(\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2] + (\gamma(\epsilon))^2 N_{\mathbf{T},2}\right)\epsilon} \\
&\quad\times\sqrt{\sum_{\mathbf{c}\in\mathcal{C}}\frac{1}{\chi_{\mathbf{c}}}\left((b_{\mathbf{c}})^2 + \langle \mathbf{d}, C_{\max}\mathbf{1}\rangle^2\right)}.
\end{aligned} \tag{23}$$

Thus, if $\lim_{\epsilon\downarrow 0}\overline{B}^{(\epsilon)} = 0$, then the RSG Algorithm is heavy-traffic optimal by Definition 1. Noting that $N_{\mathbf{T},2}$ is independent of ϵ, to satisfy $\lim_{\epsilon\downarrow 0}\overline{B}^{(\epsilon)} = 0$, it is sufficient to have

$$\lim_{\epsilon\downarrow 0}\epsilon\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2] = 0 \text{ and } \lim_{\epsilon\downarrow 0}\epsilon\left(\gamma(\epsilon)\right)^2 = 0. \tag{24}$$

By using the state-space collapse result established by Proposition 3, it is easy to see that $\gamma(\epsilon) = O\left(\frac{1}{\sqrt[5]{\epsilon}}\right)$ meets the above requirements.

VII. Conclusion

We studied the heavy-traffic behavior of the recently proposed maximum-weight type scheduling algorithm, called Regular Service Guarantee (RSG) Algorithm, where each link weight consists of its own queue length and a counter, namely the Time-Since-Last-Service (TSLS), which tracks the time since the last service. The RSG Algorithm has been shown to achieve throughput optimality while providing regular service guarantees. In this paper, we further showed that the RSG Algorithm is heavy-traffic optimal as long as its design parameter weighting for the TSLS $\gamma = O\left(\frac{1}{\sqrt[5]{\epsilon}}\right)$, where ϵ is the heavy-traffic parameter characterizing the closeness of the arrival rate vector to the boundary of the capacity region. Noting that the service regularity improves with increasing γ, our result reveals that the RSG Algorithm with a carefully selected parameter γ can achieve the best service regularity performance among the class of the RSG Algorithms without sacrificing the mean delay optimality under heavy-traffic conditions.

Appendix A
Proof of Lemma 1

We will omit ϵ associated with queue length processes, TSLS counters, and parameter $\gamma(\epsilon)$ for brevity. Also, we will use $\mathbf{I}^+$ to denote $\mathbf{I}[t+1]$ for any vector $\mathbf{I}$ in the rest of proof.

First, we note that

$$\begin{aligned}
&\Delta V_\perp(\mathbf{Q},\mathbf{T})\\
&\triangleq V_\perp(\mathbf{Q}^+,\mathbf{T}^+) - V_\perp(\mathbf{Q},\mathbf{T})\\
&=\sqrt{\|(\mathbf{Q}_\perp^+,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}^+})\|^2}-\sqrt{\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|^2}\\
&\overset{(a)}{\le} \frac{\|(\mathbf{Q}_\perp^+,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}^+})\|^2-\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|^2}{2\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\overset{(b)}{=} \frac{(W(\mathbf{Q}^+,\mathbf{T}^+)-W(\mathbf{Q},\mathbf{T}))-\big(W_\|(\mathbf{Q}^+,\mathbf{T}^+)-W_\|(\mathbf{Q},\mathbf{T})\big)}{2\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\overset{(c)}{=} \frac{\Delta W(\mathbf{Q},\mathbf{T})-\Delta W_\|(\mathbf{Q},\mathbf{T})}{2\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|},
\end{aligned}\tag{25}$$

where the step (a) follows from the fact that $f(x)=\sqrt{x}$ for $x\ge 0$ so that $f(y)-f(x)\le f'(x)(y-x)=\frac{y-x}{2\sqrt{x}}$ with $y=\|(\mathbf{Q}_\perp^+,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}^+})\|^2$ and $x=\|(\mathbf{Q}_\perp,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|^2$; (b) is true since $\|(\mathbf{I},\mathbf{J})\|^2=\|\mathbf{I}\|^2+\|\mathbf{J}\|^2$ for any vectors $\mathbf{I},\mathbf{J}$, $\|\mathbf{I}\|^2=\|\mathbf{I}_\|\|^2+\|\mathbf{I}_\perp\|^2$, $W(\mathbf{Q},\mathbf{T})\triangleq\|(\mathbf{Q},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|^2$, and $W_\|(\mathbf{Q},\mathbf{T})\triangleq\|\mathbf{Q}_\|\|^2$; (c) is true by letting

$$\begin{aligned}
\Delta W(\mathbf{Q},\mathbf{T}) &\triangleq W(\mathbf{Q}^+,\mathbf{T}^+)-W(\mathbf{Q},\mathbf{T})\\
\Delta W_\|(\mathbf{Q},\mathbf{T}) &\triangleq W_\|(\mathbf{Q}^+,\mathbf{T}^+)-W_\|(\mathbf{Q},\mathbf{T}).
\end{aligned}$$

Having inequality (25), we can first study the conditional expectation of $\Delta W(\mathbf{Q},\mathbf{T})$ and $\Delta W_\|(\mathbf{Q},\mathbf{T})$ instead of directly studying the conditional expectation of $\Delta V_\perp(\mathbf{Q},\mathbf{T})$. We first focus on $\mathbb{E}\left[\Delta W(\mathbf{Q},\mathbf{T})|\,\mathbf{Q},\mathbf{T}\right]$

$$\begin{aligned}
&\mathbb{E}\left[\Delta W(\mathbf{Q},\mathbf{T})|\,\mathbf{Q},\mathbf{T}\right]\\
&\quad=\mathbb{E}\big[\|(\mathbf{Q}^+,\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}^+})\|^2\\
&\qquad-\|(\mathbf{Q},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|^2|\mathbf{Q},\mathbf{T}\big]\\
&\quad=\mathbb{E}[\|\mathbf{Q}^+\|^2+2\gamma C_{\max}\|\boldsymbol{\beta}\cdot\mathbf{T}^+\|_1\\
&\qquad-\|\mathbf{Q}\|^2-2\gamma C_{\max}\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1|\mathbf{Q},\mathbf{T}]\\
&\quad=\mathbb{E}[\|\mathbf{Q}+\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}+\mathbf{U}\|^2-\|\mathbf{Q}\|^2\\
&\qquad+2\gamma C_{\max}(\|\boldsymbol{\beta}\cdot\mathbf{T}^+\|_1-\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1)|\mathbf{Q},\mathbf{T}]\\
&\quad=\mathbb{E}[\|\mathbf{Q}+\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\|^2+2\langle\mathbf{Q}+\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C},\mathbf{U}\rangle\\
&\qquad+\|\mathbf{U}\|^2-\|\mathbf{Q}\|^2+2\gamma C_{\max}(\|\boldsymbol{\beta}\cdot\mathbf{T}^+\|_1-\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1)|\mathbf{Q},\mathbf{T}]\\
&\quad\overset{(a)}{\le}\mathbb{E}[\|\mathbf{Q}+\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\|^2-\|\mathbf{Q}\|^2\\
&\qquad+2\gamma C_{\max}(\|\boldsymbol{\beta}\cdot\mathbf{T}^+\|_1-\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1)|\mathbf{Q},\mathbf{T}]\\
&\quad\overset{(b)}{=}\mathbb{E}\Bigg[2\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle+\|\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\|^2\\
&\qquad+2\gamma C_{\max}\left(\sum_{l=1}^{L}\beta_l-\sum_{l\in\mathbf{H}^*}\beta_l-\sum_{l\in\mathbf{H}^*}\beta_l T_l\right)\Bigg|\mathbf{Q},\mathbf{T}\Bigg]\\
&\quad\overset{(c)}{\le}2\mathbb{E}\left[\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]+K_1\\
&\qquad-2\gamma\mathbb{E}\left[\langle\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]
\end{aligned}\tag{26}$$

where the step (a) uses the fact that $U_l[t](Q_l[t]+A_l[t]-S_l^*[t]C_l[t])=-U_l^2[t]$ for each l; (b) follows from the fact that $\sum_{l=1}^{L}\beta_l T_l^+-\sum_{l=1}^{L}\beta_l T_l=\sum_{l=1}^{L}\beta_l-\sum_{l\in\mathbf{H}^*}\beta_l-\sum_{l\in\mathbf{H}^*}\beta_l T_l$ and $\mathbf{H}^*\triangleq\{l\in\mathcal{L}:S_l^*C_l>0\}$; (c) is true for $K_1\triangleq L\max\{A_{\max}^2,C_{\max}^2\}+2\gamma C_{\max}\sum_{l=1}^{L}\beta_l$.

Next, we consider $\mathbb{E}\left[\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]$. By using the definition of projection $\boldsymbol{\lambda}^{(0)}$, we have

$$\begin{aligned}
&\mathbb{E}\left[\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]\\
&\quad=\langle\mathbf{Q},\boldsymbol{\lambda}^{(0)}-\epsilon\mathbf{d}\rangle-\mathbb{E}[\langle\mathbf{Q},\mathbf{S}^*\cdot\mathbf{C}\rangle|\mathbf{Q},\mathbf{T}]\\
&\quad=-\epsilon\|\mathbf{Q}_\|\|+\langle\mathbf{Q},\boldsymbol{\lambda}^{(0)}\rangle-\mathbb{E}[\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{S}^*\cdot\mathbf{C}\rangle|\mathbf{Q},\mathbf{T}]\\
&\qquad+\gamma\mathbb{E}[\langle\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{S}^*\cdot\mathbf{C}\rangle|\mathbf{Q},\mathbf{T}].
\end{aligned}\tag{27}$$

Given the queue length vector $\mathbf{Q}$ and TSLS vector $\mathbf{T}$ at the beginning of each slot, according to the definition of the RSG Algorithm and the capacity region $\mathcal{R}$, it is easy to see that

$$\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbb{E}[\mathbf{S}^*\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle=\max_{\mathbf{r}\in\mathcal{R}}\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{r}\rangle.\tag{28}$$

Since $\boldsymbol{\lambda}^{(0)}$ is a relative interior point of dominant hyperplane $\mathcal{H}^{(\mathbf{d})}$, there exists a small enough $\delta>0$ such that $\mathcal{B}_\delta\triangleq\mathcal{H}^{(\mathbf{d})}\bigcap\left\{\mathbf{r}\succ\mathbf{0}:\|\mathbf{r}-\boldsymbol{\lambda}^{(0)}\|\le\delta\right\}$ representing the set of vectors on the hyperplane $\mathcal{H}^{(\mathbf{d})}$ that are within δ distance from $\boldsymbol{\lambda}^{(0)}$, lies strictly within the face $\mathcal{F}^{(\mathbf{d})}\triangleq\mathcal{H}^{(\mathbf{d})}\bigcap\mathcal{R}$. Therefore, we have

$$\begin{aligned}
\max_{\mathbf{r}\in\mathcal{R}}\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{r}\rangle&\ge\max_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{r}\rangle\\
&\ge\max_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q},\mathbf{r}\rangle+\langle\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{r}^*\rangle
\end{aligned}\tag{29}$$

where $\mathbf{r}^*\in\arg\max_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q},\mathbf{r}\rangle$.

Since normal vector $\mathbf{d}\succeq\mathbf{0}$ and arrival rate vector $\boldsymbol{\lambda}\succ\mathbf{0}$, we have $\boldsymbol{\lambda}^{(0)}\succ 0$. Therefore, we can find a $\delta>0$ sufficiently small such that $r_l\ge r_{\min}$ for all $\mathbf{r}=(r_l)_{l\in\mathcal{L}}\in\mathcal{B}_\delta$ and some $r_{\min}>0$. Hence, by (28) and (29), we have

$$\langle\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T},\mathbb{E}[\mathbf{S}^*\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle\ge\max_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q},\mathbf{r}\rangle+\gamma r_{\min}\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1.$$

By substituting above inequality into (27), we have

$$\begin{aligned}
&\mathbb{E}\left[\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]\\
&\quad\le-\epsilon\|\mathbf{Q}_\|\|+\min_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q},\boldsymbol{\lambda}^{(0)}-\mathbf{r}\rangle\\
&\qquad-\gamma r_{\min}\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1+\gamma\mathbb{E}[\langle\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{S}^*\cdot\mathbf{C}\rangle|\mathbf{Q},\mathbf{T}].
\end{aligned}$$

Since $\boldsymbol{\lambda}^{(0)}-\mathbf{r}$ is perpendicular to the normal vector $\mathbf{d}$ for $\mathbf{r}\in\mathcal{B}_\delta$, we have

$$\min_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q},\boldsymbol{\lambda}^{(0)}-\mathbf{r}\rangle=\min_{\mathbf{r}\in\mathcal{B}_\delta}\langle\mathbf{Q}_\perp,\boldsymbol{\lambda}^{(0)}-\mathbf{r}\rangle=-\delta\|\mathbf{Q}_\perp\|.$$

Hence, we have

$$\begin{aligned}
\mathbb{E}\left[\langle\mathbf{Q},\mathbf{A}-\mathbf{S}^*\cdot\mathbf{C}\rangle|\,\mathbf{Q},\mathbf{T}\right]&\le-\epsilon\|\mathbf{Q}_\|\|-\delta\|\mathbf{Q}_\perp\|\\
&\quad-\gamma r_{\min}\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1+\gamma\mathbb{E}[\langle\boldsymbol{\beta}\cdot\mathbf{T},\mathbf{S}^*\cdot\mathbf{C}\rangle|\mathbf{Q},\mathbf{T}].
\end{aligned}\tag{30}$$

Thus, by substituting (30) into (26), we have

$$\begin{aligned}
\mathbb{E}[\Delta W(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}]&\le-2\epsilon\|\mathbf{Q}_\|\|-2\delta\|\mathbf{Q}_\perp\|\\
&\quad-2\gamma r_{\min}\|\boldsymbol{\beta}\cdot\mathbf{T}\|_1+K_1.
\end{aligned}\tag{31}$$

Next, we lower-bound $\mathbb{E}[\Delta W_{\parallel}(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}]$

$$\begin{aligned}
&\mathbb{E}[\Delta W_{\parallel}(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}]\\
&\quad = \mathbb{E}[\langle \mathbf{d},\mathbf{Q}^{+}\rangle^2 - \langle \mathbf{d},\mathbf{Q}\rangle^2|\mathbf{Q},\mathbf{T}]\\
&\quad = \mathbb{E}[\langle \mathbf{d},\mathbf{Q}+\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}+\mathbf{U}\rangle^2 - \langle \mathbf{d},\mathbf{Q}\rangle^2|\mathbf{Q},\mathbf{T}]\\
&\quad = \mathbb{E}[\langle \mathbf{d},\mathbf{Q}+\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle^2 + \langle \mathbf{d},\mathbf{U}\rangle^2\\
&\qquad\quad + 2\langle \mathbf{d},\mathbf{Q}+\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle\langle \mathbf{d},\mathbf{U}\rangle - \langle \mathbf{d},\mathbf{Q}\rangle^2|\mathbf{Q},\mathbf{T}]\\
&\quad = \mathbb{E}[2\langle \mathbf{d},\mathbf{Q}\rangle\langle \mathbf{d},\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle + \langle \mathbf{d},\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle^2\\
&\qquad\quad + 2\langle \mathbf{d},\mathbf{Q}+\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle\langle \mathbf{d},\mathbf{U}\rangle + \langle \mathbf{d},\mathbf{U}\rangle^2|\mathbf{Q},\mathbf{T}]\\
&\quad = 2\langle \mathbf{d},\mathbf{Q}\rangle\langle \mathbf{d},\boldsymbol{\lambda}-\mathbb{E}[\mathbf{S}^{*}\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle\\
&\qquad - 2\mathbb{E}[\langle \mathbf{d},\mathbf{S}^{*}\cdot\mathbf{C}\rangle\langle \mathbf{d},\mathbf{U}\rangle|\mathbf{Q},\mathbf{T}] + \mathbb{E}[\langle \mathbf{d},\mathbf{U}\rangle^2|\mathbf{Q},\mathbf{T}]\\
&\qquad + \mathbb{E}[\langle \mathbf{d},\mathbf{A}-\mathbf{S}^{*}\cdot\mathbf{C}\rangle^2 + 2\langle \mathbf{d},\mathbf{Q}+\mathbf{A}\rangle\langle \mathbf{d},\mathbf{U}\rangle|\mathbf{Q},\mathbf{T}]\\
&\quad \geq 2\langle \mathbf{d},\mathbf{Q}\rangle\langle \mathbf{d},\boldsymbol{\lambda}-\mathbb{E}[\mathbf{S}^{*}\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle\\
&\qquad - 2\mathbb{E}[\langle \mathbf{d},\mathbf{S}^{*}\cdot\mathbf{C}\rangle\langle \mathbf{d},\mathbf{U}\rangle|\mathbf{Q},\mathbf{T}]\\
&\quad \overset{(a)}{\geq} 2\langle \mathbf{d},\mathbf{Q}\rangle\langle \mathbf{d},\boldsymbol{\lambda}-\mathbb{E}[\mathbf{S}^{*}\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle - K_2\\
&\quad \overset{(b)}{=} -2\epsilon\|\mathbf{Q}_{\parallel}\| - K_2\\
&\qquad + 2\|\mathbf{Q}_{\parallel}\|\left(\langle \mathbf{d},\boldsymbol{\lambda}^{(0)}\rangle - \langle \mathbf{d},\mathbb{E}[\mathbf{S}^{*}\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}]\rangle\right)\\
&\quad \overset{(c)}{\geq} -2\epsilon\|\mathbf{Q}_{\parallel}\| - K_2
\end{aligned}\tag{32}$$

where the step (a) is true for $K_2 \triangleq 2LC_{\max}^2$; step (b) uses the definition of projection $\boldsymbol{\lambda}^{(0)}$; step (c) follows from the fact that $\mathbb{E}[\mathbf{S}^{*}\cdot\mathbf{C}|\mathbf{Q},\mathbf{T}] \in \mathcal{R}$, $\mathcal{R} \subset \{\mathbf{r}\succeq \mathbf{0} : \langle \mathbf{d},\mathbf{r}\rangle \leq b\}$ and $b = \langle \mathbf{d}, \boldsymbol{\lambda}^{(0)}\rangle$.

By using (31), (32), and (25), we have

$$\begin{aligned}
&\mathbb{E}[\Delta V_{\perp}(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}]\\
&\quad\leq \frac{\mathbb{E}\left[\Delta W(\mathbf{Q},\mathbf{T}) - \Delta W_{\parallel}(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}\right]}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\quad\leq \frac{-2\delta\|\mathbf{Q}_{\perp}\| - 2\gamma r_{\min}\sum_{l=1}^{L}\beta_l T_l + K_1 + K_2}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}.
\end{aligned}$$

Note that

$$\begin{aligned}
\gamma\beta_l T_l &\geq \gamma\beta_l T_l \mathbb{1}_{\{\alpha T_l \geq 1\}} \geq \sqrt{\gamma\beta_l T_l}\mathbb{1}_{\{\gamma\beta_l T_l\geq 1\}}\\
&= \sqrt{\gamma\beta_l T_l} - \sqrt{\gamma\beta_l T_l}\mathbb{1}_{\{\gamma\beta_l T_l < 1\}} \geq \sqrt{\gamma\beta_l T_l} - 1
\end{aligned}$$

and $\|\mathbf{Q}_{\perp}\| \geq \frac{1}{\sqrt{L}}\|\mathbf{Q}_{\perp}\|_1$, where $\mathbb{1}_{\{.\}}$ is an indicator function. Thus, we have

$$\begin{aligned}
&\mathbb{E}[\Delta V_{\perp}(\mathbf{Q},\mathbf{T})|\mathbf{Q},\mathbf{T}]\\
&\quad\leq \frac{-\frac{2\delta}{\sqrt{L}}\|\mathbf{Q}_{\perp}\|_1 - 2r_{\min}\sum_{l=1}^{L}\sqrt{\gamma\beta_l T_l} + K_1 + K_2 + 2Lr_{\min}}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\quad\leq -\min\left\{\frac{2\delta}{\sqrt{L}},\frac{\sqrt{2}r_{\min}}{\sqrt{C_{\max}}}\right\}\frac{\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|_1}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\qquad + \frac{K_1+K_2+2Lr_{\min}}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}\\
&\quad\leq -\min\left\{\frac{\delta}{\sqrt{L}},\frac{r_{\min}}{\sqrt{2C_{\max}}}\right\} + \frac{K_1+K_2+2Lr_{\min}}{2\|(\mathbf{Q}_{\perp},\sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|}
\end{aligned}$$

where the last step uses the fact that $\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|$ for any vector $\mathbf{x}$. Hence, for any $0 < \varsigma < \min\left\{\frac{\delta}{\sqrt{L}},\frac{r_{\min}}{\sqrt{2C_{\max}}}\right\}$, by taking

$$\kappa \triangleq \frac{K_1+K_2+2Lr_{\min}}{2\left(\min\left\{\frac{\delta}{\sqrt{L}},\frac{r_{\min}}{\sqrt{2C_{\max}}}\right\} - \varsigma\right)}\tag{33}$$

we have the desired result.

Appendix B
Proof of Lemma 2

If the event

$$\mathcal{E}_l[j] \triangleq \{C_l[j] > 0, C_i[j] = 0 \quad \forall i \neq l; A_l[j-1] > 0\}\tag{34}$$

happens for some $j \in [t-m+1, t)$, then under the RSG Algorithm, link l should be scheduled at least once during the past m slots, and thus $T_l^{(\epsilon)}[t] < m$. This implies

$$\begin{aligned}
&\Pr\{T_l^{(\epsilon)}[t] \geq m\}\\
&\quad\leq \Pr\{\mathcal{E}_l[j] \text{ does not happen for all } j \in [t-m+1,t)\}\\
&\quad= \vartheta_l^m
\end{aligned}$$

where $\vartheta_l \triangleq 1-(1-q_l)(1-p_l)\Pi_{i\neq l}p_i$, and $q_l \triangleq \Pr\{A_l[t] = 0\}, \forall l \in \mathcal{L}$. Since $\lambda_l > 0$, we have $q_l < 1$ and thus $\vartheta_l \in (0,1)$. Thus, by taking $\vartheta \triangleq \max_{l\in\mathcal{L}}\vartheta_l$, we have the desired result.

Appendix C
Proof of Proposition 3

In the rest of the proof, we will omit ϵ associated with the queue length processes, the TSLS counters, and parameter $\gamma(\epsilon)$ for brevity. As we pointed it out in Section VI, it is challenging to directly provide an upper bound on $\mathbb{E}[\|\overline{\mathbf{Q}}_{\perp}\|^2]$. Therefore, we will first upper-bound the moment generation function of $\|\overline{\mathbf{Q}}_{\perp}\|$, and then establish the relationship between the moments of a random variable and its moment generation function to upper-bound $\mathbb{E}[\|\overline{\mathbf{Q}}_{\perp}\|^2]$.

Lemma 3: If $\Pr\{C_l[t] = 0\} > 0, \forall l \in \mathcal{L}$, then, under the RSG Algorithm, there exists a $\eta > 0$ satisfying $\vartheta e^{\eta F_1} < 1$ and $e^{\eta G_2} - \eta(G_2+\varsigma) < 1$ such that

$$\mathbb{E}\left[e^{\eta\|\overline{\mathbf{Q}}_{\perp}\|}\right] \leq \frac{G}{1-\rho}\tag{35}$$

where

$$G \triangleq e^{\eta G_1} + \frac{Le^{\eta F_2}}{1-e^{\eta F_1}\vartheta}\tag{36}$$

$$\rho \triangleq e^{\eta G_2} - \eta(G_2+\varsigma) \in (0,1)\tag{37}$$

$$F_1 \triangleq L\sqrt{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l} + \sqrt{1+2\gamma C_{\max}L\sum_{l=1}^{L}\beta_l}\tag{38}$$

$$F_2 \triangleq \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} + 2L\max\{A_{\max}, C_{\max}\}\tag{39}$$

$$\begin{aligned}
G_1^2 \triangleq{}& \kappa^2 + 2\left(A_{\max} + 2\sqrt{L}C_{\max}\right)\kappa\sqrt{L}\\
&+ L\left(A_{\max}+2\sqrt{L}C_{\max}\right)^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l
\end{aligned}\tag{40}$$

$$G_2 \triangleq 2L\max\{A_{\max}, C_{\max}\} + \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} + 2\gamma C_{\max} L \sum_{l=1}^{L}\beta_l \tag{41}$$

ϑ is defined in Lemma 2, ς and κ are defined in (33).

In Lemma 1, we have show that $V_\perp(\mathbf{Q},\mathbf{T}) \triangleq \|(\mathbf{Q}_\perp, \sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}})\|$ has a negative drift whenever $V_\perp(\mathbf{Q},\mathbf{T})$ is large enough. However, as we pointed it out in Section VI, we cannot directly provide an upper bound on the moment generation function of $V_\perp(\mathbf{Q},\mathbf{T})$ by [7, Theorem 2.3], which requires either boundedness or the exponential tail of the Lyapunov drift. This is due to the abrupt dynamics of TSLS counters. We resolve this issue by exploiting the coupling between the queue length processes and TSLS counters. Please see Appendix D for details.

Since we are interested in the scaling law of γ to preserve heavy-traffic optimality under the RSG Algorithm, we will write an upper bound on the moment generation function of $\|\overline{\mathbf{Q}}_\perp\|$ as a function of γ based on Lemma 3.

First, it is easy to see that $\kappa = O(\gamma)$, $G_1 = O(\gamma)$, $G_2 = O(\gamma)$, $F_1 = O(\sqrt{\gamma})$, and $F_2 = O(1)$. Note that we need to choose a $\eta > 0$ such that

$$\vartheta e^{\eta F_1} < 1 \tag{42}$$

$$e^{\eta G_2} - \eta(G_2+\varsigma) < 1. \tag{43}$$

It is not hard to verify that

$$0 < \eta \le \frac{1}{2}\min\left\{\frac{1}{F_1}\ln\frac{1}{\vartheta}, \frac{1}{G_2}\ln\frac{G_2+\varsigma}{G_2}\right\} \tag{44}$$

satisfies above requirements. If γ is large enough such that $\frac{\varsigma}{G_2} < 1$ and $G_2 \gg F_1$, then we have

$$\frac{1}{G_2}\ln\frac{G_2+\varsigma}{G_2} \le \frac{1}{F_1}\ln\frac{1}{\vartheta}. \tag{45}$$

Thus, we can take $\eta^* \triangleq \frac{1}{2G_2}\ln\frac{G_2+\varsigma}{G_2}$ to meet the above requirements, and hence $\eta^* = O\left(\frac{1}{\gamma^2}\right)$.

Taking $\eta = \eta^*$ and noting that $\eta^* < \frac{1}{2F_1}\ln\frac{1}{\vartheta}$, we have

$$\begin{aligned}\frac{G}{1-\rho} &= \frac{e^{\eta^* G_1} + \frac{Le^{\eta^* F_2}}{1-\vartheta e^{\eta^* F_1}}}{1-\left(e^{\eta^* G_2} - \eta^*(G_2+\varsigma)\right)} \\ &\le \frac{e^{\eta^* G_1} + \frac{Le^{\eta^* F_2}}{1-\sqrt{\vartheta}}}{1-\left(e^{\eta^* G_2} - \eta^*(G_2+\varsigma)\right)} \\ &= \frac{e^{\eta^* G_1} + \frac{Le^{\eta^* F_2}}{1-\sqrt{\vartheta}}}{1-\left(1+\frac{\varsigma}{G_2}\right)^{\frac{1}{2}} + \frac{1}{2}\left(1+\frac{\varsigma}{G_2}\right)\ln\left(1+\frac{\varsigma}{G_2}\right)} \\ &\overset{(a)}{=} O\left(\frac{1}{1-\left(1+\frac{\varsigma}{2G_2}\right)+\frac{1}{2}\left(1+\frac{\varsigma}{G_2}\right)\frac{\varsigma}{G_2}}\right) = O\left(\gamma^2\right)\end{aligned}$$

where the step (a) uses $\eta^* G_1 = O\left(\frac{1}{\gamma}\right)$ and $\eta^* F_2 = O\left(\frac{1}{\gamma^2}\right)$. Thus, by Lemma 3, we have $\mathbb{E}\left[e^{\eta^*\|\overline{\mathbf{Q}}_\perp\|}\right] = O(\gamma^2)$.

Having obtained the upper bound on $\mathbb{E}\left[e^{\eta^*\|\overline{\mathbf{Q}}_\perp\|}\right]$, we need to establish the relationship between the moments of a random variable and its moment generation function to upper-bound $\mathbb{E}[\|\overline{\mathbf{Q}}_\perp\|^2]$, as shown in the following lemma.

Lemma 4: For a random variable X with $\mathbb{E}[e^{\eta X}] < \infty$ for some $\eta > 0$, we have

$$\mathbb{E}[X^n] \le \frac{1}{\eta^n}\left(\log\left(e^{n-1}\mathbb{E}[e^{\eta X}]\right)\right)^n \tag{46}$$

for $n = 1, 2, 3\ldots$.

Please see Appendix E for the proof of Lemma 4.

By taking $n = 2$ in Lemma 4, we have the desired result.

Appendix D
Proof of Lemma 3

Let $\mathbf{Z}[t] \triangleq \left(\mathbf{Q}_\perp[t], \sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\mathbf{T}[t]}\right)$. We first give an upper bound on $\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|}\middle|\mathbf{Q}[t],\mathbf{T}[t]\right]$. To that end, let $l^*[t] \in \arg\max_l \beta_l T_l[t]$. We partition $(\mathbf{Q}_\perp[t], \mathbf{T}[t])$ into sets $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$, where

$$\mathcal{F}_1 \triangleq \{\|\mathbf{Z}[t]\| \le \kappa\}; \mathcal{F}_2 \triangleq \left\{\|\mathbf{Z}[t]\| > \kappa, \|\mathbf{Q}_\perp[t]\| > T_{l^*[t]}[t]\right\}$$
$$\mathcal{F}_3 \triangleq \left\{\|\mathbf{Z}[t]\| > \kappa, \|\mathbf{Q}_\perp[t]\| \le T_{l^*[t]}[t]\right\}.$$

Then, we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|}\middle|\mathbf{Q}[t],\mathbf{T}[t]\right] = \sum_{i=1}^{3}\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|}; \mathcal{F}_i\middle|\mathbf{Q}[t],\mathbf{T}[t]\right]. \tag{47}$$

Next, we consider each term in (47) individually.

i) On event $\mathcal{F}_1$, we have

$$\|\mathbf{Z}[t]\| = \sqrt{\|\mathbf{Q}_\perp[t]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t]} \le \kappa$$

which implies $\|\mathbf{Q}_\perp[t]\| \le \kappa$. For $\mathbf{Q}_\perp[t+1]$, we have

$$\begin{aligned}\mathbf{Q}_\perp[t+1] &= \mathbf{Q}[t+1] - \langle\mathbf{d},\mathbf{Q}[t+1]\rangle\mathbf{d} \\ &= (\mathbf{Q}[t]+\mathbf{A}[t]-\mathbf{S}[t]+\mathbf{U}[t]) \\ &\quad - \langle\mathbf{d},\mathbf{Q}[t]+\mathbf{A}[t]-\mathbf{S}[t]+\mathbf{U}[t]\rangle\mathbf{d} \\ &= \mathbf{Q}[t] - \langle\mathbf{d},\mathbf{Q}[t]\rangle\mathbf{d} + (\mathbf{A}[t]+\mathbf{U}[t]+\langle\mathbf{d},\mathbf{S}[t]\rangle\mathbf{d}) \\ &\quad - (\mathbf{S}[t]+\langle\mathbf{d},\mathbf{A}[t]+\mathbf{U}[t]\rangle\mathbf{d}) \\ &\overset{(a)}{\preceq} \mathbf{Q}_\perp[t] + A_{\max}\mathbf{1} + C_{\max}\mathbf{1} + \sqrt{L}C_{\max}\mathbf{d} \\ &\preceq \mathbf{Q}_\perp[t] + (A_{\max} + 2\sqrt{L}C_{\max})\mathbf{1}\end{aligned} \tag{48}$$

where step (a) uses the following inequality:

$$\langle\mathbf{d},\mathbf{S}[t]\rangle \le \|\mathbf{d}\|\|\mathbf{S}[t]\| \le \sqrt{L}C_{\max}.$$

Hence, we have

$$\begin{aligned}
\|\mathbf{Z}[t+1]\|^2 &= \|\mathbf{Q}_\perp[t+1]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t+1] \\
&\overset{(a)}{\leq} \|\mathbf{Q}_\perp[t] + (A_{\max} + 2\sqrt{L}C_{\max})\mathbf{1}\|^2 \\
&\quad + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l (T_l[t]+1) \\
&= \|\mathbf{Z}[t]\|^2 + 2\left(A_{\max} + 2\sqrt{L}C_{\max}\right)\|\mathbf{Q}_\perp[t]\|_1 \\
&\quad + L\left(A_{\max} + 2\sqrt{L}C_{\max}\right)^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l \\
&\overset{(b)}{\leq} \kappa^2 + 2\left(A_{\max} + 2\sqrt{L}C_{\max}\right)\kappa\sqrt{L} \\
&\quad + L\left(A_{\max} + 2\sqrt{L}C_{\max}\right)^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l \triangleq G_1^2
\end{aligned} \tag{49}$$

where step (a) uses the inequality (48); step (b) utilizes the inequality $\|\mathbf{x}\|_1 \leq \sqrt{L}\|\mathbf{x}\|$ for any L-dimensional vector $\mathbf{x}$. Hence, we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|};\mathcal{F}_1 \,\middle|\, \mathbf{Q}[t],\mathbf{T}[t]\right] \leq e^{\eta G_1}. \tag{50}$$

In order to analyze other two terms in (47), we need the following lemma.

Lemma 5: Under the RSG Algorithm, if $\|\mathbf{Z}[t]\| > \kappa$, then

$$\begin{aligned}
&\big|\|\mathbf{Z}[t+1]\| - \|\mathbf{Z}[t]\|\big| \\
&\quad \leq 2L\max\{A_{\max}, C_{\max}\} + \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} \\
&\quad + 2\gamma C_{\max}\frac{\sum_{l\in\mathbf{H}^*}\beta_l T_l[t]}{\sqrt{\|\mathbf{Q}_\perp[t]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t]}}
\end{aligned} \tag{51}$$

where $\mathbf{H}^* \triangleq \{l : S_l^*[t]C_l[t] > 0\}$.

The proof is omitted here due to space limitations and is available in our technical report [14].

ii) On event $\mathcal{F}_2$, we have

$$\frac{\sum_{l\in\mathbf{H}^*}\beta_l T_l[t]}{\sqrt{\|\mathbf{Q}_\perp[t]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t]}} \leq \frac{L\beta_{l^*[t]}T_{l^*[t]}[t]}{\|\mathbf{Q}_\perp[t]\|} \leq L\sum_{l=1}^{L}\beta_l.$$

By substituting above inequality into (51), we get

$$\big|\|\mathbf{Z}[t+1]\| - \|\mathbf{Z}[t]\|\big| \leq G_2 \tag{52}$$

where

$$G_2 \triangleq 2L\max\{A_{\max}, C_{\max}\} + \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} + 2\gamma C_{\max}L\sum_{l=1}^{L}\beta_l.$$

Noting that (12) and (52) satisfy conditions of [7, Lemma 2.2], there exists $\eta_1 > 0$ such that

$$\mathbb{E}\left[e^{\eta(\|\mathbf{Z}[t+1]\| - \|\mathbf{Z}[t]\|)};\mathcal{F}_2 \,\middle|\, \mathbf{Q}[t],\mathbf{T}[t]\right] \leq \rho \quad \forall 0 < \eta < \eta_1,$$

where $\rho \triangleq e^{\eta G_2} - \eta(G_2 + \varsigma) \in (0,1)$, independent of ϵ. Thus, we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|};\mathcal{F}_2 \,\middle|\, \mathbf{Q}[t],\mathbf{T}[t]\right] \leq \rho e^{\eta\|\mathbf{Z}[t]\|}. \tag{53}$$

iii) On event $\mathcal{F}_3$, we have

$$\begin{aligned}
&\frac{\sum_{l\in\mathbf{H}^*}\beta_l T_l[t]}{\sqrt{\|\mathbf{Q}_\perp[t]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t]}} \\
&\leq \frac{L\beta_{l^*[t]}T_{l^*[t]}[t]}{\sqrt{2\gamma C_{\max}\beta_{l^*[t]}T_{l^*[t]}[t]}} \\
&= \frac{L}{\sqrt{2\gamma C_{\max}}}\sqrt{\beta_{l^*[t]}T_{l^*[t]}[t]} \\
&\leq \frac{L\sqrt{\sum_{l=1}^{L}\beta_l}}{\sqrt{2\gamma C_{\max}}}\sqrt{T_{l^*[t]}[t]} \\
&\leq \frac{L\sqrt{\sum_{l=1}^{L}\beta_l}}{\sqrt{2\gamma C_{\max}}}T_{l^*[t]}[t].
\end{aligned} \tag{54}$$

By substituting (54) into (51), we get

$$\begin{aligned}
&\big|\|\mathbf{Z}[t+1]\| - \|\mathbf{Z}[t]\|\big| \\
&\quad \leq 2L\max\{A_{\max}, C_{\max}\} + \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} \\
&\quad + T_{l^*[t]}[t]L\sqrt{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}.
\end{aligned} \tag{55}$$

In addition, on event $\mathcal{F}_3$, we have

$$\begin{aligned}
\|\mathbf{Z}[t]\| &= \sqrt{\|\mathbf{Q}_\perp[t]\|^2 + 2\gamma C_{\max}\sum_{l=1}^{L}\beta_l T_l[t]} \\
&\leq \sqrt{T_{l^*[t]}^2[t] + 2\gamma C_{\max}L\beta_{l^*[t]}T_{l^*[t]}[t]} \\
&\leq \sqrt{T_{l^*[t]}^2[t] + 2\gamma C_{\max}LT_{l^*[t]}[t]\sum_{l=1}^{L}\beta_l} \\
&\leq T_{l^*[t]}[t]\sqrt{1 + 2\gamma C_{\max}L\sum_{l=1}^{L}\beta_l}.
\end{aligned} \tag{56}$$

Hence, by utilizing (55) and (56), we have

$$\begin{aligned}
\|\mathbf{Z}[t+1]\| &\leq \|\mathbf{Z}[t]\| + \big|\|\mathbf{Z}[t+1]\| - \|\mathbf{Z}[t]\|\big| \\
&\leq F_1 T_{l^*[t]}[t] + F_2
\end{aligned} \tag{57}$$

where $F_1 \triangleq L\sqrt{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l} + \sqrt{1+2\gamma C_{\max}L\sum_{l=1}^{L}\beta_l}$ and $F_2 \triangleq \frac{2\gamma C_{\max}\sum_{l=1}^{L}\beta_l}{\kappa} + 2L\max\{A_{\max}, C_{\max}\}$. Thus

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|}; \mathcal{F}_3 \middle| \mathbf{Q}[t], \mathbf{T}[t]\right] \le e^{\eta F_2} e^{\eta F_1 T_{l^*[t]}[t]}. \quad (58)$$

By substituting (50), (53), and (58) into (47), we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|} | \mathbf{Q}[t], \mathbf{T}[t]\right] \le e^{\eta G_1} + \rho e^{\eta\|\mathbf{Z}[t]\|} + e^{\eta F_2} e^{\eta F_1 T_{l^*[t]}[t]}.$$

By taking expectation on both sides, we have

$$\begin{aligned}\mathbb{E}&\left[e^{\eta\|\mathbf{Z}[t+1]\|}\right] \\ &\le e^{\eta G_1} + \rho\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t]\|}\right] + e^{\eta F_2}\mathbb{E}\left[e^{\eta F_1 T_{l^*[t]}[t]}\right] \\ &\le e^{\eta G_1} + \rho\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t]\|}\right] + e^{\eta F_2}\sum_{l=1}^{L}\mathbb{E}\left[e^{\eta F_1 T_l[t]}\right]. \quad (59)\end{aligned}$$

Next, we will upper-bound the term $\mathbb{E}\left[e^{\eta F_1 T_l[t]}\right]$

$$\begin{aligned}\mathbb{E}\left[e^{\eta F_1 T_l[t]}\right] &\overset{(a)}{=} \sum_{m=0}^{t} e^{\eta F_1 m}\Pr\{T_l[t]=m\} \\ &\le \sum_{m=0}^{t} e^{\eta F_1 m}\Pr\{T_l[t]\ge m\} \\ &\overset{(b)}{\le} \sum_{m=0}^{t} e^{\eta F_1 m}\vartheta^m \\ &\le \sum_{m=0}^{\infty} e^{\eta F_1 m}\vartheta^m \\ &\overset{(c)}{=} \frac{1}{1-e^{\eta F_1}\vartheta} \quad (60)\end{aligned}$$

where the step (a) uses the fact that $T_l[t] \le t$ for any $t \ge 0$; (b) follows from Lemma 2; (c) is true for $0 < \eta < \eta_2$ and $\vartheta e^{\eta_2 F_1} < 1$.

By substituting (60) into (59), we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t+1]\|}\right] \le \rho\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t]\|}\right] + G \quad (61)$$

holding for $0 < \eta < \eta_0 \triangleq \min\{\eta_1, \eta_2\}$, where $G \triangleq e^{\eta G_1} + \frac{Le^{\eta F_2}}{1-e^{\eta F_1}\vartheta}$. By using (61) and iterating over t, we have

$$\mathbb{E}\left[e^{\eta\|\mathbf{Z}[t]\|}\right] \le \rho^t e^{\eta\|\mathbf{Z}[0]\|} + \frac{1-\rho^t}{1-\rho}G.$$

Letting $t \to \infty$ on both sides of the above inequality, we have

$$\mathbb{E}\left[e^{\eta\|(\overline{\mathbf{Q}}_\perp, \sqrt{2\gamma C_{\max}\boldsymbol{\beta}\cdot\overline{\mathbf{T}}})\|}\right] \le \frac{G}{1-\rho}$$

which implies the desired result.

Appendix E
Proof of Lemma 4

$$\begin{aligned}\mathbb{E}[X^n] = \frac{1}{\eta^n}\mathbb{E}\left[\left(\log e^{\eta X}\right)^n\right] &\overset{(a)}{\le} \frac{1}{\eta^n}\mathbb{E}\left[\left(\log\left(e^{n-1}e^{\eta X}\right)\right)^n\right] \\ &\overset{(b)}{\le} \frac{1}{\eta^n}\left(\log\left(e^{n-1}\mathbb{E}[e^{\eta X}]\right)\right)^n\end{aligned}$$

where the step (a) follows from the fact that $f(y) = (\log y)^n$ is increasing in $y \in [1,\infty)$ for $n = 1, 2, \ldots$; (b) uses the fact that $g(y) = \left(\log\left(e^{n-1}y\right)\right)^n$ is concave in $[1,\infty)$ for $n = 1, 2, \ldots$, and Jensen's Inequality.

Appendix F
Proof of Inequality (18)

To show inequality (18), we need the following lemma.

Lemma 6: Let

$$\pi_{\mathbf{c}} \triangleq \Pr\left\{\langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle = b_{\mathbf{c}} \middle| \mathbf{C} = \mathbf{c}\right\}$$

and

$$\chi_{\mathbf{c}} \triangleq \min\left\{b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{r}\rangle : \text{for all } \mathbf{r} \in \mathcal{S}^{(\mathbf{c})} \setminus \{\mathbf{w} : b_{\mathbf{c}} = \langle \mathbf{d}, \mathbf{w}\rangle\}\right\}.$$

Then, for each channel state $\mathbf{c} \in \mathcal{C}$, and any $\epsilon \in (0, \chi_{\mathbf{c}}\psi_{\mathbf{c}})$, we have

$$1 - \pi_{\mathbf{c}} \le \frac{\epsilon}{\chi_{\mathbf{c}}\psi_{\mathbf{c}}} \quad (62)$$

where we recall that $\psi_{\mathbf{c}} \triangleq \Pr\{\mathbf{C} = \mathbf{c}\}$.

The proof mainly follows from the stability condition, i.e., $\mathbb{E}\left[\langle \mathbf{d}, \mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right] \ge \langle \mathbf{d}, \boldsymbol{\lambda}\rangle = b - \epsilon$, and is similar to that of [4, Claim 1]. We omit the proof here for conciseness. Lemma 6 implies that

$$\begin{aligned}\mathbb{E}&\left[\left(b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle\right)^2 \middle| \mathbf{C} = \mathbf{c}\right] \\ &= \mathbb{E}\left[\left(b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle\right)^2 \middle| \langle \mathbf{d}, \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle \ne b_{\mathbf{c}}\right] \\ &\quad \times (1 - \pi_{\mathbf{c}}) \\ &\le \frac{\epsilon}{\chi_{\mathbf{c}}\psi_{\mathbf{c}}}\left((b_{\mathbf{c}})^2 + \langle \mathbf{d}, C_{\max}\mathbf{1}\rangle^2\right). \quad (63)\end{aligned}$$

Similarly, we have

$$\mathbb{E}\left[b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle \middle| \mathbf{C} = \mathbf{c}\right] \le \frac{\epsilon b_{\mathbf{c}}}{\chi_{\mathbf{c}}\psi_{\mathbf{c}}}. \quad (64)$$

For $\mathbb{E}\left[\langle \mathbf{d}, \mathbf{A} - \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle^2 \middle| \mathbf{C} = \mathbf{c}\right]$, we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d}, \mathbf{A} - \mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{C})\rangle^2 \middle| \mathbf{C} = \mathbf{c}\right] \\ &= \mathbb{E}\left[\left(\langle \mathbf{d}, \mathbf{A}\rangle - b + b - b_{\mathbf{c}} + b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right)^2\right] \\ &= \mathbb{E}\left[\left(\langle \mathbf{d}, \mathbf{A}\rangle - b\right)^2\right] + (b - b_{\mathbf{c}})^2 \\ &\quad + \mathbb{E}\left[\left(b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right)^2\right] \\ &\quad + 2\left(\langle \mathbf{d}, \boldsymbol{\lambda}\rangle - b\right)(b - b_{\mathbf{c}}) \\ &\quad + 2\left(\langle \mathbf{d}, \boldsymbol{\lambda}\rangle - b\right)\mathbb{E}\left[b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right] \\ &\quad + 2(b - b_{\mathbf{c}})\mathbb{E}\left[b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right]. \quad (65)\end{aligned}$$

Next, we give upper bounds for each individual term in the right-hand side of (65). We will repeatedly use the identity

$$\langle \mathbf{d}, \boldsymbol{\lambda}\rangle - b = -\epsilon \quad (66)$$

where it follows from the definition of $\boldsymbol{\lambda}^{(0)}$ (see Fig. 2). By noting that $b_{\mathbf{c}} \triangleq \max_{\mathbf{s}\in\mathcal{S}^{(\mathbf{c})}}\langle \mathbf{d}, \mathbf{c}\cdot\mathbf{s}\rangle$, we have

$$\begin{aligned}(\langle \mathbf{d}, \boldsymbol{\lambda}\rangle - b)\,&\mathbb{E}\left[b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right] \\ &= -\epsilon\mathbb{E}\left[b_{\mathbf{c}} - \langle \mathbf{d}, \mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}}, \overline{\mathbf{T}}, \mathbf{c})\rangle\right] \le 0. \quad (67)\end{aligned}$$

In addition, by using inequality (64), we have

$$(b-b_{\mathbf{c}})\,\mathbb{E}\left[b_{\mathbf{c}}-\langle \mathbf{d},\mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{c})\rangle\right] \le b\mathbb{E}\left[b_{\mathbf{c}}-\langle \mathbf{d},\mathbf{c}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{c})\rangle\right] \le b\frac{\epsilon b_{\mathbf{c}}}{\chi_{\mathbf{c}}\psi_{\mathbf{c}}}. \quad (68)$$

For $\mathbb{E}\left[(\langle \mathbf{d},\mathbf{A}\rangle-b)^2\right]$, we have

$$\begin{aligned}\mathbb{E}&\left[(\langle \mathbf{d},\mathbf{A}\rangle-b)^2\right]\\ &=\mathbb{E}\left[(\langle \mathbf{d},\mathbf{A}\rangle-\langle \mathbf{d},\boldsymbol{\lambda}\rangle+\langle \mathbf{d},\boldsymbol{\lambda}\rangle-b)^2\right]\\ &=\mathbb{E}\left[(\langle \mathbf{d},\mathbf{A}\rangle-\langle \mathbf{d},\boldsymbol{\lambda}\rangle-\epsilon)^2\right]\\ &=\left\langle \mathbf{d}^2,\left(\boldsymbol{\sigma}^{(\epsilon)}\right)^2\right\rangle+\epsilon^2. \end{aligned} \quad (69)$$

Thus, by substituting (63) and (66)–(69) into (65), we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\mathbf{A}-\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle^2\middle|\,\mathbf{C}=\mathbf{c}\right]\\ &\le\left\langle \mathbf{d}^2,\left(\boldsymbol{\sigma}^{(\epsilon)}\right)^2\right\rangle+\epsilon^2+(b-b_{\mathbf{c}})^2\\ &\quad+\frac{\epsilon}{\chi_{\mathbf{c}}\psi_{\mathbf{c}}}\left(2bb_{\mathbf{c}}+(b_{\mathbf{c}})^2+\langle \mathbf{d},C_{\max}\mathbf{1}\rangle^2\right)-2\epsilon\,(b-b_{\mathbf{c}}).\end{aligned}$$

By taking the expectation on both sides of the above inequality, we have the desired result.

APPENDIX G
PROOF OF INEQUALITY (21)

Let $\mathbf{R}^*(\mathbf{Q}[t],\mathbf{T}[t])\triangleq\mathbb{E}[\mathbf{C}\cdot\mathbf{S}^*(\mathbf{Q}[t],\mathbf{T}[t],\mathbf{C}[t])|\mathbf{Q}[t],\mathbf{T}[t]]$. For the face $\mathcal{F}\triangleq\mathcal{H}^{(\mathbf{d})}\cap\mathcal{R}$ of the region $\mathcal{R}$, there exists an angle $\theta\in(0,\frac{\pi}{2}]$ such that $\langle \mathbf{d},\mathbf{R}^*(\mathbf{Q},\mathbf{T})\rangle=b$, for all $\mathbf{Q}$ and $\mathbf{T}$ satisfying $\frac{\|(\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T})_{\|}\|}{\|\mathbf{Q}+\gamma\boldsymbol{\beta}\cdot\mathbf{T}\|}\ge\cos(\theta)$. Note that

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\overline{\mathbf{Q}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]\\ &=\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]\\ &\quad-\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right].\end{aligned} \quad (70)$$

For $\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]$, we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]\\ &=\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle(b-\langle \mathbf{d},\mathbf{A}\rangle)\right]\\ &\quad-\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\left(b-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)\right].\end{aligned} \quad (71)$$

By using the fact that the arrivals are independent of system state, we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle(b-\langle \mathbf{d},\mathbf{A}\rangle)\right]\\ &=\epsilon\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\right]\\ &=\epsilon\mathbb{E}\left[\|\overline{\mathbf{Q}}_{\|}\|\right]+\gamma\epsilon\mathbb{E}\left[\|(\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\|}\|\right].\end{aligned} \quad (72)$$

For $\mathbb{E}\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\left(b-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)\right]$, we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\left(b-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)\right]\\ &=\mathbb{E}\left[\mathbb{E}\left[\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\|}\|\left(b-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)\middle|\,\overline{\mathbf{Q}},\overline{\mathbf{T}}\right]\right]\\ &=\mathbb{E}\left[\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\|}\|\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)\right]\\ &\overset{(a)}{=}\mathbb{E}\left[\|\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\|\cos(\phi)\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)\right]\\ &\overset{(b)}{=}\mathbb{E}\left[\|\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\|\cos(\phi)\mathbb{1}_{\{\phi>\theta\}}\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)\right]\\ &\overset{(c)}{=}\mathbb{E}\left[\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\perp}\|\cot(\phi)\mathbb{1}_{\{\phi>\theta\}}\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)\right]\\ &\overset{(d)}{\le}\cot(\theta)\mathbb{E}\left[\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\perp}\|\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)\right]\\ &\overset{(e)}{\le}\cot(\theta)\sqrt{\mathbb{E}\left[\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\perp}\|^2\right]\mathbb{E}\left[\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)^2\right]}\\ &\le\cot(\theta)\sqrt{\mathbb{E}\left[\|(\overline{\mathbf{Q}}_{\perp}\|^2+\gamma^2\|\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\|^2\right]}\\ &\quad\times\sqrt{\mathbb{E}\left[\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)^2\right]}\end{aligned} \quad (73)$$

where the step (a) is true for that ϕ is the angle between vector $\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}$ and the normal vector $\mathbf{d}$; (b) follows from the definition of θ; (c) uses $\|(\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\perp}\|=\|\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\|\sin(\phi)$; (d) follows from the fact that cotangent function is decreasing in $\left(0,\frac{\pi}{2}\right]$; (e) uses Cauchy–Schwartz Inequality.

Next, let us consider $\mathbb{E}\left[\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)^2\right]$

$$\begin{aligned}\mathbb{E}&\left[\left(b-\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\right)^2\right]\\ &\overset{(a)}{=}\mathbb{E}\left[\left(\mathbb{E}\left[\Psi-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\middle|\,\overline{\mathbf{Q}},\overline{\mathbf{T}}\right]\right)^2\right]\\ &\overset{(b)}{\le}\mathbb{E}\left[\mathbb{E}\left[\left(\Psi-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)^2\middle|\,\overline{\mathbf{Q}},\overline{\mathbf{T}}\right]\right]\\ &=\mathbb{E}\left[\left(\Psi-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)^2\right]\\ &=\sum_{\mathbf{c}\in\mathcal{C}}\psi_{\mathbf{c}}\mathbb{E}\left[\left(b_{\mathbf{c}}-\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})\rangle\right)^2\middle|\,\mathbf{C}=\mathbf{c}\right]\\ &\overset{(c)}{\le}\epsilon\sum_{\mathbf{c}\in\mathcal{C}}\frac{1}{\chi_{\mathbf{c}}}\left((b_{\mathbf{c}})^2+\langle \mathbf{d},C_{\max}\mathbf{1}\rangle^2\right)\end{aligned} \quad (74)$$

where the step (a) follows from the definition of Ψ with distribution $\Pr\{\Psi=b_{\mathbf{c}}\}=\psi_{\mathbf{c}}$ for $\mathbf{c}\in\mathcal{C}$, and the definition of $\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})$; (b) uses Jensen's Inequality; (c) uses (63).

Thus, by substituting (72)–(74) into (71), we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\overline{\mathbf{Q}}+\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]\\ &\ge\epsilon\mathbb{E}\left[\|\overline{\mathbf{Q}}_{\|}\|\right]+\gamma\epsilon\mathbb{E}\left[\|(\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\|}\|\right]\\ &\quad-\cot(\theta)\sqrt{\left(\mathbb{E}\left[\|\overline{\mathbf{Q}}_{\perp}\|^2\right]+\gamma^2\mathbb{E}\left[\|\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\|^2\right]\right)\epsilon}\\ &\quad\times\sqrt{\sum_{\mathbf{c}\in\mathcal{C}}\frac{1}{\chi_{\mathbf{c}}}\left((b_{\mathbf{c}})^2+\langle \mathbf{d},C_{\max}\mathbf{1}\rangle^2\right)}.\end{aligned} \quad (75)$$

For $\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]$, we have

$$\begin{aligned}\mathbb{E}&\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\right]\\ &=\mathbb{E}\left[\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\langle \mathbf{d},\mathbf{C}\cdot\mathbf{S}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}},\mathbf{C})-\mathbf{A}\rangle\middle|\,\overline{\mathbf{Q}},\overline{\mathbf{T}}\right]\right]\\ &=\mathbb{E}\left[\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\left(\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle-\langle \mathbf{d},\mathbf{A}\rangle\right)\middle|\,\overline{\mathbf{Q}},\overline{\mathbf{T}}\right]\right]\\ &=\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle\left(\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle-b+b-\langle \mathbf{d},\boldsymbol{\lambda}\rangle\right)\right]\\ &\overset{(a)}{\le}\mathbb{E}\left[\langle \mathbf{d},\gamma\boldsymbol{\beta}\cdot\overline{\mathbf{T}}\rangle(b-\langle \mathbf{d},\boldsymbol{\lambda}\rangle)\right]\\ &\overset{(b)}{=}\gamma\epsilon\mathbb{E}\left[\|(\boldsymbol{\beta}\cdot\overline{\mathbf{T}})_{\|}\|\right]\end{aligned} \quad (76)$$

where the step (a) uses the fact that $\langle \mathbf{d},\mathbf{R}^*(\overline{\mathbf{Q}},\overline{\mathbf{T}})\rangle\le b$; (b) uses (66). By substituting (75) and (76) into (70), we have the desired result.

REFERENCES

[1] M. Bramson, "State space collapse with application to heavy traffic limits for multiclass queueing networks," *Queueing Syst.*, vol. 30, no. 1, pp. 89–140, 1998.

[2] L. Bui, R. Srikant, and A. Stolyar, "Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2936–2940.

[3] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length based scheduling and congestion control," in *Proc. IEEE INFOCOM*, Miami, FL, USA, Mar. 2005, pp. 1794–1803.

[4] A. Eryilmaz and R. Srikant, "Asymptotically tight steady-state queue length bounds implied by drift conditions," *Queueing Syst.*, vol. 72, pp. 311–359, 2012.

[5] G. Foschini and J. Salz, "A basic dynamic routing problem and diffusion," *IEEE Trans. Commun.*, vol. COM-26, no. 3, pp. 320–327, Mar. 1978.

[6] G. R. Gupta and N. Shroff, "Delay analysis for multi-hop wireless networks," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2356–2364.

[7] B. Hajek, "Hitting-time and occupation-time bounds implied by drift analysis with applications," *Adv. Appl. Probab.*, vol. 14, no. 3, pp. 502–525, 1982.

[8] I. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS for wireless," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 486–494.

[9] J. Jaramillo, R. Srikant, and L. Ying, "Scheduling for optimal rate allocation in ad hoc networks with heterogeneous delay constraints," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 5, pp. 979–987, May 2011.

[10] B. Ji, C. Joo, and N. B. Shroff, "Delay-based back-pressure scheduling in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1539–1552, Oct. 2013.

[11] B. Li and A. Eryilmaz, "Optimal distributed scheduling under time-varying conditions: A Fast-CSMA algorithm with applications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3278–3288, Jul. 2013.

[12] B. Li, R. Li, and A. Eryilmaz, "Heavy-traffic-optimal scheduling design with regular service guarantees in wireless networks," in *Proc. ACM MobiHoc*, Jul. 2013, pp. 79–88.

[13] B. Li, R. Li, and A. Eryilmaz, "Throughput-optimal scheduling design with regular service guarantees in wireless networks," *IEEE/ACM Trans. Netw.*, 2014, DOI: 10.1109/TNET.2014.2333008, to be published.

[14] B. Li, R. Li, and A. Eryilmaz, "Wireless scheduling design for optimizing both service regularity and mean delay in heavy-traffic regimes," Tech. Rep., 2015 [Online]. Available: http://www2.ece.ohio-state.edu/~eryilmaz/HeavyTrafficRegularscheduling_report.pdf

[15] R. Li, A. Eryilmaz, and B. Li, "Throughput-optimal scheduling with regulated inter-service times," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 2616–2624.

[16] X. Lin and N. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proc. IEEE Conf. Decision Control*, Paradise Island, Bahamas, Dec. 2004, pp. 1484–1489.

[17] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1260–1267, Aug. 1999.

[18] M. Neely, E. Modiano, and C. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proc. IEEE INFOCOM*, Miami, FL, USA, Mar. 2005, pp. 1723–1734.

[19] D. Shah and D. Wischik, "Switched networks with maximum weight policies: Fluid approximation and multiplicative state space collapse," *Ann. Appl. Probab.*, vol. 22, no. 1, pp. 70–127, 2012.

[20] A. Stolyar, "Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 1–53, 2004.

[21] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Syst.*, vol. 50, no. 4, pp. 401–457, 2005.

[22] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 36, no. 12, pp. 1936–1948, Dec. 1992.

[23] W. Whitt, "Weak convergence theorems for priority queues: Preemptive-resume discipline," *J. Appl. Probab.*, vol. 8, no. 1, pp. 74–94, 1971.

[24] R. Williams, "Diffusion approximations for open multiclass queueing networks: Sufficient conditions involving state space collapse," *Queueing Syst.*, vol. 30, no. 1, pp. 27–88, 1998.

[25] H. Xiong, R. Li, A. Eryilmaz, and E. Ekici, "Delay-aware cross-layer design for network utility maximization in multi-hop networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 5, pp. 951–959, May 2011.

[26] L. Ying, S. Shakkottai, and A. Reddy, "On combining shortest-path and back-pressure routing over multihop wireless networks," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 1674–1682.

Bin Li (S'12–M'15) received the B.S. degree in electronic and information engineering and M.S. degree in communication and information engineering from Xiamen University, Xiamen, China, in 2005 and 2008, respectively, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2014.

He is currently a Postdoctoral Researcher with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA. His research interests include wireless communication and networks, cloud computing, resource allocation and management, distributed algorithm design, queueing theory, and optimization theory.

Dr. Li received the Presidential Fellowship from The Ohio State University and Chinese Government Award for Outstanding Ph.D. Students Abroad in 2013.

Ruogu Li (S'10) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2007, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2012.

His Ph.D. dissertation discussed the principles and methods of adaptive network algorithm design under various quality-of-service requirements. He is now a Staff System Engineer with ASSIA, Inc., Redwood City, CA, USA. His research interests include optimal network control, wireless communication networks, low-delay scheduling scheme design, and cross-layer algorithm design.

Atilla Eryilmaz (S'00–M'06) received the B.S. degree in electrical and electronics engineering from Bogaziçi University, Istanbul, Turkey, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2001 and 2005, respectively.

Between 2005 and 2007, he worked as a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. He is currently an Associate Professor of electrical and computer engineering with The Ohio State University, Columbus, OH, USA, where he has been a faculty member since 2007. His research interests include design and analysis for complex networked systems with focus on wireless communication and power networks, optimal control of stochastic networks, optimization theory, distributed algorithms, network pricing, and information theory.

Dr. Eryilmaz served as a TPC Chair for the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2015, and has been an Associate Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING since 2015. He was a coauthor of the paper that won the Best Student Paper Award in WiOpt 2012. He received the NSF CAREER Award in 2010 and two Lumley Research Awards for Research Achievements in 2010 and 2015.