OPTIMAL NETWORK CONTROL FOR BIG AND HEAVY DATA DELIVERY

A Dissertation by

Shuang Xia

Master of Science, Wichita State University, 2014

Bachelor of Engineering, Wuhan University of Science and Technology, 2010

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

July 2018

OPTIMAL NETWORK CONTROL FOR BIG AND HEAVY DATA DELIVERY

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Electrical Engineering and Computer Science.

Hyuck M. Kwon, Committee Chair

Pu Wang, Committee Member

John Watkins, Committee Member

Yanwu Ding, Committee Member

Xiaoming Hu, Committee Member

Accepted for the College of Engineering

Royce Bowden, Dean

Accepted for the Graduate School

Dennis Livesay, Dean

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Dr. Hyuck M. Kwon and Dr. Pu Wang, for their continuous guidance and support. Dr. Kwon is the greatest and kindest professor in the world, always trying his best to prepare for class, and to explain very careful every equation and concept to his students. Thus, his students can always learn a lot from his class. Moreover, Dr. Kwon is the greatest advisor. He guided me through my research and helped me with proofreading my draft papers step by step. His professional expertise, perceptiveness, and passion will always be an inspiration to me. I always believe that, during our lifetime, we will encounter many significant people who provide us with considerable assistance in terms of our research, studies, and career. Dr. Kwon is one of those significant people in my life. He has been a tremendous help when I have had a difficult time, and I will never forget this. In addition, Dr. Wang is the person who opened the door to my research and provided much assistance with my research. There is no doubt that without the advising, support, and encouragement from Drs. Kwon and Wang, I could not have accomplished this dissertation. Moreover, I would like to thank Dr. John Watkins, Dr. Yanwu Ding, and Dr. Xiaoming Hu for their kind help and for serving on my dissertation committee. In addition, I also thank Ms. Kristie Bixby for proofreading my paper.

Furthermore, I thank my wife Shan Shi for her endless support and encouragement. Meeting her has been the most fortunate thing in my life. I also thank my parents, Guorong Xia and Rongjian Yang, for all of their kind support during my life. In addition, I want to say thank you to my uncle Rongde Yang for his help in making it possible for me to study in the United States. Finally, I thank my grandparents, Surong Yuan and Yingmin Yang, for their support.

This dissertation is supported in part by the U.S. Air Force (USAF) Research Laboratory under Grant Nos. FA9453-17-1-0020 and FA2386-14-1-0026, and the U.S. National Science Foundation (NSF) under Grant Nos. 1547373 and 1446557.

ABSTRACT

Due to new emerging applications and technologies, data traffic has increased dramatically. However, wireless network spectrum resources are very limited. To overcome this challenge, optimal network control, which can regulate optimal transmission times for network clients and support the largest set of client traffic rates while maintaining network stability, is required. However, the most optimal network control policies have been developed under light-tailed (LT) traffic conditions. When arrival traffic involves heavy-tailed (HT) traffic, the network fails to achieve queue stability under these policies. Thus, for a single-hop network, distributed maximum weight (Max-Weight) scheduling, which solves the above issue by adding alpha power on the queue backlog, has been proposed. Nevertheless, even though maximum weight-alpha scheduling overcomes this existing issue, it still must to decide the value for different kinds of arrival traffic, which is not easy to implement in a real environment. Hence, to further solve the problems faced here, we propose a time-average stochastic gradient scheduling algorithm (TA-SGSA), which decouples the queue-length update process and the dual-variable update process in such a way that both queue stabilities can be achieved. Similarly, for a multi-hop network, we propose a time-average stochastic gradient routing algorithm (TA-SGRA), which operates under similar principle but considers a cross-layer network system model. In addition, in order to explore the convergence and stability of the proposed algorithms, we use the ordinary differential equation (ODE) method since the HT traffic has an unbounded mean and variance. Finally, simulation results are presented to verify theoretical results for the proposed TA-SGSA and TA-SGRA algorithms, whereby an LT traffic and an LT routing flow queue can share the network with bounded mean and bounded variance in queue length, even in the presence of HT traffic and an HT routing flow queue.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

CHAPTER 1

INTRODUCTION

1.1 Background

Over the last few years, data traffic has dramatically increased due to a change in the way that today′s society creates, shares, and consumes information. However, wireless network spectrum resources are very limited. To address such challenges, the optimal network control, which can regulate the optimal transmission times for network clients and support the largest set of client traffic rates, while maintaining network stability, is urgent. Currently, one of the important and extremely popular classes of throughput optimal network control is the Lyapunov optimal control [1] [2], whereby you can choose the control policy that minimizes drift of the Lyapunov function, which represents a scalar measure of the network states such as queue length and queueing delay. In particular, for single-hop wireless networks (e.g., WiFi and cellular), the Lyapunov optimal control policy is reduced to the maximum-weight (Max-Weight) scheduling algorithm that serves users and clients who have the maximum product of queue length and link rate. Similarly, for multi-hop wireless networks (e.g., wireless sensor and wireless ad hoc), the Lyapunov optimal control policy is equivalent to the backpressure routing algorithm, which finds the largest queue length of difference and routes packets in the corresponding direction. In addition, it is proven that the Lyapunov optimal control policy can prevent the queue length from going to infinity [1] [3]. Thus, the Lyapunov optimal control policy is throughput optimal [4].

Even though these Lyapunov optimal control policies have good performance, they are generally studied and developed under a light-tailed (LT) traffic condition, thus failing to counter the unique features of heavy-tailed (HT) traffic. More specifically, large volume and high variability are features of HT data traffic that can be modeled by HT distribution, but the less variable data traffic is usually modeled by LT distribution. For example, someone sends a highlight video to his friend using some chatting apps that have a highly variable file

size. This can be modeled as HT traffic [5]. In contrast, someone browses web pages with low variability, which can be considered LT traffic. In addition, HT traffic widely exists in a variety of communication [6] [7] [8] and computer networks, even processor [9] [10] [11], such as mobile ad-hoc networks [12], cellular networks [13], WiFi networks [14], data center networks [15], and generalized processor sharing [16] [17] [18]. Moreover, HT traffic is also caused by the inherent heavy-tailed distribution in the traffic source, such as the file size on internet servers, the message size of cellular base stations [19], and the frame length of variable bit rate (VBR) video streams [20], or by the network protocols themselves, such as retransmissions and random access schemes [21]. In particular, due to the high burstiness or dependence of HT traffic over a long-range time scale, the network performance significantly degrades in terms of stability [22], latency [3] [23], and connectivity [24], thereby destructively impacting the throughput optimality of routing and scheduling policies.

To counter such challenges, distributed maximum weight-α scheduling(DMWS-α) has been proposed [3], which makes scheduling decisions based on adding α power that is determined by the burstiness of HT traffic on the queue length. By properly choosing the α value, LT queues can receive more transmission opportunities compared with the conventional Max-Weight scheduling algorithm [25]. Hence, LT queues have a bounded queue length. However, even though DMWS-α has a bounded mean and is throughput optimal under HT traffic condition, it needs to decide the α value for different traffic arrivals, which is not easy to implement in a real environment. Hence, it is necessary to further develop some algorithms that can guarantee network stability and optimal network utility without having any knowledge of the statistical information of arrival traffic flows and time-varying channel conditions. Instead, only the current queue length and channel state information [26] are required.

To date, there has been active research on stochastic network utility maximization(NUM), whereby the optimal network resource allocation problems, such as congestion control [27], routing [28] [29], and scheduling [23], are formulated as a constrained maximiza-

tion of some utility function under the stochastic dynamics in users' traffic and time-varying wireless channels. By treating the stochastic NUM problem as a non-linear optimization problem, the classic stochastic gradient and subgradient algorithms can be adopted. These algorithms utilize Lagrange multipliers analysis to convert the network control decisions (e.g., scheduling, routing, and congestion control) into a Lagrange dual function [27], to treat Lagrange dual variables as a function of queue length, and to update the dual variables along the stochastic gradient or subgradient directions. In particular, the stochastic NUM problem generally aims to maximize network utility while guaranteeing queue stability, i.e., ensuring that each user has a bounded expected queue length [22]. Due to the unique stochastic features of heavy-tailed traffic, the classic subgradient/gradient algorithms can face fundamental challenges in simultaneously achieving utility optimality and queue stability. This is because queues with HT traffic arrivals (HT queues) inherently experience heavy-tailed distributed queueing delay, which implies that HT queues have a much higher chance of experiencing a very large queue length, compared with queues with LT arrivals (LT queues). However, the subgradient/gradient algorithms normally exploit queue length to update dual variables. This can lead HT queues to receive much more service opportunities with high probabilities because of their large queue lengths. Consequently, LT queues can be seriously starved, which may lead to unbounded queue length and queue instability.

1.2 Research Objectives and Solutions

To address the above challenges, this dissertation proposes two throughput and utility optimal algorithms. For a single-hop network, we propose a time-average stochastic gradient scheduling algorithm (TA-SGSA), which treats queue length as a Lagrange variable and updates the Lagrange variable along with the direction of the stochastic gradient descent to find the optimal value. For a multi-hop network, we propose a time-average stochastic gradient routing algorithm (TA-SGRA), which can guarantee that an LT flow has a bounded mean in the queue length.

1.2.1 Time-Average Stochastic Gradient Scheduling Algorithm

In this section, an effective scheduling algorithm to maximize network utility in the presence of heavy-tailed traffic and also light-tailed traffic is developed. First, it is proven that, without considering the inherent features of HT traffic, the classic stochastic subgradient scheduling algorithm (SSSA), which is utility optimal under only LT traffic, fails to achieve queue stability under both HT and LT traffic. More specifically, by exploiting the properties of regenerative processes [30] along with asymptotic queueing analysis, it is shown that under the classic subgradient algorithm, the tail distribution of the LT queue is at least one order heavier than the tail distribution of the HT arrivals. This implies by moment theory that the LT queue can be of unbounded queue length. To counter this challenge, a time-average stochastic gradient scheduling algorithm, which decouples the queue-length update process and the dual-variable update process in such a way that both utility optimality and queue stability can be simultaneously achieved, even in the presence of heavy-tailed traffic, is proposed. In particular, our theoretical analysis will show that under this algorithm, the percentage of a network resource, e.g., transmission time, allocated to each queue converges to a constant as time proceeds. Such a feature can prohibit the LT queues from competing with the HT queues, thus completely shielding those LT queues from the destructive impact of HT traffic.

1.2.2 Time-Average Stochastic Gradient Routing Algorithm

In this section, an effective routing algorithm to maximize network utility in the presence of heavy-tailed traffic as well as light-tailed traffic is developed. In particular, this dissertation considers a multi-hop wireless network, which is formed by a collection of interconnected wireless routers. Each router has a two-level queue structure. First, at the network layer, according to the destination of each flow, multiple flow queues corresponding to each flow at each router are built. For example, for one specific router, two types of flows (i.e., HT and LT) flow through it. Thus, two flow queues to record the queue-length information are built. On the other hand, at the media access control (MAC) layer, multiple

link queues for corresponding neighbor routers are built. For instance, one specific router that has two neighbor routers will create two link queues for them. In addition, through queue stability analysis, it will be determined and proven that, without considering the inherent features of HT traffic, the classic stochastic gradient algorithm (SGA), which is utility optimal under only LT traffic, fails to achieve queue stability under both HT and LT traffic flows. Furthermore, by exploiting the properties of regenerative processes [31] along with asymptotic queueing analysis [32], it will be shown that under the classic SGA, the tail distribution of the LT flow queue is at least one order heavier than the tail distribution of the HT flow arrivals. According to moment theory [33], this implies that the LT flow queue has an unbounded queue length. To counter such a challenge, at the network layer, a time-average stochastic gradient routing algorithm, which separates the flow queue-length update process and the dual-variable update process so that both utility optimality and queue stability can be simultaneously achieved even in the presence of heavy-tailed traffic, is prposed. At the MAC layer, to avoid the interference from neighbors, at each time slot, a collision-free link rate vector is selected. More specifically, by theoretical analysis, it will be found that the transmission rate for each flow converges to a constant as time proceeds under our proposed routing algorithm. Consequently, LT flow queues will no longer compete with HT flow queues for transmission time. Thus, it can be guaranteed that the LT flow queues have a bounded queue length. On the other hand, to show that our link queues are also stable, the Lyapunov method [3] is adopted to show that they have bounded means. In addition, due to the characteristics of HT traffic (e.g., unbounded mean and variance), a novel ordinary differential equation (ODE) approach to explore the convergence and stability properties of the TA-SGRA is used.

1.3 Organizations of Dissertation

The remainder of this dissertation is organized as follows: Chapter 2 will introduce the fundamentals of heavy tails. Chapter 3 proposes a time-average stochastic gradient scheduling algorithm in the presence of heavy-tailed traffic. In addition, a distributed max-

imum weight-α scheduling algorithm and a classic SGA are presented for comparison. The conventional SGA that does not work under an HT traffic condition is shown. Then, the proposed algorithm will prove that it provide good performance in the presence of heavy tails will be proven. Chapter 4 proposes time-average stochastic gradient routing algorithm in the presence of heavy-tailed traffic. Then, its throughput optimality will be proven, and the simulation results will be shown. Chapter 5 concludes the dissertation.

CHAPTER 2

FUNDAMENTALS OF HEAVY TAILS

In this dissertation, the following notations are used: for any two real functions $a(t)$ and $b(t)$, $a(t) \sim b(t)$ denotes $\lim_{t\to\infty} a(t)/b(t) = 1$. Also, let $F(x) = P(X \leq x)$ denote the cumulative distribution function (CDF) of a non-negative random variable (r.v.) X. Let $\overline{F}(x) = P(X > x)$ denote its tail distribution function.

Definition 1. *Random variable X is heavy-tailed, if for all $\theta > 0$*

$$\lim_{x\to\infty} e^{\theta x}\overline{F}(x) = \infty, \tag{2.1}$$

and random variable X is light-tailed, if it is not heavy-tailed [30].

The above definition indicates that an r.v. is HT if its tail distribution decreases slower than exponentially. An r.v. is LT if its tail distribution decreases exponentially or faster. Representative HT distributions include Pareto and log-normal, and typical LT distributions include exponential and Poisson. Figure 2.1 shows an example of HT and LT distributions. Based on the existence of the moments of an r.v., we define the tail coefficient of a non-negative random variable.

Definition 2. *The tail coefficient κ_X of a nonnegative random variable X is defined by [30]*

$$\kappa_X = \sup\{k \geq 0 : E[X^k] < \infty\}. \tag{2.2}$$

The tail coefficient defines the threshold order above which an r.v. possesses infinite moments. A certain group of HT distributions, such as Pareto, have finite tail coefficients or equivalently have infinite moments of certain orders higher than tail coefficients. In contrast, other HT distributions, such as log-normals, have infinite tail coefficients or equivalently have finite moments of all orders.

In this dissertation, an important class of HT distributions are focused, namely regularly varying distributions, which inherently have finite tail coefficients.

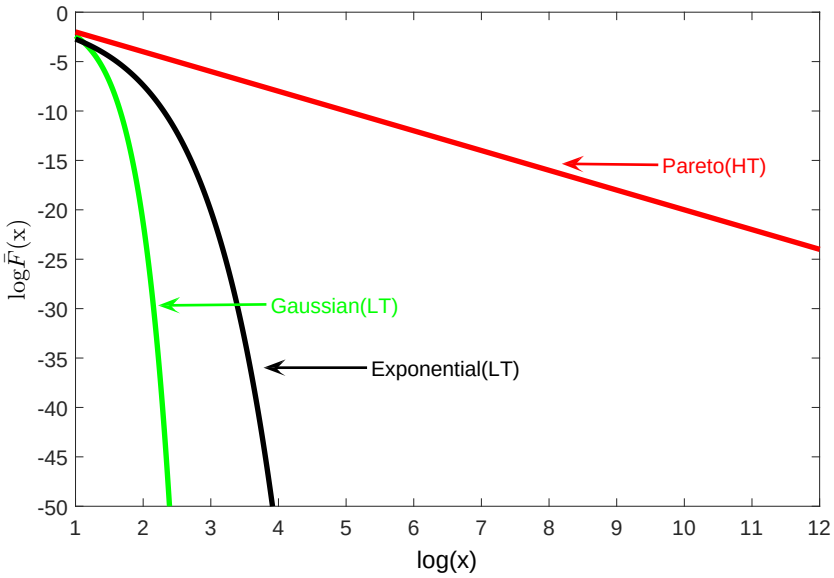

Figure 2.1: Tail distribution in log-log plot.

Definition 3. *Random variable X is called regularly varying with tail index $\beta > 0$, denoted by $X \in \mathcal{RV}(\beta)$, if*

$$\overline{F}(x) \sim x^{-\beta}\mathcal{V}(x), \tag{2.3}$$

where $\mathcal{V}(x)$ is a slowly varying function [30].

Regularly varying distributions are generalizations of Pareto/Ziph/power-law distributions, and they can effectively characterize a wide range of network attributes, such as the frame length of VBR traffic, the session duration of network users in wireless local area networks (WLANs), and the size of files at internet severs and cellular base stations. The tail index β indicates how heavy the tail distribution is, whereas smaller values of β imply heavier tails. Moreover, for an r.v. $X \in \mathcal{RV}(\beta)$, the tail coefficient κ_X of X is equal to the tail index β, which defines the maximum order of bounded moments that X can have. Specifically, if $0 < \beta < 1$, then X has an infinite mean and an infinite variance. If $1 < \beta < 2$, then X has a finite mean and an infinite variance.

CHAPTER 3

THROUGHPUT OPTIMAL SCHEDULING ALGORITHM IN THE PRESENCE OF HEAVY TAILS

This chapter presents three different scheduling algorithms: a distributed maximum weight-α scheduling algorithm, a stochastic subgradient scheduling algorithm, and the proposed time-average stochastic gradient scheduling algorithm. Their advantages and disadvantages will be analyzed.

3.1 Types of Network Stability

The next section introduces several types of network stability.

3.1.1 Strongly Stable

Definition 4. *A network user i is strongly stable if its steady-state queue length has a bounded mean, i.e.,*

$$E[Q_i(t)] < \infty. \tag{3.1}$$

And a network is strongly stable if there exists a scheduling algorithm, under which all network users are strongly stable.

The following lemma shows that strong stability is not achievable for the network user with heavy-tailed arrivals with a tail index less than three.

Lemma 1. *Let Q_i denote the steady-state queue length of user i. If user i has HT arrivals with tail index β_i, i.e., $A_i(t) \in \mathcal{RV}(\beta_i)$, then under any scheduling algorithms, Q_i necessarily follows HT distribution with the tail coefficient $\min_{j\in H} \beta_j - 1 < \kappa(Q_i) < \beta_i - 1$ [22].*

Lemma 1 and Definition 2 show that under any scheduling algorithm, if user i has an HT traffic arrival with tail index less than 2, i.e., $\beta_i < 2$, then $E[Q_i] = \infty$, and if user i has an HT traffic arrival with tail index less than 3, i.e., $\beta_i < 3$, then $E[Q_i^2] = \infty$. These

observations mean that strong stability is difficult to achieve, at least for HT users. Thus, utility-optimal scheduling with respect to strong stability may not exist in the heavy-tailed environment. To address this problem, moment stability is introduced.

3.1.2 Moment Stability

Definition 5. *A network is moment stable, if there exists a scheduling algorithm, under which the steady-state queue length of any user with light-tailed arrival has a bounded mean [22], i.e.,*

$$E[Q_i(t)] < \infty, \forall A_i(t) \in LT.$$

3.2 Distributed Maximum Weight-α Scheduling Algorithm

Xia and Wang [3] introduced DMWS-α, which is moment stable and throughput optimal.

3.2.1 System Model

There are N number of network users sharing a single fading channel with a fully connected network, and the time is slotted with a unit slot size. The term $A_i(t)$ is defined as the arrival rate, which is the number of packets that arrive at queue i during time slot t, and the mean of arrival rate is $E[A_i(t)] = a_i$. Moreover, $R_i(t)$ is used to denote the data rate of user i when it is transmitting through a fading channel during time slot t. Two types of queues are considered in the network: heavy queue and light queue.

3.2.2 Queueing Dynamics

Let $Q_i(t)$ denote the number of packets in queue i at the end of time slot t. Then, the queueing dynamic of user i can be defined as

$$Q_i(t+1) = Q_i(t) - R_i(t)H_i(t) + A_i(t) \tag{3.2}$$

where $H_i(t)$ denotes the transmission duration of user i during time slot t, and $R_i(t)$ is the transmission rate. Then, let $S_i(t)$ denote the service rate, which shows the number of

packets departing from queue i at time slot t under any scheduling policy. Thus, the queueing dynamic evolves to

$$Q_i(t+1) = Q_i(t) - S_i(t) + A_i(t). \tag{3.3}$$

Definition 6. *(**DMWS-α Algorithm**)*

***Step 1**: At the beginning of each time slot t, each user i determines the length of its carrier sensing period τ_i by independently generating an exponentially distributed random variable with the mean $\exp(w_i(t))$, i.e.,*

$$P(\tau_i > x) = \exp\Big(- exp(w_i(t)x\Big) \tag{3.4}$$

and

$$w_i(t) = \min(Q_i(t)^{\alpha_i} R_i(t), w^*) \tag{3.5}$$

where $w_i(t)$ is the queueing weight and $\alpha_i = \beta_i - 1$ if user i has HT arrivals, i.e., $A_i \in \mathcal{RV}(\beta_i)$, and $\alpha_i \geq 2$ if user i has LT arrivals, i.e., $A_i \in LT$, and w^* is the predefined maximum value of $w_i(t)$ [3].

***Step 2**: During this sensing period τ_i, if no transmissions are detected, then user i transmits its packets until the end of the time slot.*

Theorem 1. *Under the DMWS-α algorithm, the network is moment stable if*

$$\sum_{i=1}^{N} \frac{a_i}{r_i} \leq (1-\epsilon)\left(1 - \frac{N}{\exp(\epsilon w^*)}\right) \tag{3.6}$$

where r_i is the mean of the data rate, and ϵ is an arbitrary small value [3].

Proof. Theorem 1 indicates that the DMWS-α algorithm is throughput optimal, and the proof details can be found in the work of Xia and Wang [3]. □

However, even though the DMWS-α is moment stable and throughput optimal, it needs to determine the α value for different types of traffic arrivals, which is not easy to implement in a real environment.

3.3 Stochastic Network Utility Maximization in the Presence of Heavy Tails

In the previous section, the DMWS-α algorithm that is moment stable and throughput optimal is introduced, but it is difficult to find an appropriate α value. Hence, another stochastic network utility maximization algorithm is proposed.

3.3.1 System Model

Consider a single-hop network with N users, as shown in Figure 3.1. This network model is used to study the utility-optimal downlink/uplink scheduling for cellular or Wi-Fi systems. Time is slotted with a unit slot size. The channel diversity is captured by assigning user i with an unequal data rate r_i. Queue i is associated with each user i. Let $A_i(t)$ denote the number of packets that arrive at queue i during time slot t with an average traffic rate or traffic intensity of $E[A_i(t)] = a_i$. Two types of queues are considered in the network: heavy queue and light queue.

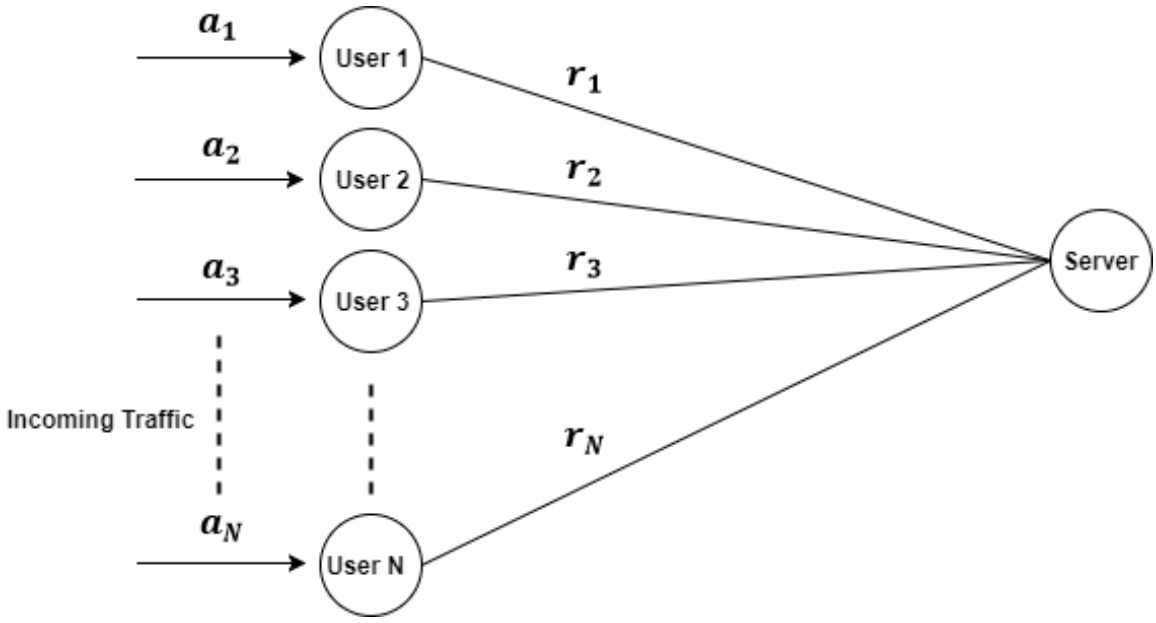

Figure 3.1: Single hop network model.

Definition 7. *(**Heavy Queue**) Queue i belongs to the heavy queue, denoted by $i \in \mathcal{H}$, if it sometimes has heavy-tailed arrivals, i.e., $A_i(t) \in \mathcal{RV}(\beta_i)$ with tail index β_i.*

Definition 8. *(**Light Queue**) Queue i belongs to the light queue, denoted by $i \in L$, if it always has light-tailed traffic arrivals, i.e., $A_i(t) \in LT$.*

Let $Q_i(t)$ denote the number of packets in queue i by the end of time slot t. Then, the queueing dynamics for user i can be represented by

$$Q_i(t+1) = Q_i(t) - W_i(t)r_i + A_i(t) \tag{3.7}$$

where $W_i(t)$ is the transmission probability of user i during time slot t with mean $E[W_i(t)] = w_i$. By defining $S_i(t)$ as the number of served packets departing from queue i at time slot t under any scheduling policy, the queueing dynamics in equation (3.7) can be rewritten by

$$Q_i(t+1) = Q_i(t) - S_i(t) + A_i(t). \tag{3.8}$$

3.3.2 Network Stability Analysis for Stochastic Subgradient Scheduling Algorithm

In this section, the utility-optimal scheduling problem is formulated as a concave optimization framework and utilizes the classic subgradient descent algorithm to solve such problem. Then, the classic subgradient descent algorithm that is not moment stable in the presence of HT traffic is proven. In particular, by defining the utility function $U_i(\cdot)$ as a concave and monotone increasing function on R_+, the utility-optimal scheduling problem is formulated as follows:

$$\begin{aligned} Maximize \quad & \sum_{i=1}^{N} U_i(w_i r_i) \\ s.t \quad & a_i \leq w_i r_i - \psi_i \qquad i = 1, 2, ...N \\ & \sum_{i \in N} w_i = 1 \\ & w_i \geq 0 \qquad\qquad i = 1, 2, ...N \\ & E[Q_i(t)] < \infty \qquad \forall A_i(t) \in LT. \end{aligned} \tag{3.9}$$

where $\psi_i \geq 0$ is a small constant. The first two constraints are flow-conservation constraints, which enforce the average arrival traffic rate a_i to be less than the average channel service

rate $w_i r_i$, and the total channel service rates to be within the network capacity region. The last constraint guarantees queue stability of LT queues in the presence of HT traffic.

To solve equation (3.9), the Lagrangian multipliers λ_i and the Lagrange function are defined as

$$\mathcal{L}(w, \lambda) = \sum_{i=1}^{N} U_i(w_i r_i) + \sum_{i=1}^{N} \lambda_i (w_i r_i - a_i - \psi_i) \tag{3.10}$$

where vectors $w \in [0,1]^N$ and $\lambda \in R^N$. Then, the primal Lagrangian maximizers are defined as follows:

$$\begin{aligned} W^*(t) &= \arg\max_{\{w_i\}_{i \le N}} \mathcal{L}(w, \lambda) \\ &= \arg\max_{\{w_i\}_{i \le N}} \sum_{i=1}^{N} U_i(w_i r_i) + \sum_{i=1}^{N} \lambda_i (w_i r_i - a_i - \psi_i) \\ &= \arg\max_{\{w_i\}_{i \le N}} \sum_{i=1}^{N} U_i(w_i r_i) + \sum_{i=1}^{N} \lambda_i w_i r_i \end{aligned} \tag{3.11}$$

where $W^*(t) = (W_1^*(t), W_2^*(t), ..., W_N^*(t)) \in [0,1]^N$, and the last equation holds because the arrival rate a_i does not affect the primal Lagrangian maximizers $W^*(t)$. Let $W_i(t)$ be a binary value, where $W_i(t) = 1$ indicates that user i is transmitting at time slot i. Then, user i with $W_i^*(t) = 1$ will be scheduled to transmit at time slot t. Next, define the stochastic subgradient $g_i(t)$ by taking the derivative of equation (3.11) with respect to λ_i as

$$g_i(t) = W_i^*(t) r_i - A_i(t) - \psi_i(t). \tag{3.12}$$

Then, update the dual variables along the stochastic subgradient direction as

$$\begin{aligned} \lambda_i(t+1) &= \lambda_i(t) - \alpha(t) g_i(t) \\ &= \lambda_i(t) - \alpha(t) \Big(W_i^*(t) r_i - A_i(t) - \psi_i(t) \Big) \end{aligned} \tag{3.13}$$

where $\alpha(t)$ is the scale step size for all users, $i = 1, ..., N$. Moreover, to guarantee $\lambda_i(t+1) \ge 0$, let $\psi_i(t) = 0$, if $\lambda_i(t) - \alpha(t) W_i^*(t) r_i + \alpha(t) A_i(t) > 0$. Otherwise, let $\psi_i(t) = \alpha(t) W_i^*(t) r_i - \alpha(t) A_i(t) - \lambda_i(t)$ if $\lambda_i(t) - \alpha(t) W_i^*(t) r_i + \alpha(t) A_i(t) < 0$. The above stochastic subgradient

algorithm is summarized in Algorithm 1. By letting $\lambda_i(0) = Q_i(0)$ for all N users, it can be seen that the queue-length evolution process defined in equation (3.7) corresponds to the dual-variable update process. It is easy to prove that Algorithm 1 is utility optimal. Next, it is shown that Algorithm 1 is not moment stable in the presence of HT traffic.

Algorithm 1 Stochastic Subgradient Scheduling Algorithm

Observe $Q_i(0)$. Initialize $\lambda_i(0) = Q_i(0)$ for all users
for $t = 0, 1, 2, ...$ **do**
  for $i = 1, 2, 3, ...$ **do**
    Find transmission probability $W_i^*(t)$
    $W_i^*(t) = \arg\max_{w_i} \sum_{i=1}^{N} U_i(w_i r_i) + \sum_{i=1}^{N} \lambda_i w_i r_i$
    Transmit $W_i^*(t) r_i$ packets for node i
  end for
  Update dual variable $\lambda_i(t+1) = \left[\lambda_i(t) - (W_i^*(t) r_i - A_i(t))\right]^+$
end for

Theorem 2 (**Network Instability**). *Under the stochastic subgradient scheduleing algorithm, if there exists a heavy queue with the tail index smaller than two, i.e.,*

$$\min_{i \in \mathcal{N}} \kappa(A_i(t)) < 2, \tag{3.14}$$

then the queue length of the light queue is necessarily of an unbounded mean (i.e., $E[Q_i(t)] = \infty \ \forall A_i(t) \in LT$).

Proof. Assume that the queueing system is in the steady state. Then, the queue length process is in a positive recurrent regenerative process. Define T as the time interval between two consecutive instances when all queues are empty. Then, $E[T] < \infty$. Moreover, assume that at time slot zero, queue h, having the HT arrival with the smallest tail index, i.e., $\kappa(A_h(t)) = \min_{i \in \mathcal{N}} \kappa(A_i(t))$, receives a file of size b number of packets, and all other queues receive no traffic. Let T_b be the first time slot when the weight of queue h becomes less than or equal to the weight of one of the queues $l \in \mathcal{N}$ and $l \neq h$, i.e.,

$$T_b := \min\{t > 0 | W_l^*(t) \geq W_h^*(t), l \in \mathcal{N}, l \neq h\}. \tag{3.15}$$

Under Algorithm 1, queue l gets no services but receives new arrival $A_l(t)$ until time T_b, and queue h is served at r_i rate while the initial traffic of queue h is served until time slot T_b, during which the HT queue h keeps transmitting data. This implies that $\forall l \neq h$

$$Q_l(T_b) = \sum_{t=0}^{T_b-1} A_l(t). \tag{3.16}$$

According to the strong law of large numbers (SLLN), $\sum_{t=0}^{T_b-1} A_l(t) \geq a_l T_b - \delta_b$, with probability 1. Thus, according to Algorithm 1, we have the following relationship, which shows that the first time slot of transmission probability of LT queue is larger than the transmission probability of HT queue, as

$$U(r_l) + (a_l T_b - \delta_b) r_l \geq U(r_h) + (b - r_h T_b) r_h \tag{3.17}$$

which implies that there exists a constant c such that

$$\begin{aligned} T_b &\geq \frac{b r_h + U(r_h) - U(r_l) + \delta_b r_l}{a_l r_l + r_h^2} \\ &\geq \frac{b r_h}{a_l r_l + r_h^2} = cb \end{aligned} \tag{3.18}$$

where $c = \frac{r_h}{a_l r_l + r_h^2}$ because $U(r_h) - U(r_l) > 0$ and $\delta_b r_l > 0$. In addition, by the property of the regenerative process with cycle length T, we have

$$\Pr(Q_l > \frac{cba_l}{2}) = \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{I}_{\{Q_l(\tau) > \frac{cba_l}{2}\}} = \frac{E[\sum_{t=0}^{T} \mathbb{I}_{\{Q_l(t) > \frac{cba_l}{2}\}}]}{E[T]}. \tag{3.19}$$

Then, we have

$$\begin{aligned} & E[\sum_{t=0}^{T} \mathbb{I}_{\{Q_l(t) > \frac{cba_l}{2}\}}] \\ & \geq E[\mathbb{I}(A_h(0) > b) \sum_{t=0}^{T} \mathbb{I}(Q_l(t) > \frac{cba_l}{2})] \\ & \geq \Pr(A_h(0) > b) \sum_{t=\frac{T_b}{2}}^{T_b} \Pr(Q_l(t) > \frac{cba_l}{2}). \end{aligned} \tag{3.20}$$

By the queueing dynamic in equation (3.8) and $Q_l(0) = 0$, we have $Q_l(t) = \sum_{\tau=0}^{t-1}[A_l(\tau) - S_l(\tau)\mathbb{I}_{\{Q_l(\tau)>0\}}]$, which implies that

$$\lim_{b\to\infty} \frac{1}{b} \sum_{t=\frac{T_b}{2}}^{T_b} \Pr(Q_l(t) > \frac{cba_l}{2}) \\ \geq \lim_{b\to\infty} \frac{1}{b} \sum_{t=\frac{cb}{2}}^{cb} \Pr\big(\sum_{\tau=0}^{t-1} A_l(\tau) > \frac{cba_l}{2}\big) = \frac{c}{2}. \tag{3.21}$$

The last equality in equation (3.21) holds, due to the fact that the heavy-tailed queue h occupies the entire channel service during the time interval $0 \leq t \leq cb/2$ so that $\Pr\big(\sum_{\tau=0}^{t-1} A_l(\tau) > \frac{cba_l}{2}\big) = 1$ when $cb/2 \leq t \leq cb$ always holds. Then, we have

$$Pr(Q_l > \frac{cba_l}{2}) \geq \frac{\Pr(A_h(0) > b) \sum_{t=\frac{T_b}{2}}^{T_b} \Pr(Q_l(t) > \frac{cba_l}{2})}{E[T]} \tag{3.22}$$

according to Nair et al. [34] and combining equation (3.19) with equations. (3.20)-(3.21), it follows from the assumption that $\kappa(A_h(t)) = \min_{i\in N} \kappa(A_i(t)) < 2$ that

$$\lim_{b\to\infty} \frac{\log[\Pr(Q_l > \frac{cba_l}{2})]}{\log[\frac{cba_l}{2}]} \geq -\min_{i\in N} \kappa(A_i(t)) + 1 \geq -1$$

which, by applying the moment theorem in the work of Daley and Goldie [33] implies that $E[Q_l]$ is infinite. □

3.3.3 Time-Average Stochastic Gradient Scheduling Algorithm

In this section, a time-average stochastic gradient scheduling algorithm, which can achieve utility optimality and queue stability simultaneously even under a heavy-tailed environment, is proposed. First, an object Lagrange function is redefined as

$$\bar{\mathcal{L}}(w, \lambda, \mu, \upsilon) = \sum_{i=1}^{N} U_i(w_i r_i) + \sum_{i=1}^{N} \lambda_i(w_i r_i - a_i - \psi_i) + \mu(1 - \sum_{i=1}^{N} w_i) + \sum_{i=1}^{N} \upsilon_i w_i \tag{3.23}$$

where λ_i, μ, and υ_i are the Lagrange multiplier; $w_i r_i - a_i - \psi_i$ comes from the first inequality constraint in equation (3.9); $\sum_{i=1}^{N} w_i = 1$ comes from the second constraint in equation (3.9);

and $w_i \geq 0$ comes from the third constraint in equation (3.9). Then, we have a Lagrange dual function

$$D(\lambda, \mu, \upsilon) = \sup_{w} \bar{\mathcal{L}}(w, \lambda, \mu, \upsilon) \tag{3.24}$$

by which, we can further define the dual problem as

$$\min_{\lambda>0} D(\lambda, \mu, \upsilon) = \min_{\lambda>0} \sup_{w} \bar{\mathcal{L}}(w, \lambda, \mu, \upsilon) \tag{3.25}$$

since the primal problem in equation (3.9) is convex, and the constraint qualification satisfies Slater's condition [35, Section 5.2]. Thus, strong duality exists, and there is no gap between the primal and dual functions so that we can find the maximum value of the primal function by minimizing the dual function using a descent algorithm. To find the minimal values, first we should compute Lagrangian maximizers. Thus, taking the first derivative of the Lagrange function in equation (3.23) with respect to w_i, we have

$$\frac{\partial \hat{\mathcal{L}}(w, \lambda, \mu, \upsilon)}{\partial w_i} = r_i U_i^{'}(w_i r_i) + \lambda_i r_i - \mu + \upsilon_i. \tag{3.26}$$

Let $\frac{\partial \hat{\mathcal{L}}(w,\lambda,\mu,\upsilon)}{\partial w_i} = 0$. Then, we obtain

$$r_i U_i^{'}(w_i r_i) + \lambda_i r_i - \mu + \upsilon_i = 0. \tag{3.27}$$

Then, we apply the complementary slackness condition

$$\begin{aligned} \upsilon_i w_i = 0 \quad and \quad \upsilon_i \geq 0, \\ \mu(1 - \sum_{i=1}^{N} w_i) = 0 \quad and \quad \mu \geq 0. \end{aligned} \tag{3.28}$$

From equation (3.27), we can determine υ_i as follows:

$$\upsilon_i = \mu - \lambda_i r_i - r_i U_i^{'}(w_i r_i). \tag{3.29}$$

Since $\upsilon_i \geq 0$, the following relationship exists:

$$U_i^{'}(w_i r_i) \leq \frac{\mu - \lambda_i r_i}{r_i}. \tag{3.30}$$

Since $U_i(\cdot)$ is a concave increasing, and $U_i^{'}(\cdot)$ is monotone, decreasing, and positive, we get

$$w_i r_i \geq U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}). \tag{3.31}$$

Hence,

$$w_i \geq \frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}). \tag{3.32}$$

If $\frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}) > 0$, then $w_i > 0$. Therefore, to meet the complementary slackness condition in equation (3.28), we conclude that $v_i = 0$. Then, according to equation (3.27), we have

$$w_i = \frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}). \tag{3.33}$$

If $\frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}) \leq 0$, then from equation (3.31), $w_i \leq 0$. Thus, we can conclude that $w_i = 0$ because w_i should be non-negative. Then, we get

$$w_i = \begin{cases} \frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}), & \frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}) > 0 \\ 0 & \frac{1}{r_i} U_i^{'-1}(\frac{\mu - \lambda_i r_i}{r_i}) \leq 0. \end{cases} \tag{3.34}$$

Therefore, using a simple expression, we have

$$w_i = \left[\frac{1}{r_i} U_i^{'-1}\big(\frac{\mu - \lambda_i r_i}{r_i}\big) \right]^+ \tag{3.35}$$

$$s.t. \quad \sum_{i=1}^{N} w_i = 1. \tag{3.36}$$

Taking the summation on both sides of equation (3.35) and according to the second equality constraint in equation (3.9), we have

$$\sum_{i=1}^{N} \left[\frac{1}{r_i} U_i^{'-1}\big(\frac{\mu - \lambda_i r_i}{r_i}\big) \right]^+ = 1. \tag{3.37}$$

Therefore, according to the second complementary slackness condition in equation (3.28), we know the Lagrange variable $\mu > 0$. Then, we can calculate μ using equation (3.37). Therefore, we can obtain the transmission probability $W_i(\lambda, \mu, t)$ by

$$W_i(t) = \frac{1}{r_i} U_i^{'-1}\big(\frac{\mu(t) - \lambda_i(t) r_i}{r_i}\big). \tag{3.38}$$

From $W_i(t)$, we can obtain the stochastic gradient $g_i(t)$ of the dual function in equation (3.24), i.e.,

$$g_i(t) = W_i(t)r_i - A_i(t) - \psi_i(t). \tag{3.39}$$

We further calculate the time average of transmission probability using

$$\overline{W}_i(t) = \frac{1}{t}\sum_{j=1}^{t} W_i(j) = \frac{t-1}{t}\overline{W}_i(t-1) + \frac{1}{t}W_i(t), \quad t \geq 1 \tag{3.40}$$

and update our queue function using

$$Q_i(t+1) = \left[Q_i(t) - \overline{W}_i(t)r_i + A_i(t)\right]^+. \tag{3.41}$$

Moreover, we can update multipliers along the stochastic subgradient descent direction using

$$\begin{aligned}\lambda_i(t+1) &= \lambda_i(t) - \alpha(t)g_i(t) \\ &= \lambda_i(t) - \alpha(t)\Big(W_i(t)r_i - A_i(t) - \psi_i(t)\Big)\end{aligned} \tag{3.42}$$

where $\alpha(t)$ is the scale step size for all users i. Moreover, to guarantee $\lambda_i(t+1) \geq 0$, we let $\psi_i(t) = 0$, if $\lambda_i(t) - \alpha(t)W_i^*(t)r_i + \alpha(t)A_i(t) > 0$. Otherwise, $\psi_i(t) = \alpha(t)W_i^*(t)r_i - \alpha(t)A_i(t) - \lambda_i(t)$ if $\lambda_i(t) - \alpha(t)W_i^*(t)r_i + \alpha(t)A_i(t) < 0$. Note that $\overline{W}_i(t)$ is used in the queue length update only in equation (3.41) and not in the stochastic gradient descent in equation (3.42). The proposed time-average stochastic gradient algorithm is summarized in Algorithm 2. Initially, $\lambda_i(0) = Q_i(0)$ is set for all users. Then, we calculate the Lagrange variable μ in step 4. Based on μ, we can compute the time-average transmission probability for each user in step 9, which leads to the transmission rate of $\overline{W}_i(t)r_i$. Then, each node can update its queue function in step 12 and the dual-variable in step 13. One of the key features of this algorithm is that the queue-length update process in equation (3.41) is decoupled from the dual-variable update process in equation (3.42). We can exploit the time-average Lagrange maximizer to regulate the transmission rates, and we can use the instantaneous Lagrange maximizer to control the dual variables. We will show that such a unique feature can guarantee the moment stability in the presence of HT traffic.

Algorithm 2 Time-Average Stochastic Gradient Scheduling Algorithm

Observe $Q_i(0)$. Initialize $\lambda_i(0) = Q_i(0)$ for all users
for $t = 0, 1, 2, ...$ **do**
  Compute $\mu(t) > 0$ such that
    $\sum_{i=1}^{N} \left[\frac{1}{r_i} U_i'^{-1}(\frac{\mu(t)-\lambda_i(t)r_i}{r_i}) \right]^+ = 1$
  for $i = 1, 2, 3, ...$ **do**
    Compute Lagrange maximizer $W_i(t)$, where
      $W_i(t) = \frac{1}{r_i} U_i'^{-1}(\frac{\mu(t)-\lambda_i(t)r_i}{r_i})$
    Compute time-average transmission probability $\overline{W}_i(t)$, where
      $\overline{W}_i(t) = \begin{cases} 0, & t = 0 \\ \frac{t-1}{t}\overline{W}_i(t-1) + \frac{1}{t}W_i(t), & t \geq 1 \end{cases}$
    Transmit packets at rate $\overline{W}_i(t)r_i$
  end for
  Update queue function $Q_i(t+1) = [Q_i(t) - \overline{W}_i(t)r_i + A_i(t))]^+$
  Update dual variable $\lambda_i(t+1) = \left[\lambda_i(t) - \alpha(t)(W_i(t)r_i - A_i(t)) \right]^+$
end for

3.3.4 Log Utility Function

Consider the proportional fairness metric inthe work of Tychogiorgos et al. [36]. Then, we can further explain our proposed algorithm using a log utility function. Hence, let $U(w_i r_i) = \log(w_i r_i)$. From equation (3.23) we obtain

$$\bar{\mathcal{L}}(w, \lambda, \mu) = \sum_{i=1}^{N} \log(w_i r_i) + \sum_{i=1}^{N} \lambda_i (w_i r_i - a_i - \psi_i) + \mu(1 - \sum_{i=1}^{N} w_i). \tag{3.43}$$

According to equation (3.34), we have transmission probability w_i as

$$w_i = \begin{cases} \frac{1}{\mu - \lambda_i r_i}, & \frac{1}{\mu-\lambda_i r_i} > 0 \\ 0 & \frac{1}{\mu-\lambda_i r_i} < 0 \quad or \quad \mu = \lambda_i r_i. \end{cases} \tag{3.44}$$

Then using the simple expression yields

$$w_i = \left[\frac{1}{\mu - \lambda_i r_i} \right]^+ \tag{3.45}$$

since the summation of all users' transmission probability should be equal to 1. Therefore,

$$\sum_{i=1}^{N} \left[\frac{1}{\mu - \lambda_i r_i} \right]^+ = 1. \tag{3.46}$$

Algorithm 3 Searching Algorithm for μ Value

for $t = 0, 1, 2, ...$ **do**
  for $i = 1, 2, 3, ...$ **do**
    Initial $\mathcal{K} = [\lambda_1(t)r_1, \lambda_2(t)r_2, \lambda_3(t)r_3, \cdots\cdots]$.
  end for
  Find maximum value $\mathcal{K}^*$
  Let $\mu = \mathcal{K}^* + 1$
  while 1 **do**
    if $\sum_{i=1}^{N} \frac{1}{\mu - \lambda_i(t)r_i} > 1$ **then**
      $\mu = \mu + \eta$
    else
      break
    end if
  end while
  Return μ
end for

From equation (3.46), we can find the μ value using Algorithm 3. Here η is a positive searching step size. Moreover, when the searching step size is small enough, we can obtain a more accurate μ value so that equation (3.46) will approach 1. More specifically, if the summation in equations (3.37) or (3.46) is slightly smaller than but very close to 1, then Algorithms 2 and 3 stop in practice.

3.3.5 Utility Optimal Analysis

In this section, the network utility of our proposed TA-SGSA algorithm will be analyzed. Since our algorithm employs an iterative method, it generates a sequence of improving approximate solutions for a class of problems. Hence, to prove that TA-SGSA is utility optimal we prove that the utility converges to an optimal value as time proceeds. However, according to Boyd and Mutapcic [37], the classical convergence analysis is based on the Euclidean distance to the optimal set, which no longer works in our case because it relies on the assumption that the arrivals have bounded variance. Therefore, to analyze the utility convergence for TA-SGSA, we adopt the ordinary differential equation method from Borkar [38, Section 2], which treats the discrete stochastic approximation scheme as a

discretization version of the ODE with a HT noise, and shows that the HT noise becomes asymptotically negligible in the εth mean. Then, we have the following theorem:

Theorem 3. *The time-average stochastic gradient scheduling algorithm is utility optimal.*

Proof. To prove that our time-average stochastic gradient scheduling algorithm is utility optimal, an ordinary differential equation approach by Borkar [38] to explore convergence and stability properties of our algorithm is used.

From equation (3.42), we have vector $\lambda(t+1)$ as

$$\lambda(t+1) = \lambda(t) - \alpha(t)(W(t)r - A(t)). \tag{3.47}$$

Then, we do the simple computation to obtain

$$\begin{aligned}\lambda(t+1) &= \lambda(t) - \alpha(t)(W(t)r - a - A(t) + a) \\ &= \lambda(t) + \alpha(t)(a - W(t)r) + \alpha(t)(A(t) - a) \\ &= \lambda(t) + \alpha(t)(h(t) + \hat{A}(t))\end{aligned} \tag{3.48}$$

where $h(t) := a - W(t)r$, $W(t) \in [0\ 1]$, $E[A(t)] = a$, and r is a constant. Thus, h is in a strong form of uniform continuity, i.e., Lipschitz [39, Section 12.3], and $\hat{A}(t)$ is a heavy-tailed arrival noise. Since this is a stochastic gradient algorithm, the associate ODE is $\dot{\lambda}(t) = -g(\lambda(t))$. Therefore, we obtain the following relationship

$$\dot{\lambda}(t) = h(\lambda(t)) \tag{3.49}$$

where $\dot{\lambda}(t)$ is the first derivative of independent variable. Thus, the equation (3.49) reflects the relationship between the independent variable and its derivative. Moreover, after we obtain the solution of equation (3.49), we can find the trajectory of the independent variable.

The idea here is to treat the discrete stochastic approximation scheme in equation (3.47) as a discretization version of the ODE in equation (3.49) with a heavy-tailed noise [40]. Then, we show under certain hypotheses that these discretization and noise errors become asymptotically negligible in the εth mean for $1 < \varepsilon < \alpha$, where α is the tail index of the HT

noise. Finally, we show that the stochastic approximation has the same asymptotic limit with the ODE in the εth mean. The proving steps are as follows:

First, we define $j(t) = \sum_{k=0}^{t-1} \alpha(k)$. Therefore, the continuous piecewise linear interpolated version of $\lambda(t)$ can be defined by $\bar{\lambda}(j(t)) = \lambda(t)$. Letting $\lambda^t(j)$, $j \geq j(t)$, is the continuous ODE trajectory with $\lambda^t(j(t)) = \bar{\lambda}(j(t)) := \lambda(t)$. In other words, both the discrete and continuous versions have the same value at $j(t)$. Moreover, we define $b = \inf\{k : j(k) > j(t) + T\}$ and $T = j(b) - j(t)$. Then, we want to find the supreme of the difference norm $\sup_{t \leq k \leq b} ||\bar{\lambda}(j(k)) - \lambda^t(j(k))||$ when $t \leq k \leq b$. Therefore, we rewrite $\bar{\lambda}(j(k))$ and $\lambda^t(j(k))$, respectively, as

$$\bar{\lambda}(j(k)) = \bar{\lambda}(j(t)) + \sum_{i=t}^{k-1} \alpha(i) h(\bar{\lambda}(j(i)) + \sum_{i=t}^{k-1} \alpha(i) \hat{A}(i) \tag{3.50}$$

and

$$\begin{aligned}
\lambda^t(j(k)) &= \bar{\lambda}(j(t)) + \int_{j(t)}^{j(k)} h(\lambda^t(v)) dv \\
&= \bar{\lambda}(j(t)) + \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} h(\lambda^t(v)) dv \\
&= \bar{\lambda}(j(t)) + \sum_{i=t}^{k-1} \left\{ \alpha(i) h(\lambda^t(j(i))) + \int_{j(i)}^{j(i+1)} \big(h(\lambda^t(v)) - h(\lambda^t(j(i))) \big) dv \right\}
\end{aligned} \tag{3.51}$$

where $\int_{j(i)}^{j(i+1)} dv = \alpha(i)$. Then, we subtract equation (3.51) from equation (3.50), obtaining

$$\begin{aligned}
&\sup_{t \leq k \leq b} ||\bar{\lambda}(j(k)) - \lambda^t(j(k))|| \\
&\leq \sup_{t \leq k \leq b} \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h(\lambda^t(v)) - h(\lambda^t(j(i)))|| dv + \sup_{t \leq k \leq b} || \sum_{i=t}^{k-1} \alpha(i) \hat{A}(i) ||.
\end{aligned} \tag{3.52}$$

Let

$$T_I = \sup_{t \leq k \leq b} \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h(\lambda^t(v)) - h(\lambda^t(j(i)))|| dv \tag{3.53}$$

and

$$T_{II} = \sup_{t \leq k \leq b} || \sum_{i=t}^{k-1} \alpha(i) \hat{A}(i) ||. \tag{3.54}$$

For term T_I, we define $C_0 = \lambda^t(j(t)) = \bar{\lambda}(j(t)) = \lambda(t) \leq \sup_t ||\lambda(t)||$, which is finite. Then,

$$\begin{aligned}\lambda^t(j) &= \overline{\lambda}(j(t)) + \int_{j(t)}^{j(b)} h(\lambda^t(v))dv \\ &= \lambda(t) + \int_{j(t)}^{j(b)} h(\lambda^t(v))dv \end{aligned} \tag{3.55}$$

where $j \in [j(t), j(b)]$ and $\overline{\lambda}(j(t)) = \lambda(t)$. Taking the norm on both sides of equation (3.55) yields

$$||\lambda^t(j)|| \leq ||\lambda(t)|| + \int_{j(t)}^{j(b)} ||h(\lambda^t(v))||dv. \tag{3.56}$$

Since h is Lipschitz and grows linearly, we have $||h(x) - h(0)|| \leq L||x||$ and $||h(x)|| \leq ||h(0)|| + L||x||$, where $L > 0$ denotes the Lipschitz constant. Then, we have

$$||h(\lambda^t(v))|| \leq ||h(0)|| + L||\lambda^t(v)||. \tag{3.57}$$

Thus, equation (3.56) can be rewritten as

$$\begin{aligned}||\lambda^t(j)|| &\leq ||\lambda(t)|| + \int_{j(t)}^{j(b)} ||h(0) + L(\lambda^t(v))||dv \\ &\leq ||\lambda(t)|| + T||h(0)|| + L\int_{j(t)}^{j(b)} ||h(\lambda^t(v))||dv \\ &= (C_0 + T||h(0)||) + L\int_{j(t)}^{j(b)} ||h(\lambda^t(v))||dv. \end{aligned} \tag{3.58}$$

Moreover, using Gronwall's inequality from the work of Coddington and Levinson [41, Section 1.6], we obtain

$$\begin{aligned}||\lambda^t(j)|| &\leq (C_0 + T||h(0)||)e^{L\int_{j(t)}^{j(b)} ||h(\lambda^t(v))||dv} \\ &\leq (C_0 + T||h(0)||)e^{LT} \end{aligned} \tag{3.59}$$

which holds since we let $h(\cdot) = 1$, $\int_{j(t)}^{j(b)} ||h(\lambda^t(v))||dv = j(b) - j(t) = T$. Then, according to the Lipschitz continuous condition,

$$\begin{aligned}||h(\lambda^t(j))|| &\leq ||h(0)|| + L||\lambda^t(j)|| \\ &\leq ||h(0)|| + L(C_0 + T||h(0)||)e^{LT} = C. \end{aligned} \tag{3.60}$$

When $j \in [j(i), j(i+1)]$, this can be simplified as

$$||\lambda^t(j) - \lambda^t(j(i))|| \leq \int_{j(i)}^{j(i+1)} ||h(\lambda^t(v))||dv \leq C\alpha(i). \tag{3.61}$$

Since h has Lipschitz continuity,

$$\begin{aligned} &\int_{j(i)}^{j(i+1)} ||h(\lambda^t(v)) - h(\lambda^t(j(i)))||dv \\ &\leq \int_{j(i)}^{j(i+1)} L||\lambda^t(v) - \lambda^t(j(i))||dv \leq CL(\alpha(i))^2. \end{aligned} \tag{3.62}$$

Using the summation,

$$\begin{aligned} &\sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h(\lambda^t(v)) - h(\lambda^t(j(i))||dv \\ &\leq \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} L||\lambda^t(v) - \lambda^t(j(i))||dv \leq \sum_{i=t}^{k-1} CL(\alpha(i))^2. \end{aligned} \tag{3.63}$$

Taking the limit on both sides,

$$\lim_{t\to\infty} \sup_{t\leq k\leq b} \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h(\lambda^t(v)) - h(\lambda^t(j(i)))|| \leq \lim_{t\to\infty} \sum_{i=t}^{k-1} CL(\alpha(i))^2 = 0 \tag{3.64}$$

since $\lim_{t\to\infty} \sum_{i=t}^{k-1} (\alpha(i))^2 = 0$.

For T_{II}, according to [42],

$$P(\sup_{t\leq k\leq b} ||\sum_{i=t}^{k-1} \alpha(i)\hat{A}(i)|| \geq x) \leq \frac{K\big(\sum_{i=t}^{b} (\alpha(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{\beta}{\beta+1}}}{x^\beta} \tag{3.65}$$

for $x > K\big(\sum_{i=t}^{b} (\alpha(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{1}{\beta+1}}$, where β is larger than 1. We define

$$\begin{aligned} \mu(t) &:= K\big(\sum_{i=t}^{b} (\alpha(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{1}{\beta+1}} \\ &= K\big(\sum_{i=t}^{b} \alpha(i)^{\frac{\beta^2-1}{\beta}} \alpha(i)\big)^{\frac{1}{\beta+1}} \\ &= K\big(\alpha(t)^{\frac{\beta^2-1}{\beta}} \alpha(t) + \alpha(t+1)^{\frac{\beta^2-1}{\beta}} \alpha(t+1), ..., +\alpha(b)^{\frac{\beta^2-1}{\beta}} \alpha(b)\big)^{\frac{1}{\beta+1}} \\ &\leq K\big((T+1)\alpha(t)^{\frac{\beta^2-1}{\beta}} \alpha(t)\big)^{\frac{1}{\beta+1}} \\ &\leq K(T+1)^{\frac{1}{\beta+1}} \alpha(t)^{\frac{\beta-1}{\beta}} \end{aligned} \tag{3.66}$$

When $t \to \infty, \mu(t) \to 0$. Furthermore, using a method similar to the one by Anantharam and Borkar [40], for $1 < \varepsilon < \beta$, we get

$$
\begin{aligned}
&E[\sup_{t\le k\le b} ||\sum_{i=t}^{k-1} \alpha(i)\hat{A}(i)||^{\varepsilon}] \\
&\le K\int_0^{\infty} x^{\varepsilon-1} P\big(\sup_{t\le k\le b} ||\sum_{i=t}^{k-1} \alpha(i)\hat{A}(i)|| \ge x\big)dx \\
&= K\int_0^{\mu(t)} x^{\varepsilon-1} P\big(\sup_{t\le k\le b} ||\sum_{i=t}^{k-1} \alpha(i)\hat{A}(i)|| \ge x\big)dx + K\int_{\mu(t)}^{\infty} x^{\varepsilon-1} P\big(\sup_{t\le k\le b} ||\sum_{i=t}^{k-1} \alpha(i)\hat{A}(i)|| \ge x\big)dx \\
&\le K\mu(t)^{\varepsilon} + K\int_{\mu(t)}^{\infty} x^{\varepsilon-1}\big(\frac{\mu(t)^{\beta}}{x^{\beta}}\big)dx \\
&= K\mu(t)^{\varepsilon} + K\mu(t)^{\beta}\int_{\mu(t)}^{\infty} x^{\varepsilon-\beta-1}dx \\
&= K\mu(t)^{\varepsilon} - \frac{1}{\varepsilon-\beta-1}K\mu(t)^{\varepsilon} \\
&= \hat{K}\mu(t)^{\varepsilon}
\end{aligned} \tag{3.67}
$$

where $\hat{K} = K(1 - \frac{1}{\varepsilon-\beta-1})$. In addition, when $t \to \infty$, $T_{II} \to 0$. Then, combining equations (3.64) and (3.67) yields

$$\lim_{t\to\infty} \sup_{t\le k\le b} ||\bar{\lambda}(j(k)) - \lambda^t(j(k))|| = 0. \tag{3.68}$$

From the linear interpolation error [38, Section 2.1], if $j(k) \le j \le j(k+1)$,

$$\bar{\lambda}(j) = \kappa\bar{\lambda}(j(k)) + (1-\kappa)\bar{\lambda}(j(k+1)) \tag{3.69}$$

for some $\kappa \in [0,1]$. Thus,

$$||\bar{\lambda}(j) - \lambda^t(j)|| = ||\kappa(\bar{\lambda}(j) - \lambda^t(j(k))) + (1-\kappa)(\bar{\lambda}(j) - \lambda^t(j(k+1)))||. \tag{3.70}$$

Then the linear interpolation error can be upper bounded as

$$
\begin{aligned}
&||\bar{\lambda}(j) - \lambda^t(j)|| \\
&\le \kappa||\bar{\lambda}(j(k)) - \lambda^t(j(k))|| + (1-\kappa)||\bar{\lambda}(j(k+1)) - \lambda^t(j(k+1))|| \\
&\qquad + \kappa\int_{j(k)}^{j} ||h(\lambda^t(v))||dv + (1-\kappa)\int_{j}^{j(k+1)} ||h(\lambda^t(v))||dv.
\end{aligned} \tag{3.71}
$$

Thus,

$$\lim_{t\to\infty}\sup_{j\in[j(t),j(b)]}||\bar{\lambda}(j)-\lambda^t(j)||\leq\lim_{t\to\infty}\sup_{t\leq k\leq b}||\bar{\lambda}(j(k))-\lambda^t(j(k))||=0. \tag{3.72}$$

Then, we take the εth mean on both sides for $1<\varepsilon<\beta$ and obtain

$$\lim_{t\to\infty}E\left[\sup_{j\in[j(t),j(b)]}||\bar{\lambda}(j)-\lambda^t(j)||^{\varepsilon}\right]\leq\lim_{t\to\infty}E\left[\sup_{t\leq k\leq b}||\bar{\lambda}(j(k))-\lambda^t(j(k))||^{\varepsilon}\right]=0 \tag{3.73}$$

because $\lambda^t(\cdot)$ is the solution of equation (3.49). Then, we obtain

$$\lim_{t\to\infty}E[||\lambda(t)-\lambda^*||^{\varepsilon}]=0. \tag{3.74}$$

This implies

$$\lim_{t\to\infty}E[||D(\lambda(t))-D^*||^{\varepsilon}]=0. \tag{3.75}$$

Therefore, due to the strong duality,

$$\lim_{t\to\infty}E[||U(\bar{W}(t)r)-U^*||^{\varepsilon}]=0 \tag{3.76}$$

which concludes that the time-average stochastic gradient descent algorithm is utility optimal and completes the proof. □

Theorem 4. *Under Algorithm 2, the time-average transmission probability $\overline{W}(t)$ converges to a constant, i.e.,*

$$\lim_{t\to\infty}\overline{W}(t)=\frac{U^{-1}(D^*)}{r}. \tag{3.77}$$

Proof. From the strong convexity [35], we have

$$f(y)-f(x)\leq\left\langle f^{'}(x),y-x\right\rangle+\frac{L}{2}||y-x||^2 \tag{3.78}$$

where $<\cdot,\cdot>$ denotes the inner product, and L is a positive constant. Furthermore, consider $\lambda(t+1)=\lambda(t)-\alpha(t)g(t)$. Then, from equation (3.78), we have the following relationship:

$$\begin{aligned}D(\lambda(t+1))&\leq D(\lambda(t))+\left\langle D^{'}(\lambda(t)),\lambda(t+1)-\lambda(t)\right\rangle+\frac{L}{2}||\lambda(t+1)-\lambda(t)||^2\\&=D(\lambda(t))-\alpha(t)||D^{'}(\lambda(t))||^2+\frac{L}{2}\alpha(t)^2||D^{'}(\lambda(t))||^2\\&=D(\lambda(t))-(1-\frac{L}{2}\alpha(t))\alpha(t)||D^{'}(\lambda(t))||^2\end{aligned} \tag{3.79}$$

where $g(t) = D^{'}(\lambda(t))$. Therefore,

$$E[D(\lambda(t)) - D(\lambda(t+1))] \geq (1 - \frac{L}{2}\alpha(t))\alpha(t)E[||g(t)^2||]. \tag{3.80}$$

When $t \to \infty$, from Theorem 3, this implies that

$$\lim_{t\to\infty} E[D(\lambda(t)) - D(\lambda(t+1))] = \lim_{t\to\infty} E[D(\lambda(t)) - D^*] = 0 \tag{3.81}$$

because $E[||g(t)^2||] \to 0$ as $t \to \infty$, where D^* is the optimal value of the dual function. Therefore, from the first-order optimality condition, we can conclude that

$$\lim_{t\to\infty} E[||g(\lambda(t))||] = 0. \tag{3.82}$$

According to the strong duality, there is no gap between the dual and the primal domain solutions, and the minimum value of the dual function is the optimal solution of the primal function. Therefore,

$$U(w^*r) = D^* \tag{3.83}$$

and

$$w^* = \frac{U^{-1}(D^*)}{r}. \tag{3.84}$$

Finally, since $\lim_{t\to\infty} \overline{W}(t) = w^*$, we obtain

$$\lim_{t\to\infty} \overline{W}(t) = \frac{U^{-1}(D^*)}{r}. \tag{3.85}$$

□

Theorem 3 shows that the service rate will approach the arrival rate over time. Moreover, as the service rate converges to a constant, both a light queue and a heavy queue evolve as a G/G/1 queue. Thus, we have the following theorem:

Theorem 5. *The time-average stochastic gradient scheduling algorithm is moment stable.*

Proof. It follows from Theorem 4 that when $t \to \infty$, the transmission probability of each queue converges to a constant $\overline{W}_i(t)$. In this case, both the light and heavy queues evolve as a $G/G/1$ queue with a constant service rate. The queueing analysis for the G/G/1 queue with heavy-tailed arrivals indicates that the tail distribution of the heavy queue is one order heavier than the HT arrivals [43], i.e., $\lim_{t\to\infty} \frac{\log[P(Q_i>t)]}{t} = -\beta_i + 1$. This implies that all heavy queues have the bounded moments of order $\beta_i - 1$. However, by the classic queueing theory of a G/G/1 queue with light-tailed arrivals, it directly follows that the light queue has LT distributed queue length. Because all LT distributed random variables have a bounded mean, all light queues are of a bounded average queue length, i.e., $E[Q_i(t)] < \infty$, $\forall A_i(t) \in LT$. This indicates that the network is moment stable. □

3.3.6 Simulation Results

In this section, we use simulations to illustrate our theoretical results. More specifically, we choose Pareto and exponential distributions to represent HT and LT distributions, respectively. We refer to a random variable $X \in \mathcal{PAR}(\alpha, x_m)$, if it follows a Pareto distribution with parameters α and x_m, i.e., $P(X > x) = (x_m/x)^\alpha$. We refer to a random variable $X \in \mathcal{EXP}(\lambda)$, if it follows an exponential distribution with parameter λ, i.e., $P(X > x) = e^{-\lambda x}$. All of the following simulation results are plotted on log-log coordinates, whereby an HT distribution can manifest itself as a straight line with a slope equal to the negative value of the tail index β.

We first evaluate the stability performance of the stochastic subgradient scheduling algorithm by considering a scenario where three users are sharing one channel. We assume that user 1 is transmitting heavy-tailed traffic with arrival processes $A_1(t) \in \mathcal{PAR}(1.5, 1)$, and users 2 and 3 are sending light-tailed traffic with $A_2(t) \in \mathcal{EXP}(1/3)$ and $A_3(t) \in \mathcal{EXP}(1/2)$, respectively. Consequently, we have $a_1 = E[A_1(t)] = 3$, $a_2 = E[A_3(t)] = 3$, and $a_3 = E[A_3(t)] = 2$, respectively. Assume that the data rate of the wireless channel is 15 for all three users. Moreover, we define the utility function as $U_i(w_i r_i) = log(w_i r_i)$. In this case, according to Theorem 1, users 2 and 3 will have unbounded average queue lengths. As shown

in Figure 3.2, the queue lengths of users 2 (i.e., LT1) and 3 (i.e., LT2) have a tail distribution that exhibits itself as a straight line parallel to that of the reference Pareto distribution with tail index 0.5. Similarly, the queue delay is also unbounded, which is shown in Figure 3.3 that users 2 and 3 have the same slope as user 1. This indicates that the queueing lengths of users 2 and 3 have unbounded means, and thus the moment stability is not achievable under the classic SSSA.

Furthermore, we investigate the stability performance of the proposed time-average stochastic gradient scheduling algorithm under the same network settings as the previous case. Specifically, we define the step size $\alpha(t) = 1/(t+1)$, which satisfies $\sum_{t=0}^{\infty} \alpha(t) = \infty$ and $\sum_{t=0}^{\infty} \alpha(t)^2 < \infty$, and show in Figure 3.4 that the tail distributions of users 2 and 3 have a slope or a decaying rate smaller than that of the reference tail index 0.5, which implies that the queue lengths of users 2 and 3 have a finite means. This is because Theorem 3 indicates that under the proposed TA-SGSA, the service rate of each user converges to a constant, and each queue will evolve as a G1/G1/1 queue. This, combined with Theorem 4, leads to a bounded average queue length for light queues. The converged transmission rate of each user is shown in Figure 3.6. Likewise, we find that the queue delay of users 2 and 3 decreases faster than that of reference $\alpha = 1$ in Figure 3.5, which means that the queue delays for users 2 and 3 are also bounded under the proposed TA-SGSA. Moreover, to show that the proposed TA-SGSA has better performance regarding queue delay, we plot the instantaneous queue delay in the time domain. From Figure 3.8, we can see that the queue delay under the classic SSSA reaches to around a 230-packet delay. However, the queue delay under the proposed TA-SGSA is much lower than that, with only a maximum six-packet delay, as shown in Figure 3.9. Thus, we can conclude that the proposed TA-SGSA not only can guarantees a bounded mean for LT users but also provides them with a very small delay. To show further the difference the between two algorithms, we plot the queue delays of two algorithms delay together in Figure 3.10, which shows that the queue delay under the TA-SGSA is much smaller than that under the classic SSSA.

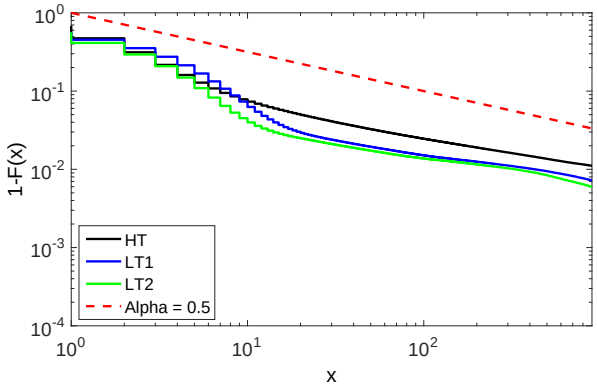

Figure 3.2: Queue-length tail distribution under conventional SSSA.

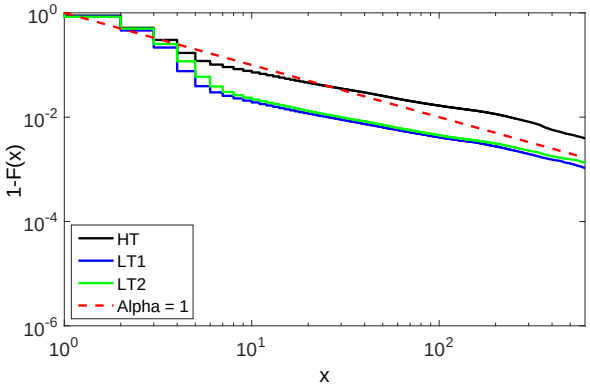

Figure 3.3: Queue-delay tail distribution under conventional SSSA.

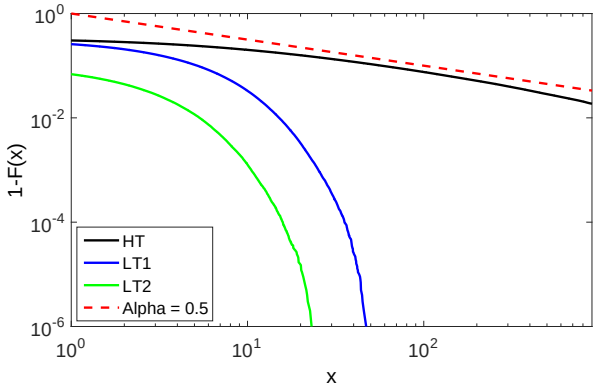

Figure 3.4: Queue-length tail distribution under proposed TA-SGSA.

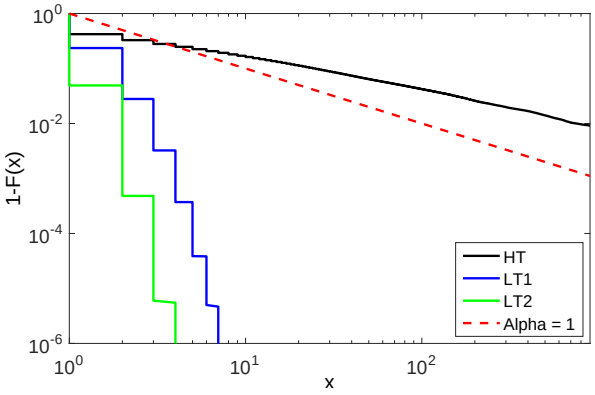

Figure 3.5: Queue-delay tail distribution under proposed TA-SGSA.

Finally, to show that the proposed time-average stochastic gradient descent algorithm is utility optimal, we first compute the theoretical maximum utility value as $3 * log(5) = 4.8283$. Moreover, in Figure 3.7, it can be seen that the allocated transmission rate converges to 5, which means the channel has been divided equally for three users. Therefore, combined with Theorem 4, it can be stated that under the proposed TA-SGSA, the actual utility will gradually approach the theoretical maximum utility, which Figure 3.11 verifies.

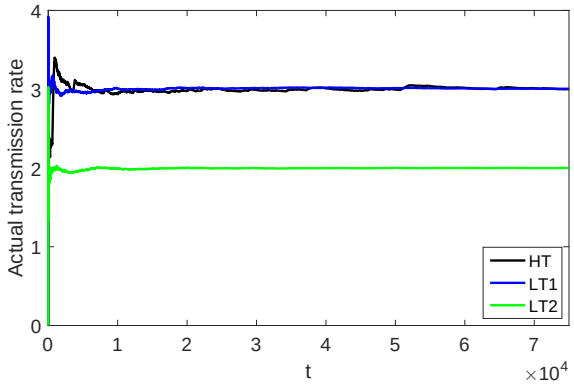

Figure 3.6: Converged transmission rate under TA-SGSA.

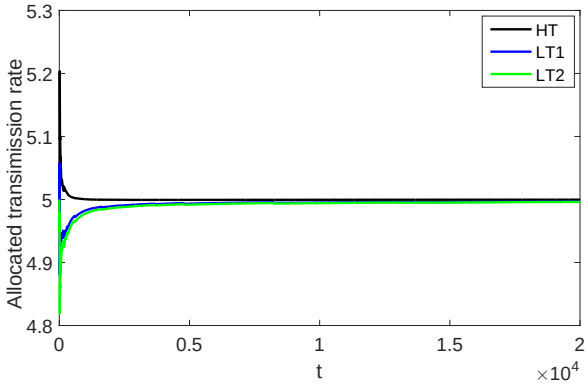

Figure 3.7: Converged allocated transmission rate under TA-SGSA.

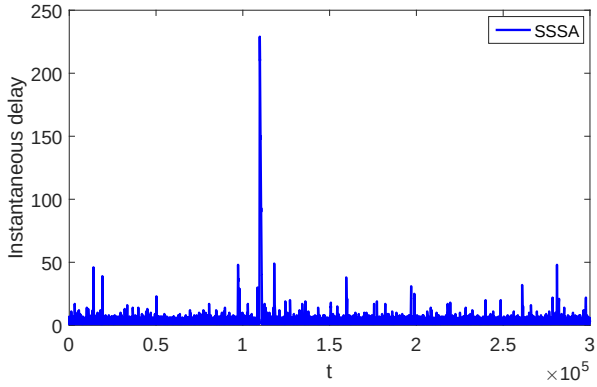

Figure 3.8: Instantaneous queue delay under SSSA.

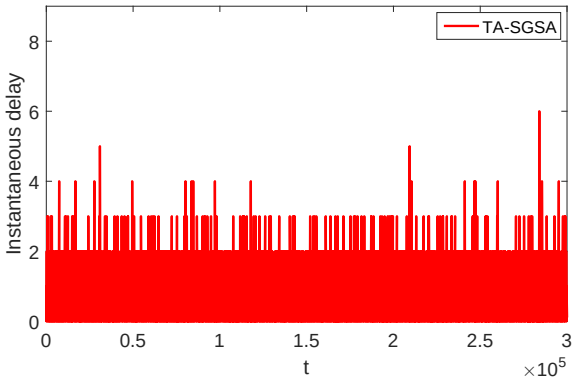

Figure 3.9: Instantaneous queue delay under TA-SGSA.

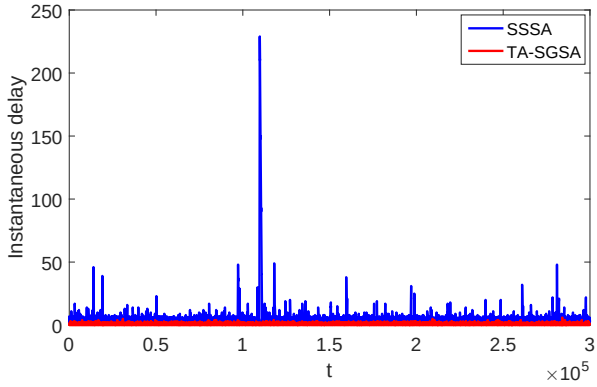

Figure 3.10: Comparing queue delay between classic SSSA and proposed TA-SGSA.

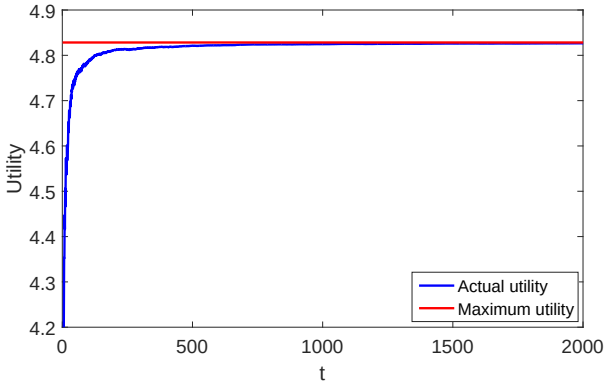

Figure 3.11: Network utility under TA-SGSA.

CHAPTER 4

THROUGHPUT OPTIMAL ROUTING ALGORITHM IN THE PRESENCE OF HEAVY TAILS

In this chapter, we will introduce several routing algorithms, and analyze their performance under heavy-tailed traffic condition.

4.1 Backpressure Routing Algorithm Review

4.1.1 System Model

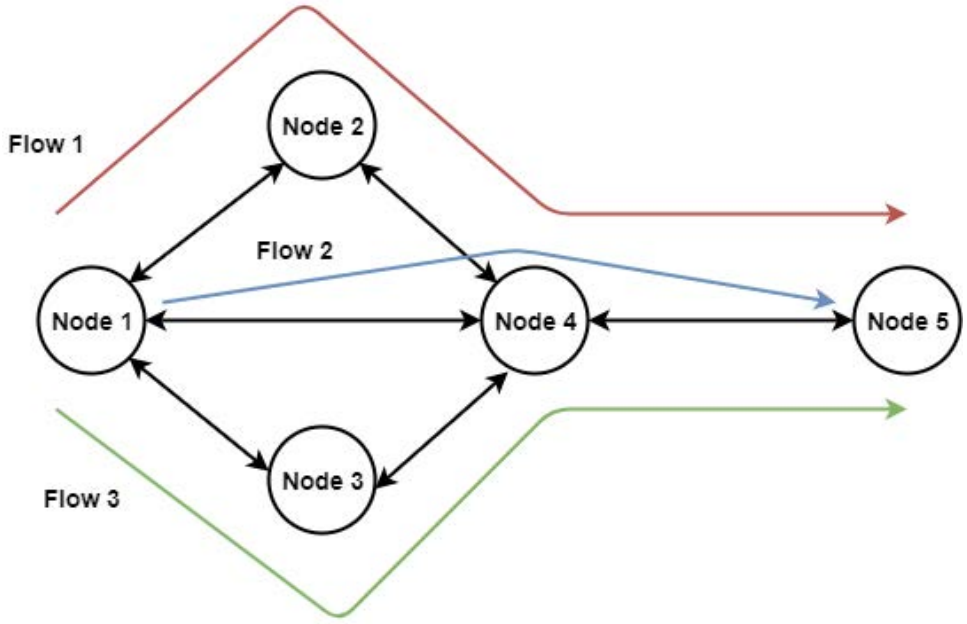

Figure 4.1: Five-node multi-hop network with three flows.

Consider a multi-hop network with N users where time is slotted with a unit slot size. On each time slot t, new packets come to the origination nodes and are delivered to the destination nodes according to the routing and transmission scheduling policy. Let $f \in \{1, ..., F\}$ denote network flows from the origination nodes to the destination nodes. Moreover, let $A_i^f(t)$ denote the number of packets that arrive at origination node i during

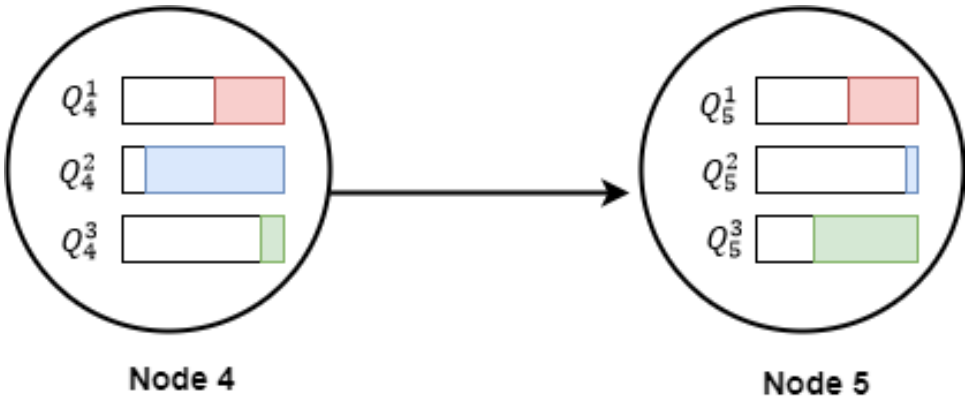

Figure 4.2: Backpressure routing between nodes 4 and 5.

time slot t for flow f with an average traffic rate or traffic intensity of $E[A_i^f(t)] = a_i^f$. Figure 4.1 shows that there are five nodes, where node 1 is the source node and node 5 is the destination node. Therefore, there exist three flows going to destination node 5, as shown in Figure 4.1. More specifically, the backpressure routing algorithm details between nodes 4 and 5 in Figure 4.2 shows three flows. It can be seen that flow 2 has the biggest pressure (i.e., difference between Q_4^f and Q_5^f), thus flow 2 of the corresponding queue Q_4^2 will transmit at the current time slot.

4.1.2 Queueing Dynamics

Let $Q_i(t)$ denote the number of packets in queue i at the end of time slot t. Then, the queueing dynamic of user i can be defined as

$$Q_i^f(t+1) = \left[Q_i^f(t) - \sum_{j \in N_c} (R_{i,j}^f(t) - R_{j,i}^f(t)) + A_i^f(t) \right]^+ \tag{4.1}$$

where $Q_i^f(t)$ denotes the queueing length of user i at time slot t for flow f, $R_{i,j}^f(t)$ is the number of transmitted packets from node i to node j, $j \in N_c$, N_c is the set of the neighbor nodes connected to node i during time slot t for flow f, and $A_i^f(t)$ is the arrival rate of user i at time slot t for flow f. Furthermore, there exist the following two constraints to make

the network stable:

$$\sum_{j \in N_c} (R_{i,j}^f(t) - R_{j,i}^f(t)) \geq A_i^f(t) \tag{4.2}$$

$$\sum_{f \in F} R_{i,j}^f(t) \leq C_{i,j}. \tag{4.3}$$

The first constraint shows that the total delivered packets of user i should be larger or equal to the number of arrival packets during time slot t for flow f. The second constraint denotes that the total transmitted packets for all flows at time slot t should be smaller than or equal to the channel capacity $C_{i,j}$ between nodes i and j. At each time slot t, the backpressure routing algorithm selects the flow f that owns the biggest "pressure" to transmit between node i and j as

$$f^* = \arg\max_f [Q_i^f(t) - Q_j^f(t)]^+. \tag{4.4}$$

If $Q_i^f(t) - Q_j^f(t) \leq C_{i,j}$, then $(Q_i^f(t) - Q_j^f(t))$ number of packets will be transmitted; if $Q_i^f(t) - Q_j^f(t) \geq C_{i,j}$, then at most $C_{i,j}$ number of packets will be transmitted.

Remark 1. *The backpressure routing algorithm was developed under the light-tailed traffic condition. When the network has a heavy-tailed traffic arrival, the whole system will be unstable. More specifically, the LT flow will be seriously starved since the HT flow has more chance of being transmitted, which will lead to an unbounded queue and queue instability.*

4.2 Stochastic Gradient Routing Algorithm

In this section, considering the NUM problem, we introduce the stochastic gradient routing algorithm, and analyze its performance under a heavy-tailed traffic condition.

4.2.1 System Model

Consider a multi-hop network $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ is the set of nodes, and $\mathcal{L}$ is the set of links between nodes with cardinality $|\mathcal{N}| = N$ and $|\mathcal{L}| = L$, respectively. This paper considers a single wireless channel model between nodes. Time is slotted with a unit slot size. We denote the capacity of link $(i, j) \in \mathcal{L}$ as $R_{i,j}$ and define the neighborhood

of node i as the set $N_i = \{j \in \mathcal{N} | (i,j) \in \mathcal{L}\}$. Let $\mathcal{F}$ denote a set of traffic flows. Each flow $f \in \mathcal{F}$ has a source node $S(f)$ and a destination node $D(f)$, where $S(f), D(f) \in \mathcal{N}$, $(S(f), D(f)) \in \mathcal{L}$, and $S(f) \neq D(f)$ for all flows f. Moreover, each node maintains two types of queues according to the corresponding layer:

At the network layer, each node maintains the flow queues for corresponding flows. Let $Q_i^f(t)$ denote the network layer flow queue, which temporally stores packets for each flow f at each node i, and let $A_i^f(t)$ denote the number of packets that arrive at queue i for flow f during time slot t with an average traffic rate or traffic intensity of $E[A_i^f(t)] = a_i^f$. Then, the flow queue $Q_i^f(t)$ of node i for flow f can be represented by

$$Q_i^f(t+1) = \left[Q_i^f(t) - \sum_{j \in N_i} \left(r_{i,j}^f(t) - r_{j,i}^f(t) \right) + A_i^f(t) \right]^+, \tag{4.5}$$

where $[\cdot]^+ = \max(\cdot, 0)$, and $r_{i,j}^f(t)$ is the routing rate from nodes i and j for flow f during time slot t with mean $E[r_{i,j}^f(t)] = r_{i,j}^f$.

At the MAC layer, each node builds link queues for its neighbor nodes. In addition, we use a protocol model [44] [45] [46] to define the collision-free link. More specifically, consider N nodes arbitrarily located on a plane, and let $d_{i,j}$ denotes the distance between nodes i and j. The communication range is denoted by Z_i, and $Z_i^{'}$ denotes the interference range for node i. Then, a successful transmission can be made by following two conditions:

$$d_{i,j} \leq Z_i \tag{4.6}$$

$$d_{k,j} \geq Z_k^{'}, \qquad \forall k \in N, \tag{4.7}$$

where the first condition means that the distance between node i and node j is within the communication range Z_i, and the second condition tells us that node j is out of the interference range of any other nodes k. We consider a collision-free MAC layer model that contains S number of collision-free link rate vectors. At a time slot, a collision-free link rate vector R^s, $s \in \mathcal{S}$, a subset of links that are activated without any collisions with each other, is selected. We denote the set of all collision-free link capacity vectors by $\mathcal{R}$ and $\mathcal{R} =$

$\{R^1, R^2, ..., R^S\}$, where $R^s = [R_1^s, R_2^s, ..., R_L^s]$. Furthermore, to guarantee network system stability, the routing rate of all flows should be within the collision-free link capacity region. We use the convex hull $Co(\mathcal{R})$ to denote our collision-free link capacity region, and $W_s(t)$ to denote a indicator function at time slot t for vector R^s. When $W_s(t) = 1$, the collision-free link capacity vector R^s is selected, and the expectation of $W_s(t)$, i.e., $E[W_s(t)] = w_s$, is the probability that the collision-free link capacity vector R^s will be selected. Then, the convex hull can be represented by

$$Co(\mathcal{R}) = \{\sum_{s=1}^{S} w_s R^s | (\forall s : w_s \geq 0) \cap \sum_{s=1}^{S} w_s = 1\}. \tag{4.8}$$

Thus, considering the link (i, j), the actual data rate can be computed by

$$C_{i,j}^s = \sum_{s=1}^{S} w_s R_{i,j}^s. \tag{4.9}$$

More specifically, Figure 4.3 shows a simple network topology with five nodes, where node 2 receives packets from node 1 with rate $C_{1,2}^s$ under a collision-free link capacity vector $R_{i,j}^s$ and also sends packets to nodes 3 and 4 with rate $C_{2,3}^s$ and $C_{2,4}^s$, respectively. Figure 4.4 shows the details of node 2 shown in Figure 4.3. In node 2, there are two types of flows: red and green. Thus, there are two flow queues in node 2. Furthermore, since node 2 has two neighbor nodes, it also maintains two link queues: $q_{2,3}$ and $q_{2,4}$. In addition, let $q_{i,j}(t)$ denote the MAC-layer link queue containing the packets stored at node i for delivery to neighboring node j. The link queue $q_{i,j}(t)$ of node i receives a number $\sum_{f \in \mathcal{F}} r_{i,j}^f(t)$ of packets from

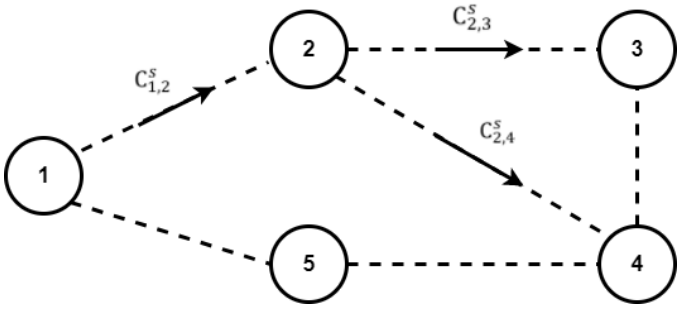

Figure 4.3: Network model.

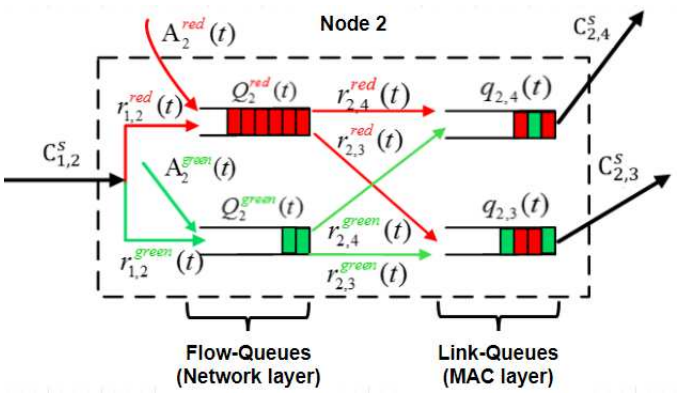

Figure 4.4: Cross-layer network model details.

its network layer, and a maximum number of $C^s_{i,j}(t)$ packets can be released from the link queue $q_{i,j}(t)$ and sent over the wireless link (i, j) between nodes i and j. As a result, the link queue $q_{i,j}(t)$ evolves as

$$q_{i,j}(t+1) = \left[q_{i,j}(t) - C^s_{i,j}(t) + \sum_{f \in \mathcal{F}} r^f_{i,j}(t) \right]^+, \tag{4.10}$$

where $C^s_{i,j}(t) = \sum_{s=1}^{S} W_s(t) R^s_{i,j}$ denotes the instantaneous data rate of the wireless link between nodes i and j under the collision-free link rate vector s.

4.2.2 Network Stability Analysis for Stochastic Gradient Routing Algorithm

In this section, we first formulate the utility routing problem as a concave optimization framework and utilize the classic gradient descent algorithm [35, Section 9.3] to solve such a problem. Then, we prove that the classic stochastic gradient routing algorithm cannot reach moment stability in the presence of HT traffic. In particular, we define a concave utility function $U(\cdot)$ that is monotone increasing on R_+. Then, we formulate the utility-optimal

routing problem as follows:

$$
\begin{aligned}
&Find && r_{i,j}^{f}, w_s && \forall (i,j) \in \mathcal{L}, \forall f \in \mathcal{F}, \forall s \in \mathcal{S} \\
&Maximize && \sum_{f,i,j} U(r_{i,j}^{f}) && \\
&s.t && \sum_{j \in N_i} (r_{i,j}^{f} - r_{j,i}^{f}) \geq a_i^{f}, && \forall f \in \mathcal{F}, i \in \mathcal{N} \\
& && \sum_{f \in \mathcal{F}} r_{i,j}^{f} \leq \sum_{s=1}^{S} w_s R_{i,j}^{s}, && \forall (i,j) \in \mathcal{L}, \forall s \in \mathcal{S} \\
& && r_{i,j}^{f} \geq 0, && \forall f \in \mathcal{F}, \forall (i,j) \in \mathcal{L} \\
& && \sum_{s=1}^{S} w_s = 1 && \forall s \in \mathcal{S} \\
& && 0 \leq w_s \leq 1, && \forall s \in \mathcal{S} \\
& && E[Q_i^{f}(t)] < \infty, && \forall f \in \mathcal{F}, i \in \mathcal{N} \\
& && E[q_{i,j}(t)] < \infty, && \forall (i,j) \in \mathcal{L}
\end{aligned}
\tag{4.11}
$$

where the first constraint is the flow-conservation constraint, which enforces, after a certain time, the average arrival traffic rate a_i^f to be less than the difference of transmission rates between nodes i and j, represented by $\sum_{j \in N_i}(r_{i,j}^f - r_{j,i}^f)$; the second constraint is a link-conservation constraint, which means the average link rate $\sum_{s=1}^{S} w_s R_{i,j}^s$ should be larger than the average arrival traffic rate $\sum_{f \in \mathcal{F}} r_{i,j}^f$; the third constraint means that the transmission rate $r_{i,j}^f$ should be larger than or equal to 0; the fourth constraint means that the summation of transmission probability for all collision-free link rate vectors is equal to 1; the fifth constraint is the transmission probability constraint; and the last two constraints are the flow and queue length constraints, which guarantee queue stability of LT queues in the presence of HT traffic.

To solve the above problem in equation (4.11), we define Lagrangian multipliers λ_i and the Lagrange function as

$$
\mathcal{L}(r, \lambda) = \sum_{f,i,j} U(r_{i,j}^{f}) + \sum_{f,i} \lambda_i^{f} \left(\sum_{j \in N_i} (r_{i,j}^{f} - r_{j,i}^{f}) - a_i^{f} \right). \tag{4.12}
$$

Then, we reorder equation (4.12) and obtain

$$\mathcal{L}(r,\lambda) = \sum_{f,i,j} \left(U(r_{i,j}^f) + r_{i,j}^f(\lambda_i^f - \lambda_j^f) \right) - \sum_{f,i} \lambda_i^f a_i^f. \tag{4.13}$$

Thus, we have the Lagrange dual function as

$$D(\lambda) = \sup_r \mathcal{L}(r,\lambda). \tag{4.14}$$

To find the primal Lagrangian maximizers $(r_{i,j}^f)^*$, we have

$$\begin{aligned}
& (r_{i,j}^f)^* = \arg\max_r \mathcal{L}(r,\lambda) \\
s.t. \quad & \sum_{f\in\mathcal{F}} r_{i,j}^f \le \sum_{s=1}^{S} w_s R_{i,j}^s, \quad \forall (i,j)\in\mathcal{L}, \forall s\in\mathcal{S} \\
& r_{i,j}^f \ge 0, \qquad\qquad \forall f\in\mathcal{F}, \forall (i,j)\in\mathcal{L} \\
& \sum_{s=1}^{S} w_s = 1, \qquad\qquad \forall s\in\mathcal{S} \\
& 0 \le w_s \le 1, \qquad\qquad \forall s\in\mathcal{S}.
\end{aligned} \tag{4.15}$$

To obtain the Lagrangian maximizers $(r_{i,j}^f)^*$, we first find the flow that has the maximum dual difference between node i and node j as

$$f^* = \arg\max_{f\in F}(\lambda_i^f - \lambda_j^f). \tag{4.16}$$

Then, we select a collision-free link rate vector R^s and obtain the primal Lagrangian maximizers as

$$\begin{aligned}
(r_{i,j}^{f^*})^* &= \arg\max_{r_{i,j}^{f^*} = w_s R_{i,j}^s} \mathcal{L}(r,w,\lambda) \\
&= \arg\max_{r_{i,j}^{f^*} = w_s R_{i,j}^s} \sum_{i,j} \left(U(r_{i,j}^{f^*}) + r_{i,j}^{f^*}(\lambda_i^{f^*} - \lambda_j^{f^*}) \right),
\end{aligned} \tag{4.17}$$

where the last equation holds because the last term of equation (4.13) does not affect the primal Lagrangian maximizers $(r_{i,j}^{f^*})^*$.

Then, we define the stochastic gradient $(g_i^{f^*})^*(t)$ as

$$(g_i^{f^*})^*(t) = \sum_{j\in N_i} \left((r_{i,j}^{f^*})^*(t) - (r_{j,i}^{f^*})^*(t) \right) - A_i^{f^*}(t). \tag{4.18}$$

Therefore, the dual variable λ can be updated in the direction of the stochastic gradient descent as

$$\begin{aligned}\lambda_i^{f^*}(t+1) &= \lambda_i^{f^*}(t) - \alpha_1(t)(g_i^{f^*})^*(t) \\ &= \lambda_i^{f^*}(t) - \alpha_1(t)\Bigg(\sum_{j\in N_i}\big((r_{i,j}^{f^*})^*(t) - (r_{j,i}^{f^*})^*(t)\big) - A_i^{f^*}(t)\Bigg)^+, \end{aligned} \tag{4.19}$$

where $\alpha_1(t)$ is the step size for all routers i. Then, we summarize the above conventional stochastic gradient routing algorithm in Algorithm 4. By letting $\lambda_i^f(0) = Q_i^f(0)$ for all flows, we can see that the flow queue-length update process defined in equation (4.5) corresponds to the dual-variable update process. Thus, it is easy to prove that Algorithm 4 is utility optimal. Next, we show that Algorithm 4 is not moment stable in the presence of HT traffic.

Algorithm 4 Conventional Stochastic Gradient Routing Algorithm

1: Observe $Q_i^f(0)$. Initialize $\lambda_i^f(0) = Q_i^f(0)$ for all nodes
2: **for** $t = 0, 1, 2, ...$ **do**
3: **for** all neighbors $j \in N_i$ **do**
4: Find largest dual-variable difference flow f^*
5: $f^* = \arg\max_{f\in F}\big(\lambda_i^f(t) - \lambda_j^f(t)\big)$
6: Find optimal primary variables $(r_{i,j}^{f^*})^*(t)$
7: $(r_{i,j}^{f^*})^*(t) = \arg\max_{r_{i,j}^{f^*}(t)=W_s(t)R_{i,j}^s}\sum_{i,j}\Bigg(U(r_{i,j}^{f^*}(t)) + r_{i,j}^{f^*}(t)\big(\lambda_i^{f^*}(t) - \lambda_j^{f^*}(t)\big)\Bigg)$
8: Transmit packets at $(r_{i,j}^{f^*})^*(t)$ between nodes i and j for flow f^*
9: **end for**
10: Update dual variable $\lambda_i^{f^*}(t+1) = \lambda_i^{f^*}(t) - \alpha_1(t)\Bigg[\sum_{j\in N_i}\big((r_{i,j}^{f^*})^*(t) - (r_{j,i}^{f^*})^*(t)\big) - A_i^{f^*}(t)\Bigg]^+$
11: **end for**

Theorem 6. *For any router, if there is a heavy-tailed flow with a tail index smaller than two, i.e.,*

$$\min_{i\in N}\kappa(A_i^{f_H}(t)) < 2, \tag{4.20}$$

then all flow queues in this router can experience unbounded queue length (i.e., $E[Q_i^f(t)] = \infty$).

Proof. See Appendix A. □

4.3 Time-Average Stochastic Gradient Routing Algorithm

In this section, to counter the instability problems of the classic stochastic gradient routing algorithm, we propose a TA-SGRA that can achieve utility optimality and queue stability simultaneously, even under a heavy-tailed environment. In particular, we first reconsider the Lagrange dual function in equation (4.14),

$$\begin{aligned} & D(\lambda) = \sup_{r} \mathcal{L}(r, \lambda) \\ s.t. \quad & \sum_{f \in \mathcal{F}} r_{i,j}^{f} \leq \sum_{s=1}^{S} w_s R_{i,j}^{s}, \quad \forall (i,j) \in \mathcal{L}, \forall s \in \mathcal{S} \\ & r_{i,j}^{f} \geq 0, \qquad \forall f \in \mathcal{F}, \forall (i,j) \in \mathcal{L} \\ & \sum_{s=1}^{S} w_s = 1, \qquad \forall s \in \mathcal{S} \\ & 0 \leq w_s \leq 1, \qquad \forall s \in \mathcal{S}. \end{aligned} \tag{4.21}$$

To find primal Lagrangian maximizers $(r_{i,j}^{f})^{*}$, we redefine a new Lagrange function as

$$\bar{\mathcal{L}}(r, w, \mu, \upsilon) = \mathcal{L}(r, \lambda) + \sum_{i,j} \mu_{i,j} \left(\sum_{s=1}^{S} w_s R_{i,j}^{s} - \sum_{f \in \mathcal{F}} r_{i,j}^{f} \right) + \sum_{i,j} \upsilon_{i,j} r_{i,j}^{f}, \tag{4.22}$$

where $\mu_{i,j}$ and $\upsilon_{i,j}$ are Lagrange multipliers. The expression $\sum_{s=1}^{S} w_s R_{i,j}^{s} - \sum_{f \in \mathcal{F}} r_{ij}^{f}$ comes from the first constraint in equation (4.21), and the last term comes from the second constraint in equation (4.21). Then, we have a new dual function of equation (4.22) as

$$\begin{aligned} & D(\mu, \upsilon) = \sup_{r,w} \bar{\mathcal{L}}(r, w, \mu, \upsilon) \\ s.t. \quad & \sum_{s=1}^{S} w_s = 1, \qquad \forall s \in \mathcal{S} \\ & 0 \leq w_s \leq 1, \qquad \forall s \in \mathcal{S}, \end{aligned} \tag{4.23}$$

and a dual problem as

$$\min_{\mu, \upsilon} D(\mu, \upsilon) = \min_{\mu, \upsilon} \sup_{r,w} \bar{\mathcal{L}}(r, w, \mu, \upsilon) \tag{4.24}$$

According to $\bar{\mathcal{L}}(r, w, \mu, \upsilon)$ defined in equation (4.22), the determination of Lagrange maximizers r and w is equivalent to the maximization of separate summands, which correspond

to optimization of the MAC layer parameter w and the network layer parameter r, respectively. The MAC-layer controls the link gate status with an average probability w_s to open the gate, whereas the network layer controls the transmission data rate r. In particular, **at the MAC-layer**, to find the optimum value of w_s^*, we have the following equations:

$$\begin{aligned} & w_s^* = \arg\max_{w_s} \sum_{i,j} \sum_{s=1}^{S} \mu_{i,j} w_s R_{i,j}^s \\ s.t. \quad & \sum_{s=1}^{S} w_s = 1, \qquad \forall s \in \mathcal{S} \\ & 0 \le w_s \le 1, \qquad \forall s \in \mathcal{S}. \end{aligned} \tag{4.25}$$

Note that, at time slot 0, we randomly select a collision-free link rate vector R^s, $s \in \mathcal{S}$. Then, considering time slot t, $t \ge 1$, $W_s(t)$ denotes an indicator function of two values, 1 and 0, depending on whether a collision-free link rate vector $R^s = \{R_1^s, R_2^s, ..., R_L^s\}$ is active or not. If the collision-free link rate vector R^s is active, then $W_s(t) = 1$; otherwise, $W_s(t) = 0$. In addition, if the link l is in the collision-free link rate vector R^s, then $R_l^s > 0$; otherwise, $R_l^s = 0$. Therefore, if it is active (i.e., $W_s(t) = 1$), then $\sum_{i,j} \mu_{i,j}(t-1) W_s(t) R_{i,j}^s$ will be maximum out of

$$\left\{ \sum_{i,j} \mu_{i,j}(t-1) W_1(t) R_{i,j}^1, \sum_{i,j} \mu_{i,j}(t-1) W_2(t) R_{i,j}^2, ..., \sum_{i,j} \mu_{i,j}(t-1) W_S(t) R_{i,j}^S \right\}, \quad t \ge 1 \tag{4.26}$$

Thus, $\sum_{i,j} \sum_{s=1}^{S} \mu_{i,j}(t-1) W_s(t) R_{i,j}^s$ will be equal to $\sum_{i,j} \mu_{i,j}(t-1) W_s^*(t) R_{i,j}^s$. Then, we denote

$$\begin{aligned} W_s^*(t) &= \arg\max_{W_s} \left\{ \sum_{i,j} \mu_{i,j}(t-1) W_1(t), \sum_{i,j} \mu_{i,j}(t-1) W_2(t), ..., \sum_{i,j} \mu_{i,j}(t-1) W_S(t) \right\} \\ &= \arg\max_{W_s} \{ \sum_{i,j} \mu_{i,j}(t-1) W_1(t) R_{i,j}^1, \sum_{i,j} \mu_{i,j}(t-1) W_2(t) R_{i,j}^2, ..., \sum_{i,j} \mu_{i,j}(t-1) W_S(t) R_{i,j}^S \Big\} \\ &= \arg\max_{W_s} \sum_{i,j} \sum_{s=1}^{S} \mu_{i,j}(t-1) W_s(t) R_{i,j}^s, \qquad t \ge 1. \end{aligned} \tag{4.27}$$

Furthermore, the time-average $w_s^*(t)$ can be computed by

$$w_s^*(t) = \begin{cases} W_s^*(t), & t = 0 \\ \sum_{k=1}^{t} \frac{W_s^*(k)}{t} & t \ge 1. \end{cases} \tag{4.28}$$

Note that we randomly select a collision-free link rate vector $R^s, s \in \mathcal{S}$ at time slot 0. **At the network layer**, to find the primal Lagrangian maximizers $(r_{i,j}^f)^*$, we first re-order equation (4.22) and obtain

$$\bar{\mathcal{L}}(r, w, \lambda, \mu, \upsilon) = \sum_{f,i,j} \big(U(r_{i,j}^f) + r_{i,j}^f(\lambda_i^f - \lambda_j^f)\big) - \sum_{f,i} \lambda_i^f a_i^f + \sum_{i,j} \mu_{i,j} \Bigg(\sum_{s=1}^{S} w_s R_{i,j}^s - \sum_{f \in \mathcal{F}} r_{i,j}^f \Bigg) + \sum_{i,j} \upsilon_{i,j} r_{i,j}^f. \tag{4.29}$$

Then, considering Karush-Kuhn-Tucker (KKT) conditions, we take the first derivative of the Lagrange function in equation (4.29) with respect to $r_{i,j}^f$, which yields

$$\frac{\partial \bar{\mathcal{L}}(r, \lambda, \mu, \upsilon)}{\partial r_{i,j}^f} = U^{'}(r_{i,j}^f) + (\lambda_i^f - \lambda_j^f) - \mu_{i,j} + \upsilon_{i,j}. \tag{4.30}$$

By the stationarity condition, we have

$$U^{'}(r_{i,j}^f) + (\lambda_i^f - \lambda_j^f) - \mu_{i,j} + \upsilon_{i,j} = 0. \tag{4.31}$$

In addition, by applying the complementary slackness condition [35, Section 5.5],

$$\upsilon_{i,j} r_{i,j}^f = 0 \quad and \qquad \upsilon_{i,j} \geq 0 \tag{4.32}$$

$$\mu_{i,j} (\sum_{s=1}^{S} w_s R_{i,j}^s - \sum_{f \in \mathcal{F}} r_{i,j}^f) = 0 \quad and \qquad \mu_{i,j} \geq 0. \tag{4.33}$$

From equation (4.31), we obtain $\upsilon_{i,j}$ as

$$\upsilon_{i,j} = \mu_{i,j} - (\lambda_i^f - \lambda_j^f) - U^{'}(r_{i,j}^f). \tag{4.34}$$

Because $\upsilon_i \geq 0$, we obtain the following relationship:

$$U^{'}(r_{i,j}^f) \leq \mu_{i,j} - (\lambda_i^f - \lambda_j^f). \tag{4.35}$$

In addition, since $U_i(\cdot)$ is a concave increasing, $U_i^{'}(\cdot)$ is monotone, decreasing, and positive. Then, we obtain

$$r_{i,j}^f \geq U^{'-1}\big(\mu_{i,j} - (\lambda_i^f - \lambda_j^f)\big). \tag{4.36}$$

If $U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big)>0$, then $r_{i,j}^f>0$. Therefore, to meet the complementary slackness condition [35, Section 5.5] in equation (4.32), we conclude that $\upsilon_{i,j}=0$. Then, according to equation (4.31), we have

$$r_{i,j}^f=U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big), \tag{4.37}$$

if $U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big)\leq 0$. This combined with (4.36) implies that $r_{i,j}^f\leq 0$. Thus, we can conclude that $r_{i,j}^f=0$. Then, we get

$$r_{i,j}^f=\begin{cases}U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big), & U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big)>0\\ 0 & U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big)\leq 0\end{cases}. \tag{4.38}$$

Therefore, equation (4.38) can be re-expressed as

$$r_{i,j}^f=\Big[U'^{-1}\big(\mu_{i,j}-(\lambda_i^f-\lambda_j^f)\big)\Big]^+. \tag{4.39}$$

Taking the summation on both sides of equation (4.39) and using the second equality constraint in equation (4.11) and the optimum w_s^* in equation (4.27) yields

$$\sum_{f\in\mathcal{F}}U'^{-1}\bigg(\Big[\big(\mu_{i,j}(t)-(\lambda_i^f(t)-\lambda_j^f(t)\big)\Big]^+\bigg)=w_s^*(t)R_{i,j}^s. \tag{4.40}$$

Since the dual variables $\lambda_i^f(t)$ and $\lambda_j^f(t)$ and link rate $w_s^*(t)R_{i,j}^s$ are known, the dual variable $\mu_{i,j}(t)$ can be calculated by using equation (4.40). In addition, according to the complementary slackness condition, we know that $\mu_{i,j}(t)>0$. Then, we can easily obtain the routing rate $r_{i,j}^f(\lambda,\mu,t)$ through

$$(r_{i,j}^f)^*(t)=(r_{i,j}^f)^*(\lambda,\mu,t)=U'^{-1}\bigg(\Big[\big(\mu_{i,j}(t)-(\lambda_i^f(t)-\lambda_j^f(t)\big)\Big]^+\bigg). \tag{4.41}$$

Taking the partial derivatives of function (4.29) with respect to $\lambda_i^f(t)$ yields the stochastic gradient $g_i^f(t)$ as

$$g_i^f(t)=\sum_{j\in N_i}\Big((r_{i,j}^f)^*(t)-(r_{j,i}^f)^*(t)\Big)-A_i^f(t). \tag{4.42}$$

Thus, we can update our stochastic gradient using Lagrangian maximizer $(r_{i,j}^f)^*(t)$ in (4.41). Moreover, we can calculate the time-average value $(\bar{r}_{i,j}^f)^*(t)$ using the following equation:

$$\begin{aligned}(\bar{r}_{i,j}^f)^*(t) &= \frac{1}{t}\sum_{k=1}^{t}(r_{i,j}^f)^*(k) \\ &= \frac{t-1}{t}(\bar{r}_{i,j}^f)^*(t-1) + \frac{1}{t}(r_{i,j}^f)^*(t), \quad t \geq 1. \end{aligned} \tag{4.43}$$

As a result, we can update our flow queue function using

$$Q_i^f(t+1) = \left[Q_i^f(t) - \sum_{j\in N_i}\left((\bar{r}_{i,j}^f)^*(t) - (\bar{r}_{j,i}^f)^*(t)\right) + A_i^f(t)\right]^+ \tag{4.44}$$

and link-queues function as

$$q_{i,j}(t+1) = \left[q_{i,j}(t) - W_s^*(t)R_{i,j}^s + \sum_{f\in\mathcal{F}}\bar{r}_{i,j}^f(t)\right]^+. \tag{4.45}$$

In addition, we can also update multiplier λ along the stochastic gradients in equation (4.42) using

$$\begin{aligned}\lambda_i^f(t+1) &= \lambda_i^f(t) - \alpha_2(t)g_i^f(t) \\ &= \left[\lambda_i^f(t) - \alpha_2(t)\left(\sum_{j\in N_i}\left((r_{i,j}^f)^*(t) - (r_{j,i}^f)^*(t)\right) - A_i^f(t)\right)\right]^+, \end{aligned} \tag{4.46}$$

where $\alpha_2(t)$ is the step size for all routers i. Note that $(\bar{r}_{i,j}^f)^*(t)$ is used in the flow-queue length and link-queue length update in equations (4.44) and (4.45), not in the stochastic gradient descent in equation (4.46).

The proposed time-average stochastic gradient routing algorithm is summarized in Algorithm 2. Initially, we set $\lambda_i^f(0) = Q_i^f(0)$ for all routers. Then, **at the network layer**, we first calculate the Lagrange dual variable $\mu_{i,j}$ in step 6. Based on the $\mu_{i,j}$, we can further compute the maximum routing rate for each router in step 9, which leads to a time-average routing rate of $(\bar{r}_{i,j}^f)^*(t)$. Then, **at the MAC layer**, we calculate the optimal value $W_s^*(t)$ in step 15 and time-average $w_s^*(t)$ in step 18. Then, each node can update its flow-queue function in step 19, and link-queue function in step 20. In addition, each node can also update the Lagrange multiples λ in step 21. More specifically, the proposed TA-SGRA has

two key features: first, we separate the routing and scheduling algorithms so that routing is implemented in the network layer and scheduling is implemented in the MAC layer. Second, we also decouple the queue-length update process in equation (4.44) and the dual-variable update process in equation (4.46) so that we can use the time-average Lagrange maximizer to manage the transmission rates and the instantaneous Lagrange maximizer to control the dual variables.

4.3.1 Utility and Stability Analysis

Here, we first prove that the TA-SGRA is utility optimal. More specifically, our algorithm uses an iterative method. This means that it generates a series of improved approximations for a class of problems. Hence, we can prove that the utility converges to the optimal value over time. In addition, the conventional convergence analysis is based on the Euclidean distance to the optimal set [37] and requires that the traffic arrival has bounded mean and variance. However, it is obvious that, in our case, if a heavy-tailed traffic arrives, then the arrival will have an unbounded mean and an unbounded variance. Therefore, to overcome this issue, we adopt the ordinary differential equation method [38, Section 2], which treats the discrete stochastic approximation scheme as a discretization version of the ODE with HT noise that is asymptotically negligible in the εth mean. Then, we establish the following theorem:

Theorem 7. *The time-average stochastic gradient routing algorithm is utility optimal.*

Proof. See Appendix B. □

The following theorem will show that the time-average transmission rate converges to some constant as time proceeds so that the LT flow queues will no longer compete with HT flow queues for transmission time.

Theorem 8. *Under the proposed Algorithm 2, the time-average transmission rate* $\overline{r}(t)$ *of*

Algorithm 5 Time-Average Stochastic Gradient Routing Algorithm

Observe $Q_i^f(0)$. Initialize $\lambda_i^f(0) = Q_i^f(0)$ for all routers

for $t = 0, 1, 2, ...$ **do**

When $t = 0$, randomly select a collision-free link rate vector $R^s, s \in \mathcal{S}$. i.e., $W_s^*(0) = 1$, and let $w_s^*(0) = W_s^*(0)$.

for $j \in N_i$ **do**

At the network layer, compute $\mu_{i,j}(t) > 0$ such that

$$\sum_{f\in\mathcal{F}} U'^{-1}\left(\left[\left(\mu_{i,j}(t) - (\lambda_i^f(t) - \lambda_j^f(t))\right]^+\right) = w_s^*(t)R_{i,j}^s$$

Compute the Lagrange maximizer $(r_{i,j}^f)^*(t)$, where

$$(r_{i,j}^f)^*(t) = U'^{-1}\left(\left[\mu_{i,j}(t) - \left(\lambda_i^f(t) - \lambda_j^f(t)\right)\right]^+\right)$$

Compute the time-average routing rate $(\overline{r}_{i,j}^f)^*(t)$, where

$$(\overline{r}_{i,j}^f)^*(t) = \begin{cases} 0, & t = 0 \\ \frac{t-1}{t}(\overline{r}_{i,j}^f)^*(t-1) + \frac{1}{t}(r_{i,j}^f)^*(t) & t \geq 1 \end{cases}$$

Transmit packets at rate $(\overline{r}_{i,j}^f)^*(t)$

end for

At the MAC layer, find an optimal $W_s^*(t+1)$ using

$$W_s^*(t+1) = \arg\max_{W_s} \sum_{i,j}\sum_{s=1}^{S} \mu_{i,j}(t-1)\cdot W_s(t+1)R_{i,j}^s$$

Compute the time-average $w_s^*(t+1)$, where

$$w_s^*(t+1) = \sum_{k=1}^{t+1} \frac{W_s^*(k)}{t+1}$$

Update the flow-queue function $Q_i^f(t+1) = \left[Q_i^f(t) - \sum_{j\in N_i}\left((\overline{r}_{i,j}^f)^*(t) - (\overline{r}_{j,i}^f)^*(t)\right) + A_i^f(t)\right]^+$

Update the link-queue function $q_{i,j}(t+1) = \left[q_{i,j}(t) - W_s^*(t)R_{i,j}^s + \sum_{f\in\mathcal{F}} \bar{r}_{i,j}^f(t)\right]^+$

Update the dual variable $\lambda_i^f(t+1) = \left[\lambda_i^f(t) - \alpha_2(t)\left(\sum_{j\in N_i}\left((r_{i,j}^f)^*(t) - (r_{j,i}^f)^*(t)\right) - A_i^f(t)\right)\right]^+$

end for

each flow, even in the presence of both LT and HT traffic, converges to a constant, i.e.,

$$\lim_{t\to\infty} \overline{r}^f(t) = U^{-1}(D^*), \tag{4.47}$$

where $D^* = U(r^{*f})$.

Proof. See Appendix C. □

Theorem 8 implies that both LT and HT traffic-flow optimization have been separated, instead of competing with each other. Thus, we have the following theorem:

Theorem 9. *Under the proposed time-average stochastic gradient routing algorithm, the LT flow queue has a bounded mean, even if an HT traffic flow is present.*

Proof. According to Theorem 8, we know that when $t \to \infty$, the transmission rate of each flow converges to a constant value, which implies that the competition between HT traffic flow and LT traffic flow has been diminished. Therefore, according to Xia and Wang [30], all LT traffic flow has a bounded mean, i.e., $E[Q_i^f(t)] < \infty, f \in LT$. □

Theorem 9 shows that at the network layer, the LT flow-queue length is bounded. Then, considering the link queue at the MAC layer, we have the following theorem:

Theorem 10. *Under the proposed time-average stochastic gradient routing algorithm, the link queue has a bounded mean.*

Proof. See Appendix D. □

According to Theorems 9 and 10, we know that both the LT traffic flow queue and link queue have bounded means. Thus, the network is moment stable.

4.4 Simulation Results

In this section, we use simulations to verify our theoretical results. We first define a log utility function as the proportional fairness metric in [36]. Thus, the utility function is set to $U(r_{i,j}^f) = \log(r_{i,j}^f)$, and from equation (4.41), we obtain the routing rate $r_{i,j}^f(t) =$

$\left[\frac{1}{\mu_{i,j}(t)-(\lambda_i^f(t)-\lambda_j^f(t))}\right]^+$. Moreover, we select Pareto and exponential distributions to represent HT and LT distributions, respectively. We refer to a random variable $X \in \mathcal{PAR}(\alpha, x_m)$, if it follows a Pareto distribution with parameters α and x_m, i.e., $P(X > x) = (x_m/x)^\alpha$. We refer to a random variable $X \in \mathcal{EXP}(\lambda)$, if it follows an exponential distribution with parameter λ, i.e., $P(X > x) = e^{-\lambda x}$. The following simulation results are plotted on log-log coordinates, by which an HT distribution can manifest itself as a straight line with the slope equal to the negative value of the tail index α.

To illustrate our theoretical results, we use a small, four-node network topology, as shown in Figure 4.5. At node 1, there is an LT flow arrival f_L, of which the destination is node 4 with arrival processes $A_1^{f_L}(t) \in \mathcal{EXP}(1/2)$ and $a_1^{f_L} = E[A_1^{f_L}] = 2$. At node 3, an HT flow is injected into the network with arrival processes $A_3^{f_H}(t) \in \mathcal{PAR}(1.5, 1)$ and $a_3^{f_H} = E[A_3^{f_H}(t)] = 3$, and the HT flow destination is also node 4. In addition, we assume that the link capacity $R_{i,j}$ between two nodes is 16, and $R^s = [R_1^s, R_2^s, R_3^s, R_4^s]$. From Figure 4.5, we can conclude that there are three collision-free link rate vectors: R^1, R^2 and R^3, which are [16 0 0 16], [0 16 0 0], and [0 0 16 0], respectively. Moreover, we observe and analyze HT and LT flows between nodes 3 and 4. In this case, **at the network layer**, according to Theorem 6, the light-tailed traffic flow f_L will have unbounded average queue lengths. As shown in Figure 4.6, the queue length of the LT flow (i.e., f_L) has a tail distribution that exhibits itself as a straight line parallel to that of the reference Pareto distribution with tail index $\alpha = 0.5$. In addition, in Figure 4.7, we see that the instaneous queue length of LT flow has reached 250 packets between time slots 88000 and 89000. The reason for this phenomenon is that under the conventional SGRA the HT flow seizes more transmission opportunities. As shown in Figure 4.8, between time slots 88290 and 88500, the LT flow of transmission rate is 0, which leads to a long queue length. This indicates that queueing lengths of the LT flows have unbounded means, and thus moment stability is not achievable under the conventional SGRA.

In addition, we verify the stability performance of the proposed TA-SGRA under the

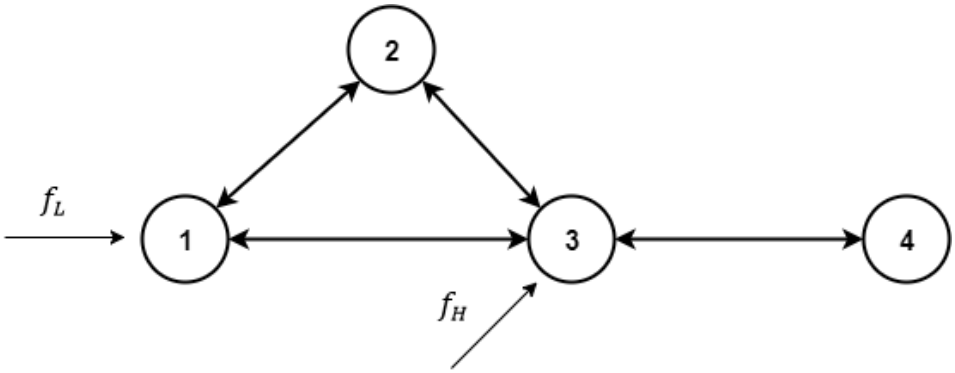

Figure 4.5: Simulation model.

same network settings as the previous case. Specifically, we define the step size in Algorithm 2 as $\alpha_2(t) = 1/(t+1)$, which satisfies $\sum_{t=0}^{\infty} \alpha_2(t) = \infty$ and $\sum_{t=0}^{\infty} \alpha_2(t)^2 < \infty$. It is shown in Figure 4.9 that the tail distributions of LT traffic flow f_L have a slope or a decaying rate larger than that of the reference tail index 1. Moreover, in Figure 4.10, we find the maximum queue length of LT flow is around 25 packets, compared with the LT flow of queue length under the conventional SGRA, where the latter is 10 times higher than the former, which implies that the queue lengths of LT flow f_L is of a finite mean under the proposed TA-SGRA. This is because of Theorem 8, which indicates that under the proposed TA-SGRA, the routing rate of each flow at one node converges to a constant. This, combined with Theorem 9, leads to a bounded average queue length for an LT traffic flow. Figure 4.11 verifies the converged routing rate of each flow.

Furthermore, to show that the proposed TA-SGRA is utility optimal, we first show that the probability of the selected collision-free link rate vectors R^1, R^2, and R^3 are also converged. As shown in Figure 4.12, the probabilities of w_1, w_2, and w_3 converge to 0.6010, 0.1994, and 0.1996 respectively. Thus, the mean of the data rate between nodes 3 and 4 is 16*0.6010 = 9.616. Thus, in Figure 4.13, it can be seen that HT flow and LT flow of the allocated routing rate converge to 4.8, which means that the TA-SGRA algorithm can reach the maximum network utility $2 * ln(4.8) = 3.1372$ and verifies our Theorem 7. Therefore, the network is stable and utility optimal.

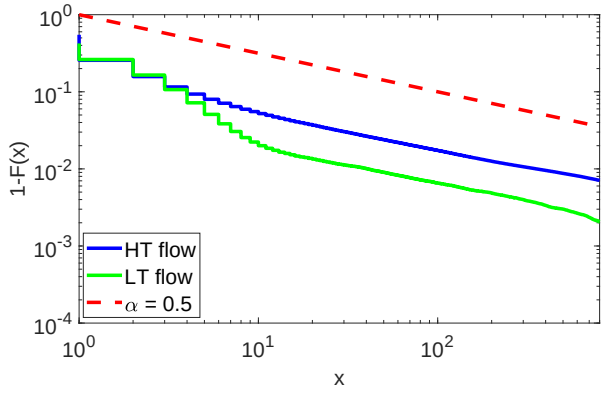

Figure 4.6: Queue-length tail distribution under conventional SGRA.

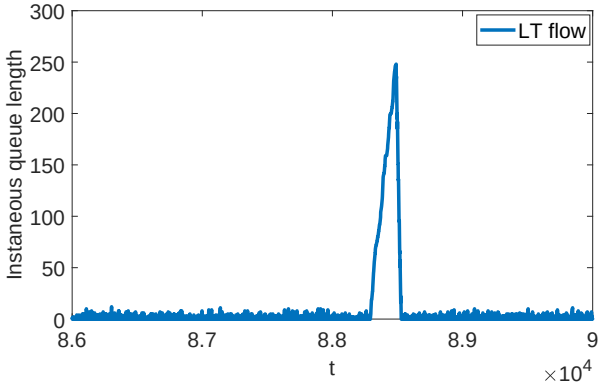

Figure 4.7: Instaneous queue length under conventional SGRA.

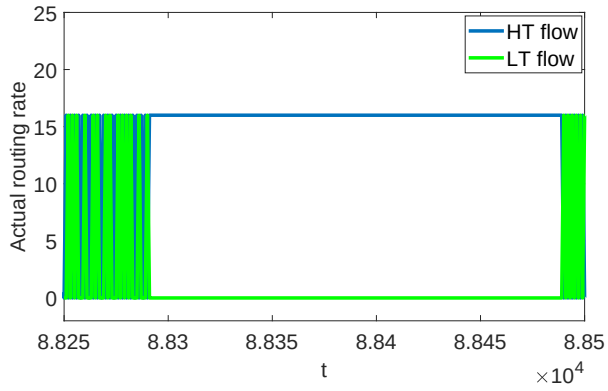

Figure 4.8: Routing rate under conventional SGRA.

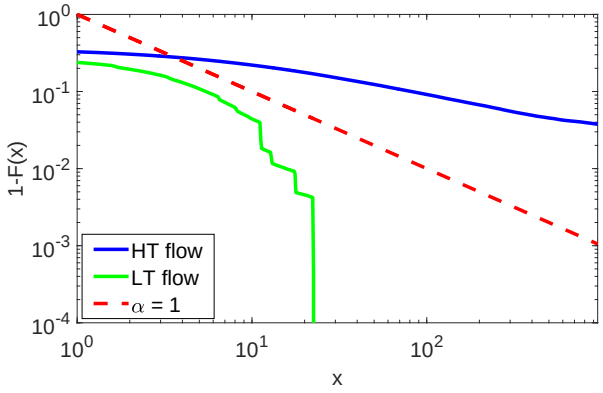

Figure 4.9: Queue-length tail distribution under proposed TA-SGRA.

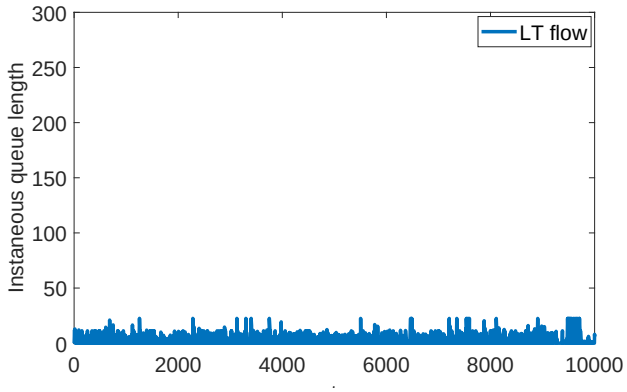

Figure 4.10: Instaneous queue length under proposed TA-SGRA.

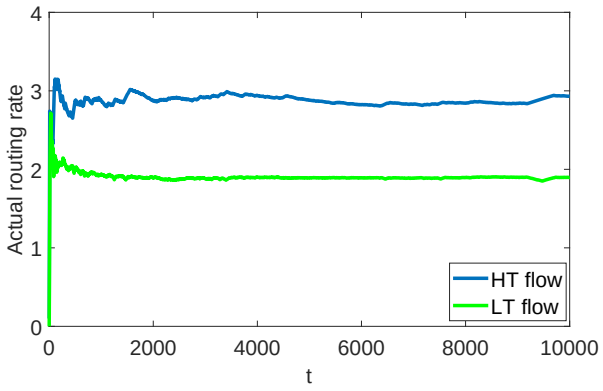

Figure 4.11: Convergence of routing rate under proposed TA-SGRA.

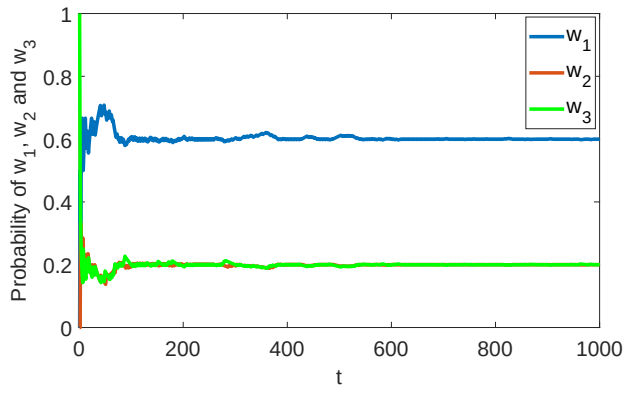

Figure 4.12: Convergence of probability of selected collision-free link rate vector.

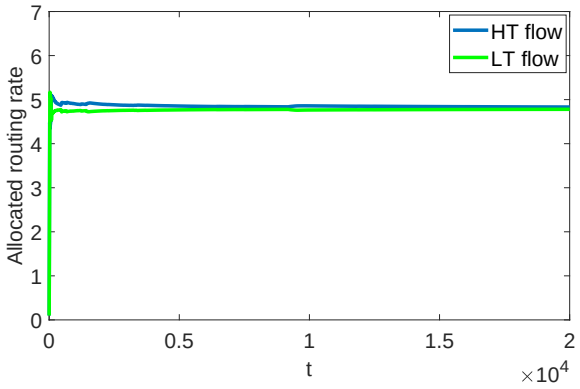

Figure 4.13: Convergence of allocated routing rate under proposed TA-SGRA.

CHAPTER 5

CONCLUSIONS

In this dissertation, two optimal network control policies were developed. For a single network, a time-average stochastic gradient scheduling algorithm was proposed. In addition, the impact of heavy-tailed traffic was first analyzed on the classic stochastic network utility maximization problem. In particular, it was shown that the classic stochastic subgradient algorithm is no longer stable due to the inherent burstiness in the HT traffic. More specifically, it was shown that the tail distribution of the light queue is at least one order heavier than the tail distribution of HT traffic arrivals, which means that the light queue length is of infinite mean. To solve this problem, the time-average stochastic gradient scheduling algorithm, which exploits the decoupled queue-length and dual-variable update processes along with the proposed time-average Lagrange maximizer, were proposed. In this way, the utility optimality and the moment stability can be guaranteed simultaneously. Similarly, for a multi-hop network, a time-average stochastic gradient routing algorithm under a cross-layer network system was proposed. Furthermore, this dissertation proved that the time-average transmission rate for each flow converges to a constant as time proceeds. This implies that the issue that an HT flow queue competes with an LT flow queue for transmission time under a classic SGRA has been solved by the proposed TA-SGRA. In addition, due to the features of HT traffic, an ordinary differential equation approach was adopted to explore the convergence and stability properties of the proposed TA-SGRA. On the other hand, at the MAC layer, the classical Lyapunov drift method was used to show that the link queue is also bounded. Hence, it can be concluded that the whole network is moment stable with the proposed TA-SGSA and TA-SGRA algorithms.

REFERENCES

REFERENCES

[1] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37(12):1936–1948, 1992.

[2] Y. Chen, X. Wang, and L. Cai. On achieving fair and throughput-optimal scheduling for tcp flows in wireless networks. *IEEE Transactions on Wireless Communications*, 15(12):7996–8008, 2016.

[3] S. Xia and P. Wang. Distributed throughput optimal scheduling in the presence of heavy-tailed traffic. In *IEEE International Conference on Communications (ICC) 2015*, pages 3490–3496, London, UK, June 2015.

[4] S. Xia. Distributed throughput optimal scheduling for wireless networks. Master's thesis, Wichita State University, 2014.

[5] X. Zhou, Z. F. Zhao, R. P. Li, Y. F. Zhou, J. Palicot, and H. G. Zhang. Understanding the nature of social mobile instant messaging in cellular networks. *IEEE Communications Letters*, 18(3):389–392, 2014.

[6] J. Yang, H. M. Kwon, A. Mukherjee, and K. D. Pham. Spreading-sequence design for partially connected multirelay networks under multipath fading. *IEEE Transactions on Vehicular Technology*, 65(3):1420–1433, 2016.

[7] H. M. Kwon and K. B. Lee. A novel digital fm receiver for mobile and personal communications. *IEEE Transactions on Communications*, 44(11):1466–1476, 1996.

[8] H. M. Kwon. Third-generation tdrss-compatible direct-sequence spread-spectrum digital receiver. *IEEE Transactions on vehicular technology*, 46(4):891–899, 1997.

[9] S. Borst, D. Ooteghem, and B. Zwart. Tail asymptotics for discriminatory processor-sharing queues with heavy-tailed service requirements. *Performance Evaluation*, 61(2-3):281–298, 2005.

[10] X. Yu and M. Becchi. Gpu acceleration of regular expression matching for large datasets: exploring the implementation space. In *Proceedings of the ACM International Conference on Computing Frontiers*, page 18. ACM, 2013.

[11] X. Yu and M. Becchi. Exploring different automata representations for efficient regular expression matching on gpus. In *ACM SIGPLAN Notices*, volume 48, pages 287–288. ACM, 2013.

[12] P. Wang and I. F. Akyildiz. Spatial correlation and mobility aware traffic modeling for wireless sensor networks. *IEEE/ACM Transactions on Networking*, 19(6):1860–1873, 2011.

REFERENCES (continued)

[13] V. Ramaswami, K. Jain, R. Jana, and V. Aggarwal. Modeling heavy tails in traffic sources for network performance evaluation. *computational intelligence, cyber security and computational models*, 246:23–44, 2014.

[14] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. K. Shankaranarayanan. Modeling and characterization of large-scale WI-FI traffic in public hot-spots. In *IEEE INFOCOM*, pages 2921–2929. IEEE, 2011.

[15] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken. The nature of data center traffic: Measurements & analysis. In *ACM 9th SIGCOMM Conference on Internet Measurement*, pages 202–208, New York, NY, USA, 2009.

[16] X. Yu, H. Wang, W. Feng, H. Gong, and G. Cao. An enhanced image reconstruction tool for computed tomography on cpus. In *Proceedings of the Computing Frontiers Conference*, pages 97–106. ACM, 2017.

[17] X. Yu, K. Hou, H. Wang, and W. Feng. Robotomata: A framework for approximate pattern matching of big data on an automata processor. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 283–292. IEEE, 2017.

[18] X. Yu, B. Lin, and M. Becchi. Revisiting state blow-up: Automatically building augmented-fa while preserving functional equivalence. *IEEE Journal on Selected Areas in Communications*, 32(10):1822–1833, 2014.

[19] V. Ramaswami, K. Jain, R. Jana, and V. Aggarwal. Modeling heavy tails in traffic sources for network performance evaluation. In *computational intelligence, cyber security and computational models*, pages 23–44. Springer, 2014.

[20] K. Park and W. Willinger. *Self-similar network traffic and performance evaluation.* A Wiley-Interscience Publication, 2000.

[21] J. Tan and N. B. Shroff. Transition from heavy to light tails in retransmission durations. In *IEEE INFOCOM*, pages 1–9. IEEE, 2010.

[22] P. Wang and I. F. Akyildiz. On the stability of dynamic spectrum access networks in the presence of heavy tails. *IEEE Transactions on Wireless Communications*, 14(2):870–881, 2015.

[23] S. Xia, P. Wang, and Z. Sun. Distributed timely-throughput optimal scheduling for wireless networks. In *IEEE Global Communications Conference 2014*, pages 4820–4826, Austin, TX, USA, December 2014.

REFERENCES (continued)

[24] P. Wang and I.F. Akyildiz. Improving network connectivity in the presence of heavy-tailed interference. *IEEE Transactions on Wireless Communications*, 13(10):5427–5439, 2014.

[25] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transacations on Automatic Control*, 37(12):1936–1949, 1992.

[26] W. Xiong, A. Mukherjee, and H. M. Kwon. MIMO cognitive radio user selection with and without primary channel state information. *IEEE Transactions on Vehicular Technology*, 65(2):985–991, 2016.

[27] J. Liu, N.B. Shroff, C.H. Xia, and H.D. Sherali. Joint congestion control and routing optimization: An efficient second-order distributed approach. *IEEE/ACM Transactions on Networking*, 24(3):1404–1420, 2016.

[28] J. Liu, H.Y. Thomas, Y. Shi, and H.D. Sherali. Cross-layer optimization on routing and power control of MIMO ad hoc networks. *IEEE Journal on Selected Areas of Communications (JSAC)*, 26(6):913–926, 2008.

[29] A. Ribeiro. Stochastic soft backpressure algorithms for routing and scheduling in wireless ad-hoc networks. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 137–140, 2009.

[30] S. Xia and P. Wang. Stochastic network utility maximization in the presence of heavy-tails. In *IEEE International Conference on Communications (ICC)*, pages 1–7, Paris, France, May 2017.

[31] K. Sigman and R.W. Wolff. A review of regenerative processes. *SIAM Review*, 35(2):269–288, 1993.

[32] P. Wang and I. F. Akyildiz. Asymptotic queueing analysis for dynamic spectrum access networks in the presence of heavy tails. *IEEE Journal on Selected Areas in Communications*, 3(3):514–522, 2013.

[33] D.J. Daley and C.M. Goldie. The moment index of minima (ii). *Statistics & Probability Letters*, 76(8):831–837, 2006.

[34] J. Nair, K. Jagannathan, and A. Wierman. When heavy-tailed and light-tailed flows compete: The response time tail under generalized max-weight scheduling. *IEEE/ACM Transactions on Networking*, 24(2):982–995, 2016.

[35] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK, 2004.

REFERENCES (continued)

[36] G. Tychogiorgos, A. Gkelias, and K. Leung. Utility-proportional fairness in wireless networks. In *IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 839–844. IEEE, 2012.

[37] S. Boyd and A. Mutapcic. Subgradient methods. *Lecture notes of EE364b, Stanford University*, 2006.

[38] V. S. Borkar. *Stochastic approximation: A dynamical systems viewpoint*. Hindustan Book Agency, 2008.

[39] K. Eriksson, D. Estep, and C. Johnson. *Applied mathematics: Body and soul: Volume 1: Derivatives and geometry in IR3*. Springer, 2013.

[40] V. Anantharam and V. S. Borkar. Stochastic approximation with long range dependent and heavy tailed noise. *Queueing Systems*, 71(1-2):221–242, 2012.

[41] E. Coddington and N. Levinson. *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.

[42] A. Joulin. On maximal inequalities for stable stochastic integrals. *Potential Analysis*, 26(1):57–78, 2007.

[43] P. Wang and I.F. Akyildiz. Asymptotic queuing analysis for dynamic spectrum access networks in the presence of heavy tails. *IEEE Journal on Selected Areas in Communications*, 31(3):514–522, 2013.

[44] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. *Wireless networks*, 11(4):471–487, 2005.

[45] Y. Shi, T. Hou, J. Liu, and S. Kompella. How to correctly use the protocol interference model for multi-hop wireless networks. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, pages 239–248. ACM, 2009.

[46] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on information theory*, 46(2):388–404, 2000.

[47] S. Keshav. *Mathematical foundations of computer networking*. Wesley, 2012.

[48] S. Xia, P. Wang, and H.M. Kwon. Stochastic network utility maximization in the presence of heavy-tailed traffic. *IEEE/ACM Transactions on Mobile Computing, Submitted*, 2017.

REFERENCES (continued)

[49] F.G. Foster. On the stochastic matrices associated with certain queuing processes. *The Annals of Mathematical Statistics*, pages 355–360, 1953.

[50] B.S. Thomson, J.B. Bruckner, and A.M. Bruckner. *Elementary real analysis*. ClassicalRealAnalysis.com, 2008.

APPENDIXES

APPENDIX A

PROOF OF THEOREM 6

Proof. To prove that the queue length of the LT flow f_L is unbounded under the conventional SGRA, we consider a node i that has both HT and LT flow arrivals. In addition, we assume that the queueing system is in the steady state. Then, the queue-length process is in a positive recurrent regenerative process. Moreover, we let T denote the time interval between two consecutive instances when all queues are empty. Then, $E[T] < \infty$ always holds. More specifically, assume at time slot zero that the HT flow arrival (*i.e.*, f_H) with the smallest tail index, i.e., $\kappa(A_i^{f_H}(t)) = \min_{f \in \mathcal{F}} \kappa(A_i^f(t))$, receives a file of size M number of packets, and all other flows receive no traffic. Furthermore, let T_M denote the first time slot when the flow-queue length difference of flow f_H becomes less or equal to the flow-queue length difference of the flow f_L between nodes i and j. In particular, node j, $j \in N_i$, is the neighbor of node i. Then, we can denote T_M using the following equation:

$$T_M := \min\{t > 0 | Q_i^{f_L}(t) - Q_j^{f_L}(t) \geq Q_i^{f_H}(t) - Q_j^{f_H}(t)\} \qquad i \in N, j \in N_i, (f_L, f_H) \in \mathcal{F}. \tag{A.1}$$

Under Algorithm 1, we know not only that the LT flow will not obtain any service since it has a smaller flow-queue difference than the HT flow, but also that the LT flow will keep receiving new arrival $A_i^{f_L}(t)$ from the external network and packets from last-hop neighbor nodes j, $j \in N_i$. At the same time, the HT flow is always served at $r_{i,j}^{f_H}$ rate until time slot T_M. Thus, for the LT flow, we have

$$Q_i^{f_L}(T_M) = \sum_{t=0}^{T_M - 1} \left(A_i^{f_L}(t) + \sum_{j \in N_i} r_{j,i}^{f_L}(t) \right). \tag{A.2}$$

According to the strong law of large numbers [47, Chapter 1], we have $\sum_{t=0}^{T_M-1}(A_i^{f_L}(t) + \sum_{j \in \mathcal{N}_i} r_{j,i}^{f_L}(t)) \geq (a_i^{f_L} + \sum_{j \in \mathcal{N}_i} r_{j,i}^{f_L})T_M - \delta_M$ and $\sum_{t=0}^{T_M-1} \sum_{j \in \mathcal{N}_i} r_{i,j}^{f_H}(t) \geq \sum_{j \in \mathcal{N}_i} r_{i,j}^{f_H} T_M - \varsigma_M$, with probability 1. Thus, according to step 7 of Algorithm 4, we have the following

relationship, which shows the first time slot the queue length of LT flow larger or equal to the queue length of HT flow, as

$$\big(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L}\big)T_M - \delta_M \geq M - \big(\sum_{j\in N_i} r_{i,j}^{f_H} T_M - \varsigma_M\big), \tag{A.3}$$

where $r_{j,i}^{f_L} = \arg\max_{r_{j,i}^{f_L} = w_s R_{j,i}^s} \sum_{j,i} \big(U(r_{j,i}^{f_L}) + r_{j,i}^{f_L}(Q_j^{f_L} - Q_i^{f_L})\big)$ and $r_{i,j}^{f_H} = \arg\max_{r_{i,j}^{f_H} = w_s R_{i,j}^s} \sum_{i,j} \big(U(r_{i,j}^{f_H}) + r_{i,j}^{f_H}(Q_i^{f_H} - Q_j^{f_H})\big)$. In addition, there exists a constant K, such that

$$\begin{aligned} T_M &\geq \frac{M + \varsigma_M + \delta_M}{a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L} + \sum_{j\in N_i} r_{i,j}^{f_H}} \\ &\geq \frac{M}{a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L} + \sum_{j\in N_i} r_{i,j}^{f_H}} = KM, \end{aligned} \tag{A.4}$$

where $K = \frac{1}{a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L} + \sum_{j\in N_i} r_{i,j}^{f_H}}$ and $\delta_M, \varsigma_M > 0$. In addition, by the property of the regenerative process with cycle length T, we have

$$\begin{aligned} &\Pr(Q_i^{f_L} > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2} \\ &= \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{I}_{\{Q_i^{f_L}(\tau) > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2}\}} \\ &= \frac{E\Big[\sum_{t=0}^{T} \mathbb{I}_{\{Q_i^{f_L}(t) > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2}\}}\Big]}{E[T]}. \end{aligned} \tag{A.5}$$

Then, we obtain

$$\begin{aligned} &E\Big[\sum_{t=0}^{T} \mathbb{I}_{\{Q_i^{f_L}(t) > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2}\}}\Big] \\ &\geq E[\mathbb{I}(A_i^{f_H}(0) > M) \cdot \sum_{t=0}^{T} \mathbb{I}(Q_i^{f_L}(t) > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2})] \\ &\geq \Pr(A_i^{f_H}(0) > M) \cdot \sum_{t=\frac{T_M}{2}}^{T_M} \Pr(Q_i^{f_L}(t) > \frac{KM(a_i^{f_L} + \sum_{j\in N_i} r_{j,i}^{f_L})}{2}). \end{aligned} \tag{A.6}$$

By the queueing dynamic in equation (4.5) and $Q_i^{f_L}(0) = 0$, we have $Q_i^{f_L}(t) = \sum_{\tau=0}^{t-1}[A_i^{f_L}(\tau) - \sum_{j\in N_i}\big(r_{i,j}^{f_L}(\tau) - r_{j,i}^{f_L}(\tau)\big)\mathbb{I}_{\{Q_i^{f_L}(\tau)>0\}}]$. This implies that

$$\begin{aligned}
&\lim_{M\to\infty}\frac{1}{M}\sum_{t=\frac{T_M}{2}}^{T_M}\Pr(Q_i^{f_L}(t) > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2}) \\
&\geq \lim_{M\to\infty}\frac{1}{M}\sum_{t=\frac{KM}{2}}^{KM}\Pr\big(\sum_{\tau=0}^{t-1}A_i^{f_L}(\tau) > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2}\big) \\
&= \frac{K}{2}. \qquad \text{(A.7)}
\end{aligned}$$

The last equality in equation (A.7) holds, due to the fact that flow f_H occupies the entire channel service during the time interval $0 \leq t \leq KM/2$ so that $\Pr\big(\sum_{\tau=0}^{t-1}A_i^{f_L}(\tau) > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2}\big) = 1$ when $KM/2 \leq t \leq KM$ always holds. Then, we have

$$\begin{aligned}
&Pr(Q_i^{f_L} > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2}) \\
&\geq \frac{\Pr(A_i^{f_H}(0) > M)}{E[T]}\cdot\frac{\sum_{t=\frac{T_M}{2}}^{T_M}\Pr(Q_i^{f_L}(t) > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2})}{E[T]}, \qquad \text{(A.8)}
\end{aligned}$$

according to Nair et al. [34, page 7], and combining equation (A.5) with equations (A.6)-(A.7), it follows from the condition in equation (4.20) that

$$\lim_{M\to\infty}\frac{\log[\Pr(Q_i^{f_L} > \frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2})]}{\log[\frac{KM(a_i^{f_L}+\sum_{j\in N_i} r_{j,i}^{f_L})}{2}]} \geq -\min_{f\in\mathcal{F}}\kappa(A_i^f(t)) + 1 \geq -1. \qquad \text{(A.9)}$$

By applying the moment theorem [33], $E[Q_i^{f_L}]$ is infinite. Thus, the moment stability cannot be achieved under the classic stochastic gradient algorithm. □

APPENDIX B

PROOF OF THEOREM 7

Proof. To explore our algorithm of convergence and stability properties and prove that it is utility optimal, we adopt a technique similar to that in the work of Xia et al. [48], which uses the ordinary differential equation approach.

From equation (4.46), we obtain vector $\lambda^f(t+1)$:

$$\begin{aligned}\lambda^f(t+1) &= \lambda^f(t) - \alpha_2(t)g^f(t) \\ &= \lambda^f(t) + \alpha_2(t)\big(a_i^f - \sum_{j\in N_i}(r_{i,j}^f(t) - r_{j,i}^f(t)\big) + \alpha_2(t)\big(A^f(t) - a_i^f)\big) \\ &= \lambda^f(t) + \alpha_2(t)(h^f(t) + A'^f(t)), \end{aligned} \tag{B.1}$$

where $h^f(t) := a_i^f - \sum_{j\in\mathcal{N}_i}\big(r_{i,j}^f(t) - r_{j,i}^f(t)\big)$ and $\hat{A}^f(t) := A^f(t) - a_i^f$. Since, according to the first constraint of equation (4.11), we know $h^f(t)$ is larger than some negative constant c and smaller than 0, it is in a strong form of uniform continuity in the range [c, 0], i.e., Lipschitz [39, Section 12.3]. Moreover, for stochastic gradient algorithms, the associated ODE is $\dot{\lambda}^f(t) = -g^f(\lambda^f(t))$. Thus, we have the following relationship

$$\dot{\lambda}^f(t) = h^f(\lambda^f(t)). \tag{B.2}$$

Then, we define $\bar{\lambda}^f(j(t)) = \lambda^f(t)$ as a discretely interpolated version of $\lambda^f(t)$, where $j(t) = \sum_{k=0}^{t-1}\alpha_2(k)$, and $\lambda_t^f(j)$, $j \geq j(t)$ as a continuous ODE trajectory with $\lambda_t^f(j(t)) = \bar{\lambda}^f(j(t)) := \lambda^f(t)$, which leads both the discrete and continuous versions to have the same value at $j(t)$. In addition, we let $n = \inf\{k : j(k) > j(t) + T\}$ and $T = j(n) - j(t)$. Thus, to find the supreme of the norm difference $\sup_{t\leq k\leq n}||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))||$ when $t \leq k \leq n$, we rewrite $\bar{\lambda}^f(j(k))$ and $\lambda_t^f(j(k))$, respectively, as

$$\bar{\lambda}^f(j(k)) = \bar{\lambda}^f(j(t)) + \sum_{i=t}^{k-1}\alpha_2(i)\big(h^f(\bar{\lambda}^f(j(i)) + A'^f(i)\big) \tag{B.3}$$

and

$$\begin{aligned}\lambda_t^f(j(k)) &= \bar{\lambda}^f(j(t)) + \int_{j(t)}^{j(k)} h^f(\lambda_t^f(v))dv \\ &= \bar{\lambda}^f(j(t)) + \sum_{i=t}^{k-1}\Big\{\alpha_2(i)h^f(\lambda_t^f(j(i))) + \int_{j(i)}^{j(i+1)}\big(h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i))\big)dv\Big\},\end{aligned} \tag{B.4}$$

where $\int_{j(i)}^{j(i+1)} dv = \alpha_2(i)$. Then, we subtract equation (B.4) from equation (B.3), obtaining

$$\begin{aligned}&\sup_{t\le k\le n} ||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))|| \\ &\le \sup_{t\le k\le n}\sum_{i=t}^{k-1}\int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i)))||dv + \sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)A'^f(i)||.\end{aligned} \tag{B.5}$$

Then, we let

$$I = \sup_{t\le k\le n}\sum_{i=t}^{k-1}\int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i)))||dv \tag{B.6}$$

and

$$II = \sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)A'^f(i)||. \tag{B.7}$$

For term I, according to equation (B.4), we know that

$$\lambda_t^f(j) = \lambda^f(t) + \int_{j(t)}^{j(n)} h^f(\lambda_t^f(v))dv, \tag{B.8}$$

where $j \in [j(t), j(n)]$, and $\overline{\lambda}^f(j(t)) = \lambda^f(t)$. Then, we take the norm on both sides of equation (B.8) and obtain

$$||\lambda_t^f(j)|| \le ||\lambda^f(t)|| + \int_{j(t)}^{j(n)} ||h^f(\lambda_t^f(v))||dv. \tag{B.9}$$

Since h^f is Lipschitz continuous and grows linearly, we have $||h^f(x) - h^f(0)|| \le L||x||$ and $||h^f(x)|| \le ||h^f(0)|| + L||x||$, where $L > 0$ denotes the Lipschitz constant. Then, we obtain

$$||h^f(\lambda_t^f(v))|| \le ||h^f(0)|| + L||\lambda_t^f(v)||. \tag{B.10}$$

By defining $B_0 = \lambda_t^f(j(t)) = \lambda^f(t) \leq \sup_{t \in N} ||\lambda^f(t)||$, we can rewrite equation (B.9) as

$$\begin{aligned} ||\lambda_t^f(j)|| &\leq ||\lambda^f(t)|| + \int_{j(t)}^{j(n)} ||h^f(0) + L(\lambda_t^f(v))||dv \\ &\leq (B_0 + T||h^f(0)||) + L\int_{j(t)}^{j(n)} ||h^f(\lambda_t^f(v))||dv. \end{aligned} \tag{B.11}$$

In addition, by using Gronwall's inequality [41, Section 1.6], we further have

$$\begin{aligned} ||\lambda_t^f(j)|| &\leq (B_0 + T||h^f(0)||)e^{L\int_{j(t)}^{j(n)} ||h^f(\lambda_t^f(v))||dv} \\ &\leq (B_0 + T||h^f(0)||)e^{LT} \end{aligned} \tag{B.12}$$

where since $h^f(\cdot) = 1$, $\int_{j(t)}^{j(b)} ||h^f(\lambda_t^f(v))||dv = T$ always holds. Then, according to the Lipschitz continuity,

$$\begin{aligned} ||h^f(\lambda_t^f(j))|| &\leq ||h^f(0)|| + L||\lambda_t^f(j)|| \\ &\leq ||h^f(0)|| + L(B_0 + T||h^f(0)||)e^{LT} \\ &= B, \end{aligned} \tag{B.13}$$

where $B := ||h^f(0)|| + L(B_0 + T||h^f(0)||)e^{LT}$. Moreover, when $j \in [j(i), j(i+1)]$, we have

$$||\lambda_t^f(j) - \lambda_t^f(j(i))|| \leq \int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v))||dv \leq B\alpha_2(i). \tag{B.14}$$

Since h^f has Lipschitz continuity, we obtain the following relationship:

$$\int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i)))||dv \leq BL(\alpha_2(i))^2. \tag{B.15}$$

Then, we take the summation on both sides of equation (B.15) and obtain

$$\sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i)))||dv \leq \sum_{i=t}^{k-1} BL(\alpha_2(i))^2, \tag{B.16}$$

and by taking the limitation on both sides of equation (B.16), we have

$$\lim_{t\to\infty} \sup_{t\leq k\leq n} \sum_{i=t}^{k-1} \int_{j(i)}^{j(i+1)} ||h^f(\lambda_t^f(v)) - h^f(\lambda_t^f(j(i)))|| \leq \lim_{t\to\infty} \sum_{i=t}^{k-1} BL(\alpha_2(i))^2 = 0, \tag{B.17}$$

where $\lim_{t\to\infty}\sum_{i=t}^{k-1}(\alpha_2(i))^2 = 0$. Thus, equation (B.17) holds.

For II, according to Joulin [42], we have

$$P(\sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)\hat{A}^f(i)|| \ge x) \le \frac{K\big(\sum_{i=t}^{n}(\alpha_2(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{\beta}{\beta+1}}}{x^\beta} \tag{B.18}$$

for $x > K\big(\sum_{i=t}^{n}(\alpha_2(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{1}{\beta+1}}$, where β is larger than 1. We define the following:

$$\begin{aligned}
\mu(t) &:= K\big(\sum_{i=t}^{n}(\alpha_2(i))^{\frac{\beta^2-1}{\beta}+1}\big)^{\frac{1}{\beta+1}} \\
&= K\big(\sum_{i=t}^{n}\alpha_2(i)^{\frac{\beta^2-1}{\beta}}\alpha_2(i)\big)^{\frac{1}{\beta+1}} \\
&= K\big(\alpha_2(t)^{\frac{\beta^2-1}{\beta}}\alpha_2(t) + \alpha_2(t+1)^{\frac{\beta^2-1}{\beta}}\alpha_2(t+1), ..., +\alpha_2(n)^{\frac{\beta^2-1}{\beta}}\alpha_2(n)\big)^{\frac{1}{\beta+1}} \\
&\le K\big((T+1)\alpha_2(t)^{\frac{\beta^2-1}{\beta}}\alpha_2(t)\big)^{\frac{1}{\beta+1}} \\
&\le K(T+1)^{\frac{1}{\beta+1}}\alpha_2(t)^{\frac{\beta-1}{\beta}}
\end{aligned} \tag{B.19}$$

When $t\to\infty, \mu(t)\to 0$. Furthermore, we use a similar method in the work of Anantharam

and Borkar [40], for $1 < \varepsilon < \beta$, and we obtain the following relationship:

$$
\begin{aligned}
&E[\sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)\hat{A}^f(i)||^{\varepsilon}] \\
&\le K\int_0^{\infty} x^{\varepsilon-1}P\big(\sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)\hat{A}^f(i)||\ge x\big)dx \\
&= K\int_0^{\mu(t)} x^{\varepsilon-1}P\big(\sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)\hat{A}^f(i)||\ge x\big)dx \\
&\qquad\qquad + K\int_{\mu(t)}^{\infty} x^{\varepsilon-1}P\big(\sup_{t\le k\le n}||\sum_{i=t}^{k-1}\alpha_2(i)\hat{A}^f(i)||\ge x\big)dx \\
&\le K\mu(t)^{\varepsilon} + K\int_{\mu(t)}^{\infty} x^{\varepsilon-1}\big(\frac{\mu(t)^{\beta}}{x^{\beta}}\big)dx \\
&= K\mu(t)^{\varepsilon} + K\mu(t)^{\beta}\int_{\mu(t)}^{\infty} x^{\varepsilon-\beta-1}dx \\
&= K\mu(t)^{\varepsilon} + K\mu(t)^{\beta}\frac{1}{\varepsilon-\beta-1}x^{\varepsilon-\beta}|_{\mu(t)}^{\infty} \\
&= K\mu(t)^{\varepsilon} - \frac{1}{\varepsilon-\beta-1}K\mu(t)^{\varepsilon} \\
&= \hat{K}\mu(t)^{\varepsilon}. \qquad\qquad \text{(B.20)}
\end{aligned}
$$

where $\hat{K} = K(1 - \frac{1}{\varepsilon-\beta-1})$. Moreover, according to equation (B.19), we know that when $t \to \infty$, $T_{II} \to 0$. Then, combining equations (B.17) and (B.20), we obtain

$$\lim_{t\to\infty}\sup_{t\le k\le n}||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))|| = 0. \qquad \text{(B.21)}$$

Considering the linear interpolation error [38, Section 2.1], if $j(k) \le j \le j(k+1)$, then

$$\bar{\lambda}^f(j) = \kappa\bar{\lambda}^f(j(k)) + (1-\kappa)\bar{\lambda}^f(j(k+1)) \qquad \text{(B.22)}$$

for some $\kappa \in [0, 1]$. Thus,

$$||\bar{\lambda}^f(j) - \lambda_t^f(j)|| = ||\kappa(\bar{\lambda}^f(j) - \lambda_t^f(j(k))) + (1-\kappa)(\bar{\lambda}^f(j) - \lambda_t^f(j(k+1)))||. \qquad \text{(B.23)}$$

In addition, equation (B.23) can be upper bounded as

$$\begin{aligned}
&||\bar{\lambda}^f(j) - \lambda_t^f(j)|| \\
&\leq \kappa||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))|| + (1-\kappa)||\bar{\lambda}^f(j(k+1)) - \lambda_t^f(j(k+1))|| \\
&\qquad + \kappa \int_{j(k)}^{j} ||h^f(\lambda_t^f(v))||dv + (1-\kappa)\int_{j}^{j(k+1)} ||h^f(\lambda_t^f(v))||dv.
\end{aligned} \tag{B.24}$$

Thus, we obtain

$$\lim_{t\to\infty} \sup_{j\in[j(t),j(n)]} ||\bar{\lambda}^f(j) - \lambda_t^f(j)|| \leq \lim_{t\to\infty} \sup_{t\leq k\leq n} ||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))|| = 0. \tag{B.25}$$

By taking the εth mean on both sides of equation (B.25), we get

$$\lim_{t\to\infty} E\left[\sup_{j\in[j(t),j(n)]} ||\bar{\lambda}^f(j) - \lambda_t^f(j)||^{\varepsilon}\right] \leq \lim_{t\to\infty} E\left[\sup_{t\leq k\leq n} ||\bar{\lambda}^f(j(k)) - \lambda_t^f(j(k))||^{\varepsilon}\right] = 0, \tag{B.26}$$

where $1 < \varepsilon < \beta$. Then, according to Anantharam and Borkar [40], we know that

$$\lim_{t\to\infty} E\big[||\lambda^f(t) - \lambda^*||^{\varepsilon}\big] = 0 \tag{B.27}$$

which implies that

$$\lim_{t\to\infty} E[r^f(t) - r^{*f}] = 0. \tag{B.28}$$

Furthermore, we can conclude that

$$\lim_{t\to\infty} \Big(U(E[r^f(t)]) - U(r^{*f})\Big) = 0 \tag{B.29}$$

Thus, we can conclude that the TA-SGRA is utility optimal, since the utility converges to a optimal value U^* as time proceeds. □

APPENDIX C

PROOF OF THEOREM 8

Proof. To prove that the time-average transmission rate converges to a constant, we use the same technique [30]. Then, considering strong convexity [35], we have

$$f(y) - f(x) \leq \left\langle f^{'}(x), y - x \right\rangle + \frac{G}{2}||y - x||^2 \tag{C.1}$$

where $< \cdot, \cdot >$ denotes the inner product, and G is a positive constant. Moreover, since $\lambda^f(t+1) = \lambda^f(t) - \alpha_2(t)g^f(t)$ and according to equation (C.1), we obtain the following relationship:

$$\begin{aligned} D(\lambda^f(t+1)) &\leq D(\lambda^f(t)) + \left\langle D^{'}(\lambda^f(t)), \lambda^f(t+1) - \lambda^f(t) \right\rangle + \frac{L}{2}||\lambda^f(t+1) - \lambda^f(t)||^2 \\ &= D(\lambda^f(t)) - \alpha_2(t)||D^{'}(\lambda^f(t))||^2 + \frac{L}{2}\alpha_2(t)^2||D^{'}(\lambda^f(t))||^2 \\ &= D(\lambda^f(t)) - \big(1 - \frac{L}{2}\alpha_2(t)\big)\alpha_2(t)||D^{'}(\lambda^f(t))||^2. \end{aligned} \tag{C.2}$$

In addition, since $g^f(t) = D^{'}(\lambda^f(t), t)$, we have

$$E[D(\lambda^f(t)) - D(\lambda^f(t+1))] \geq \big(1 - \frac{L}{2}\alpha_2(t)\big)\alpha_2(t)E[||g^f(t)||]^2. \tag{C.3}$$

According to Theorem 2, when $t \to \infty$, we obtain

$$\lim_{t\to\infty} E[D(\lambda^f(t)) - D(\lambda^f(t+1))] = \lim_{t\to\infty} E[D(\lambda^f(t)) - D^*] = 0, \tag{C.4}$$

where D^* is the optimal value of the dual function. Therefore, from the first-order optimality condition, we can conclude that

$$\lim_{t\to\infty} E[||g^f(\lambda^f(t))||] = 0. \tag{C.5}$$

Since strong duality exists, the minimum value of the dual function is the optimal solution of the primal function. Then, we obtain

$$U(r^{*f}) = D^*, \tag{C.6}$$

and thus,

$$r^{*f} = U^{-1}(D^*). \tag{C.7}$$

In addition, from equation (4.43), we have

$$\overline{r}^f(t) = \frac{t-1}{t}\overline{r}^f(t-1) + \frac{1}{t}r^f(t), \quad t \geq 1. \tag{C.8}$$

and

$$t\overline{r}^f(t) - (t-1)\overline{r}^f(t-1) - r^f(t) = 0. \tag{C.9}$$

Furthermore, we take expectations on both sides of equation (C.9) to obtain

$$E[t\overline{r}^f(t) - (t-1)\overline{r}^f(t-1)] = E[r^f(t)]. \tag{C.10}$$

Then, when $t \to \infty$, taking the limitation on both sides of equation (C.10), we have the following relationship:

$$\lim_{t\to\infty} E[t\overline{r}^f(t) - (t-1)\overline{r}^f(t)] = \lim_{t\to\infty} E[r^f(t)]. \tag{C.11}$$

From equation (B.28), we know $\lim_{t\to\infty} E[r^f(t)] = r^{*f}$. Then, we have

$$\lim_{t\to\infty} E[\bar{r}^f(t)] = r^{*f}. \tag{C.12}$$

In addition, when $t \to \infty$, the $\lim_{t\to\infty} \bar{r}^f(t) = \lim_{t\to\infty} E[\bar{r}^f(t)]$. Therefore, we obtain

$$\lim_{t\to\infty} \overline{r}^f(t) = r^{*f}. \tag{C.13}$$

Thus, we have

$$\lim_{t\to\infty} \overline{r}^f(t) = U^{-1}(D^*) \tag{C.14}$$

which completes the proof. □

APPENDIX D

PROOF OF THEOREM 10

Proof. To prove that the link queue $q_{i,j}$ has a bounded mean, we adopt the Lyapunov drift theory. First, we define a quadratic Lyapunov function as

$$L(q(t)) = \sum_{i,j} q_{i,j}^2(t). \tag{D.1}$$

Then, we consider

$$\begin{aligned} & L(q(t+1)) - L(q(t)) \\ &= \sum_{i,j} (q_{i,j}(t+1) - q_{i,j}(t))(q_{i,j}(t+1) - q_{i,j}(t) + 2q_{i,j}(t)) \\ &= \sum_{i,j} (q_{i,j}(t+1) - q_{i,j}(t))^2 + \sum_{i,j} 2q_{i,j}(t)(q_{i,j}(t+1) \\ &\qquad\qquad - q_{i,j}(t)). \end{aligned} \tag{D.2}$$

Taking conditional expectations on equation (D.2), we obtain

$$\begin{aligned} & E[L(q(t+1)) - L(q(t))|q(t)] \\ &= E[\sum_{i,j} (q_{i,j}(t+1) - q_{i,j}(t))^2 | q(t)] + E[\sum_{i,j} 2q_{i,j}(t)(q_{i,j}(t+1) - q_{i,j}(t))]. \end{aligned} \tag{D.3}$$

Then, we let

$$I = E[\sum_{i,j} (q_{i,j}(t+1) - q_{i,j}(t))^2 | q(t)] \tag{D.4}$$

$$II = E[\sum_{i,j} 2q_{i,j}(t)(q_{i,j}(t+1) - q_{i,j}(t))]. \tag{D.5}$$

For I, from equation (4.10), we have

$$\begin{aligned} I &= E[\sum_{i,j} (\overline{r}_{i,j}(t) - \sum_{s \in S} W_s(t) R_{i,j}^s)^2 | q(t)] \\ &\leq E[\sum_{i,j} \overline{r}_{i,j}^2(t) | q(t)] \\ &= \sum_{i,j} E[\overline{r}_{i,j}^2(t)]. \end{aligned} \tag{D.6}$$

APPENDIX D (continued)

For II, using equation (4.10), we obtain

$$\begin{aligned} II &= E[\sum_{i,j} 2q_{i,j}(t)(\overline{r}_{i,j}(t) - \sum_{s\in S} W_s(t)R^s_{i,j})|q(t)] \\ &= E[\sum_{i,j} 2q_{i,j}(t)\overline{r}_{i,j}(t)|q(t)] - E[\sum_{i,j} 2q_{i,j}(t)\sum_{s\in S} W_s(t)R^s_{i,j}]. \end{aligned} \tag{D.7}$$

Then, we define

$$III = E[\sum_{i,j} 2q_{i,j}(t)\overline{r}_{i,j}(t)|q(t)] \tag{D.8}$$

$$IV = E[\sum_{i,j} 2q_{i,j}(t)\sum_{s\in S} W_s(t)R^s_{i,j}]. \tag{D.9}$$

For III, we have

$$III = 2q_{i,j}(t)r_{i,j}. \tag{D.10}$$

For IV, we get

$$IV = 2\sum_{i,j} q_{i,j}(t)\sum_{s\in S} w_s R^s_{i,j} \tag{D.11}$$

Therefore, we have

$$E[\sum_{i,j} 2q_{i,j}(t)(q_{i,j}(t+1) - q_{i,j}(t))] = 2\sum_{i,j} q_{i,j}(t)(r_{i,j} - \sum_{s\in S} w_s R^s_{i,j}). \tag{D.12}$$

According to the second constraint of equation (4.11), we can bound equation (D.12) as follows:

$$E[\sum_{i,j} 2q_{i,j}(t)(q_{i,j}(t+1) - q_{i,j}(t))] \leq 2\xi\sum_{i,j} q_{i,j}(t), \tag{D.13}$$

where ξ is a small constant. Then, from equations (D.3), (D.6), and (D.13), we have

$$E[L(q(t+1)) - L(q(t))|q(t)] \leq \sum_{i,j} E[\overline{r}^2_{i,j}(t)] + 2\xi\sum_{i,j} q_{i,j}(t). \tag{D.14}$$

Using Foster's criterion [49] for the ergodic Markov chain, the link queue-length process converges in distribution. Using the iterated mean and telescoping sums [50, Section 3.4], we have

$$\sum_{i,j} E[q_{i,j}(t)] \leq -\frac{1}{2\xi} \sum_{i,j} E[\overline{r}^2_{i,j}(t)], \tag{D.15}$$

which completes the proof. □