A Risk-Sensitive Approach for Packet Inter-Delivery Time Optimization in Networked Cyber-Physical Systems

Xueying Guo, *Member, IEEE*, Rahul Singh, P. R. Kumar, *Fellow, IEEE, ACM*, and Zhisheng Niu, *Fellow, IEEE*

***Abstract*—In networked cyber-physical systems, the inter-delivery time of data packets becomes an important quantity of interest. However, providing a guarantee that the inter-delivery times of the packets are "small enough" becomes a difficult task in such systems due to the unreliable communication medium and limited network resources. We design scheduling policies that meet the inter-delivery time requirements of multiple clients connected over wireless channels. We formulate the problem as an infinite-state risk-sensitive Markov decision process, where large exceedances of inter-delivery times for different clients over their design thresholds are severely penalized. We reduce the infinite-state problem to an equivalent finite-state problem and establish the existence of a stationary optimal policy and an algorithm for computing it in a finite number of steps. However, its computational complexity makes it intractable when the number of clients is of the order of 100 or so that is found in applications such as in-vehicle networks. To design computationally efficient optimal policies, we, therefore, develop a theory based on the high reliability asymptotic scenario, in which the channel reliability probabilities are close to one. We thereby obtain an algorithm of relatively low computational complexity for determining an asymptotically optimal policy. To address the remaining case when the channels are not relatively reliable, we design index-based policies for the risk sensitive case, which extends key ideas for index policies in risk-neutral multi-armed bandit problems. Simulation results are provided to show the effectiveness of our policies.**

***Index Terms*—Scheduling policy, packet inter-delivery time, high reliability asymptotic approach, index-based policy.**

I. Introduction

In CYBER-PHYSICAL systems, where physical processes are often monitored and controlled by embedded computers through feedback control over a network [2], the traditional quality of service (QoS) metrics such as delay and throughput, that are often used to judge the effectiveness of data networks [3], [4], are inadequate. As an example, a low value of end-to-end latency does not imply the absence of long time-gaps between successive packet deliveries. A recent metric of interest that has attracted attention for networks transporting sensor measurements for control, such as for in-vehicular wireless network shown in Figure 1, is the packet inter-delivery time [1], [5]–[9]. Figure 1 illustrates an in-vehicular network. In such a cyber-physical system, there are of the order of a hundred wireless sensor nodes that monitor processes such as pressure and temperature. The sensor measurements thus obtained are then transmitted to controllers so that the actuator signals can be chosen appropriately. A key distinguishing feature is that the arrival process, which corresponds to sensor measurement generations, can be controlled. When a packet transmission attempt fails, the sensor can re-sample the physical process and attempt to transmit a fresh measurement next time, instead of retransmitting the same measurement. Thus, delay experienced by packets is not the right performance measure. A more appropriate measure is the time-gap between successive packet deliveries of sensor measurements, since long gaps might cause the underlying physical processes to suffer from poor control or even instability. In this paper, we consider the problem of optimizing data networks with respect to the regularity of the inter-delivery times of measurement packets. Our goal is the design of a scheduling policy for making new measurements and transmitting the resulting data packets, which ensures that the deviations of packet inter-delivery times exceeding specified thresholds have small expected value.

We formulate the problem of obtaining an optimal policy as an infinite-state risk-sensitive Markov Decision Process (MDP) [10]–[15]. Different clients may have heterogeneous packet inter-delivery time regularity requirements manifested through different inter-delivery time thresholds τ_n. In order to severely penalize large deviations of inter-delivery times exceeding their corresponding thresholds, we employ

Manuscript received November 15, 2016; revised September 4, 2017 and February 11, 2018; accepted July 2, 2018; approved by IEEE/ACM Transactions on Networking Editor L. Ying. Date of publication August 1, 2018; date of current version August 16, 2018. This work was supported in part by the National Nature Science Foundation of China under Grant 61861136003, Grant 91638204, Grant 61571265, and Grant 61621091, in part by the NSF under Grant CNS-1719384, in part by the Science and Technology Center under Grant CCF-0939370, Grant CCF-1619085, and Grant ECCS-1646449, in part by the Office of Naval Research under Grant N00014-18-1-2048, in part by the U.S. Army Research Office under Grant W911NF-15-1-0279, in part by the NPRP through the Qatar National Research Fund, a member of the Qatar Foundation, under Grant NPRP 8-1531-2-651, and in part by Hitachi Ltd. Part of this work was published in the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing [1]. *(Corresponding author: Zhisheng Niu.)*

X. Guo was with Tsinghua University, Beijing 100084, China. She is now with the Department of Computer Science, University of California at Davis, Davis, CA 95616 USA (e-mail: guoxueying@outlook.com).

R. Singh was with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with Intel Corporation, Santa Clara, CA 95054 USA (e-mail: rahulsiitk@gmail.com).

P. R. Kumar is with Texas A&M University, College Station, TX 77843 USA (e-mail: prk.tamu@gmail.com).

Z. Niu is with the Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: niuzhs@tsinghua.edu.cn).

This paper has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNET.2018.2856883

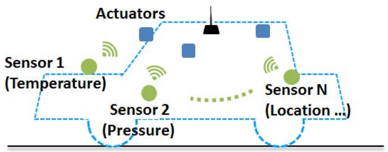


Fig. 1. An in-vehicular network with an access point and several wirelessly connected sensors and actuators.

an exponential cost function, $\mathrm{E}[\exp(\theta \cdot \sum_{n=1}^{N}(D^{(n)} - \tau_n)^+)]$, where $D^{(n)}$ is the inter-delivery time of client n, $\theta > 0$ is a risk-aversion parameter, and $(a)^+ := \max\{0, a\}$.

We begin by showing that though the above problem involves an infinite number of states, it can be solved via a finite-state MDP. We prove that there exists a stationary policy that is optimal for the original problem, and provide an algorithm that determines the optimal policy in a finite number of steps. However, this algorithm suffers from significant computational complexity when the number of sensors is of the order of a hundred nodes, as is the case in in-vehicular networks.

In order to design a policy that can be implemented in an online fashion, we analyze the "high reliability asymptote" when the channel failure probabilities for different clients are of the same order, and asymptotically approach zero. The resulting asymptotically optimal policies are expected to have near-optimal performance when the channel failure probabilities are small or even moderate. This asymptotic approach has a similar motivation as the study of high SNR asymptotics in network information theory [16]. It yields a policy that can be obtained with relatively low computational complexity. Extensive simulations show that the resulting policy performs well even in the pre-asymptotic regime, justifying the effectiveness of this approach. We also address the scheduling problem in the scenario when the channel reliabilities are arbitrary and not necessarily close to one. We generalize the well studied and easily computable "index policies" [17] for the risk neutral setting to the problems of interest. Finally, to provide a qualitative insight into the nature of policies produced by this approach, we thoroughly analyze the simple two-client scenario and show that it results in a certain "modified-least-time-to-go" policy that is both structurally clean and asymptotically optimal.

The rest of the paper is organized as follows. We present an account of related work in Section II, and the system model in Section III. In Section IV, we exhibit the equivalent finite-state problem. We establish the algorithm to determine the stationary optimal policy for the risk-sensitive MDP approach in Section V. We develop the high reliability asymptotic approach in Section VI, and design the asymptotically optimal policies in Section VII. In Section VIII, we derive the index policies for non-asymptotic cases. We provide the results of simulation testing in Section IX, and conclude in Section X.

II. Related Work

To the best of the author's knowledge, Li *et al.* [5] and Li *et al.* [18] are the first to regard inter-delivery time as a performance metric in packet transmission. They consider inter-delivery time in a different queueing system scenario, where the outdated packets cannot be replaced by newer packets. The periodic nature of wireless packet transmissions has drawn increasing attention recently, as in Sadi and Ergen [19]. Singh *et al.* [6] have addressed the trade-off between throughput and variation in the inter-delivery times. Guo *et al.* [9] have designed scheduling policies that jointly optimize inter-delivery time and transmission power consumption. However, in this work, there is no modeling effort to severely penalize the large inter-delivery exceedance. Singh and Stolyar [20] have investigated service regularity under the Max Weight policy, and shown that the service process is asymptotically smooth.

Risk-sensitive MDPs introduced by Howard and Matheson [10], have attracted increasing attention [12]–[14], [21]. However, unlike risk-neutral MDPs, a risk-sensitive MDP need not have a stationary optimal policy even with a discounted cost [11], which further complicates our problem. Risk-sensitive MDPs have also been employed in wireless communications. For example, Altman *et al.* [22] have employed a risk-sensitive MDP model to study power control strategies in delay tolerant networks.

Our high-reliability asymptotic approach is philosophically similar to the high SNR asymptotics studied in the field of network information theory. Avestimehr *et al.* [16] have obtained constant gap approximations to the capacity of wireless networks based on a deterministic channel model, which yields near optimal communication schemes for Gaussian relay systems. Kittipiyakul *et al.* [23] and Zhang and Tepedelenlioglu [24] have further analyzed error performance in fading channels by employing such an asymptotic approach.

The high reliability asymptotic approach may be compared with the periodic scheduling approach [25], [26], which focuses on the scenario of exactly zero failure probabilities. However, in contrast to the asymptotic approach proposed and analyzed in this paper, there is no guarantee that an optimal periodic scheduling policy will have good performance under non-zero but small failure probabilities. In fact, as we show in Section IX, periodic scheduling policies can have extremely bad performance even with rather small failure probabilities.

III. System Model

Consider a wireless network comprising of one access point (AP) and N sensors. Time is discretized with the network evolving over time slots indexed by $t = 1, 2, \ldots$. At the beginning of each time slot, the AP broadcasts a control signal that indicates the index of the sensor that is supposed to transmit a packet in that time-slot. This eliminates the problem of multiple clients attempting to transmit packets simultaneously, which leads to packet losses due to collisions. In Section VIII, we generalize this problem to the case where several orthogonal wireless channels are used. The probability that a packet sent by client n is delivered to the AP is p_n, and is called the channel reliability of client n. The model can be extended by considering fading channels.

Each client $n = 1, 2, \ldots, N$ has an inter-delivery threshold τ_n modeling the inter-delivery requirement of client n. The system cost incurred over T time-steps is modeled as,

$$\mathrm{E}\left[\exp\left(\theta \sum_{n=1}^{N} \sum_{i=1}^{M_T^{(n)}} (D_i^{(n)} - \tau_n)^+ + (T - t_{D^{(n)}_{M_T^{(n)}}} - \tau_n)^+\right)\right] \tag{1}$$

where $D_i^{(n)}$ is the time between the $(i-1)$-th and i-th packet deliveries of client n, $M_T^{(n)}$ is the number of packets delivered for client n by time T, $t_{D_i^{(n)}}$ is the time slot in which the i-th packet for client n is delivered, and $(a)^+ := \max\{a, 0\}$. The last term is included since, otherwise, the policy of never transmitting any packet at all will result in the least cost. This is because if there is no successful transmissions of client n in an interval of T time-steps, the total number of inter-delivery times will be 0, and thus leading to the first term $\sum_{i=1}^{M_T^{(n)}} (D_i^{(n)} - \tau_n)^+ = 0$. The parameter $\theta > 0$ is a risk-aversion parameter. The larger the value of θ, the more severely are the exceedances of inter-delivery times over the thresholds penalized. At each time slot t, the scheduling policy decides which client should transmit a packet, so as to minimize the above cost.

Now, we formulate the system as a risk-sensitive MDP. The system *state* at time-slot t is denoted by,

$$X(t) := (X_1(t), \dots, X_N(t)),$$

where $X_n(t)$ is the time elapsed since the previous delivery of a packet from client n. Consequently, the *state space* is $\{0, 1, \dots\}^N$. Although the state-space is finite when a finite time horizon problem is considered, it exponentially grows to infinity as the horizon increases. Let $U(t)$ denote the system *control* in time slot t; it specifies the client that is to transmit in slot t. Then, the system state evolves as

$$X_n(t+1) = \begin{cases} 0, & \text{if a packet is delivered for client } n \text{ in } t; \\ X_n(t)+1, & \text{otherwise.} \end{cases}$$

Thus, the system can be described by a controlled Markov Chain [27], with associated transition probabilities

$$\mathrm{P}\left[X(t+1) = \mathbf{y} \middle| X(t) = \mathbf{x}, U(t) = u\right] = \begin{cases} p_u, & \text{if } \mathbf{y} = (x_1+1, \dots, x_{u-1} + 1, 0, x_{u+1}+1, \dots, x_N+1); \\ 1-p_u, & \text{if } \mathbf{y} = \mathbf{x} + \mathbf{1}; \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{1} := (1, \dots, 1)$. The T-horizon optimal cost-to-go from initial state $\mathbf{x}$ is as follows

$$V_T(\mathbf{x}) := \min_\pi \mathrm{E}_\pi \times \left[\exp\left(\theta \sum_{t=0}^{T-1} \sum_{n=1}^{N} (X_n(t) + 1 - \tau_n)^+ \times \mathbb{1}\{X_n(t+1) = 0\} \right) \middle| X(0) = \mathbf{x} \right], \quad (2)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, and $X(T) := \mathbf{0}$ is assumed to introduce the last term in the cost (1). Our aim is to find an optimal policy amongst all history-dependent scheduling policies π, which achieves the optimal cost-to-go $V_T(\mathbf{x})$ for any value of the initial state $\mathbf{x}$.

We also consider the infinite-horizon problem when $T \to \infty$, and denote the infinite-horizon cost by $J(\pi, \mathbf{x})$(shown in equation (13) in the sequel).

We employ the following notation: Vectors are denoted in bold font, e.g., $\boldsymbol{\tau} := (\tau_1, \dots, \tau_N)$ and $\mathbf{x} := (x_1, \dots, x_N)$. Let $a_n \wedge b_n := \min\{a_n, b_n\}$, and $\mathbf{a} \wedge \mathbf{b} := (a_1 \wedge b_1, \dots, a_N \wedge b_N)$.

IV. Reduction to Finite State Problem

The problem in Section III is denoted as *MDP-1*. Although it has an infinite state space for the infinite time horizon problem, we now show that there is an equivalent finite-state problem.

The dynamic programming (DP) recursive relationship for the optimal cost-to-go functions in MDP-1 is:

$$V_T(\mathbf{x}) = \min_n \left\{ p_n \exp\left(\theta \left(x_n + 1 - \tau_n\right)^+\right) V_{T-1}\left(\mathcal{S}_n(\mathbf{x})\right) + (1-p_n) V_{T-1}\left(\mathbf{x} + \mathbf{1}\right) \right\}, \quad (3)$$

where

$$\mathcal{S}_n(\mathbf{x}) := (x_1 + 1, \dots, x_{n-1} + 1, 0, x_{n+1} + 1, \dots, x_N + 1) \quad (4)$$

is the state that succeeds the state $\mathbf{x}$ in the event of a successful transmission for client n. This directly follows from the optimal cost-to-go definition in (2).

Lemma 1: For MDP-1, the following results hold:

1) For all $n \in \{1, \dots, N\}$, and $\forall x_1, \dots, x_N \geq 0$,

$$V_T\left(x_1, \dots, x_n + \tau_n, \dots, x_N\right) = \exp\left(\theta x_n\right) \cdot V_T\left(x_1, \dots, \tau_n, \dots, x_N\right). \quad (5)$$

Further, the optimal controls in the two states, $(x_1, \dots, x_n + \tau_n, \dots, x_N)$ and $(x_1, \dots, \tau_n, \dots, x_N)$, are the same.

2) The optimal cost function starting with any system state $\mathbf{x}$ such that $x_n \leq \tau_n, \forall n$ satisfies:

$$V_T(\mathbf{x}) = \exp\left(\theta \sum_{n=1}^{N} \mathbb{1}\{x_n = \tau_n\}\right) \times \min_u \left\{ p_u V_{T-1}\left(\mathcal{S}_u(\mathbf{x}) \wedge \boldsymbol{\tau}\right) + (1-p_u) V_{T-1}\left((\mathbf{x} + \mathbf{1}) \wedge \boldsymbol{\tau}\right) \right\}, \quad (6)$$

where $\mathcal{S}_u(\mathbf{x})$ is as in (4);

3) $Y(t) := X(t) \wedge \boldsymbol{\tau}$ is a Markov Decision Process, i.e.

$$\mathrm{P}\left[Y(t+1) \middle| Y(t), \dots, Y(0), U(t), \dots, U(0)\right] = \mathrm{P}\left[Y(t+1) \middle| Y(t), U(t)\right]. \quad (7)$$

Proof: The proof is omitted due to space constraints. More details can be found in the online supplementary material. □

Now we construct a new MDP, denoted by *MDP-2*. We will show in Theorem 1 that this MDP-2 is equivalent to the MDP-1 in an appropriate sense. In the new MDP-2, we still use $U(t) \in \{1, \dots, N\}$ to denote the system *control* at time slot t. The system *state* is an N-dimensional vector $Y(t) := (Y_1(t), \dots, Y_N(t))$ with each element $Y_n(t) \in \{0, 1, \dots, \tau_n\}$. The *state space* is $\mathbb{Y} := \prod_{n=1}^{N} \{0, 1, \dots, \tau_n\}$, which is finite even for the infinite time horizon problem. The transition probabilities of the MDP-2 are set as

$$\mathrm{P}\left[Y(t+1) = \mathbf{y} \middle| Y(t) = \mathbf{x}, U(t) = u\right] = \begin{cases} p_u & \text{if } \mathbf{y} = \mathcal{S}_u(\mathbf{x}) \wedge \boldsymbol{\tau}, \\ 1 - p_u & \text{if } \mathbf{y} = (\mathbf{x} + \mathbf{1}) \wedge \boldsymbol{\tau}, \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

where $\mathcal{S}_u(\mathbf{x})$ is as in (4).

This system incurs the following cost in a T-time horizon problem with initial state $\mathbf{x} \in \mathbb{Y}$ when policy π applied:

$$V_T^\pi(\mathbf{x}) = \mathrm{E}_\pi\Big[\exp\Big(\theta \sum_{t=0}^{T-1}\sum_{n=1}^{N} \mathbb{1}\left\{Y_n(t)=\tau_n\right\}\Big)\Big|Y(0)=\mathbf{x}\Big]. \tag{9}$$

The optimal cost-to-go function is

$$\tilde{V}_T(\mathbf{x}) := \min_\pi V_T^\pi(\mathbf{x}), \forall \mathbf{x} \in \mathbb{Y}. \tag{10}$$

Here we differentiate this optimal cost-to-go function from that for the MDP-1 (recalling (2)) by the superscript tilde.

Theorem 1: The MDP-2 is equivalent to MDP-1 in the following senses:

1) The optimal cost-to-go functions of the two MDPs are equal in each time slot t for any initial state $\mathbf{x}$ such that $x_n \le \tau_n, \forall n$, i.e.

$$V_T(\mathbf{x}) = \tilde{V}_T(\mathbf{x}), \forall \mathbf{x} \in \mathbb{Y}; \tag{11}$$

2) Any optimal control for MDP-1 in state $\mathbf{x}$ is also optimal for MDP-2 in state $\mathbf{x} \wedge \boldsymbol{\tau}$, and conversely.

Proof: The DP recursion for the optimal cost in MDP-2 is

$$\tilde{V}_T(\mathbf{x}) = \exp\Big(\theta\sum_{n=1}^{N}\mathbb{1}\{x_n=\tau_n\}\Big) \times \min_u\Big\{\sum_{\mathbf{y}} P_u(\mathbf{x},\mathbf{y})\tilde{V}_{T-1}(\mathbf{y})\Big\}, \tag{12}$$

where $P_u(\mathbf{x},\mathbf{y}) := \mathrm{P}\big[Y(t+1)=\mathbf{y}\big|Y(t)=\mathbf{x}, U(t)=u\big]$, which can be obtained from (8). By comparing the r.h.s. of (12) with the r.h.s. of (6), we note that the optimal cost-to-go function $\tilde{V}_T(\cdot)$ for MDP-2 and the optimal cost-to-go function $V_T(\cdot)$ (restricted to the state space $\mathbb{Y}$) for MDP-1 evolve in exactly the same way. Further, the initial costs satisfy $V_0(\mathbf{x}) = \tilde{V}_0(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathbb{Y}$. Thus, statement 1) holds.

In order to prove statement 2), note that since we have already shown that MDP-2 and MDP-1 have identical recursive relationships on state space $\mathbb{Y}$, it directly follows that the two systems have identical optimal controls at any state $\mathbf{x} \in \mathbb{Y}$. Further recalling the first statement of Lemma 1, statement 2) directly follows. □

One can interpret the control in MDP-2 as the client to transmit in slot t, and the evolution of the MDP-2 as

$$Y_n(t+1) = \begin{cases} 0 & \text{if a packet delivered for client } n \text{ in slot } t, \\ (Y_n(t)+1)\wedge\tau_n & \text{otherwise.} \end{cases}$$

It can be observed that, under the same scheduling policy, the system state of the MDP-2 and the accompanying process of the MDP-1, defined in (7), evolve in the same way statistically. (This is also the reason why the symbol $Y(t)$ is used to denote both these processes.) Therefore, we use the symbol $U(t)$ for the controls of both MDP-1 and MDP-2.

V. The Risk-Sensitive Approach

By Theorem 1, we can focus exclusively on MDP-2 which involves only a finite number $\prod_{n=1}^{N}(\tau_n+1)$ of states. We begin with some notation. The (risk-sensitive infinite horizon) *average cost* under policy π starting at state $\mathbf{x}$ is defined as

$$J(\pi,\mathbf{x}) := \limsup_{T\to\infty}\frac{1}{\theta}\cdot\frac{1}{T}\ln V_T^\pi(\mathbf{x}), \quad \forall \mathbf{x}\in\mathbb{Y}, \tag{13}$$

where $V_T^\pi(\mathbf{x})$ is as in (9). A stationary policy can be described by a mapping f from state space $\mathbb{Y}$ to control set $\{1,\ldots,N\}$, i.e., the control $U(t) = f(Y(t))$.

We now define a special class of stationary policies, called *Non-Exclusionary (NE)* policies, which have the following property: Under an NE policy, for any n, client n is not selected to transmit when the system state is $(\tau_1,\ldots,\tau_{n-1},0,\tau_{n+1},\ldots,\tau_N)$. We can show that, any non-NE stationary policy is either out-performed by some NE policy, or associated with a cost that is trivial to obtain (see online supplementary material). Consequently, we focus exclusively on NE policies.

Denote by $\mathbf{P}^f$ the *transition probability matrix* of a stationary policy f, i.e.

$$(\mathbf{P}^f)_{\mathbf{x},\mathbf{y}} := \mathrm{P}\left[Y(t+1)=\mathbf{y}\middle|Y(t)=\mathbf{x},U(t)=f(\mathbf{x})\right].$$

Further, denote by $\mathbf{L}^f$ the *disutility matrix* of f, i.e.,

$$(\mathbf{L}^f)_{\mathbf{x},\mathbf{y}} := \exp\Big(\theta\sum_{n=1}^{N}\mathbb{1}\{x_n=\tau_n\}\Big)\cdot(\mathbf{P}^f)_{\mathbf{x},\mathbf{y}}, \quad \forall \mathbf{x},\mathbf{y}\in\mathbb{Y}.$$

Also, denote by $\rho(\mathbf{L}^f)$ the spectral radius of this matrix.

The standard notations of transient/non-transient states and communicating classes (of states) of Markov Chain are also used here. In a communicating class, any two states are accessible from each other. Further, any two different communicating classes are disjoint. The reader may refer to [28] for details.

Lemma 2: The following results hold for any NE policy f:

1) There is one and only one non-transient communicating class, which contains the state $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_N)$.
2) If a state $\mathbf{y}$ is transient, then $(\mathbf{P}^f)_{\mathbf{y},\mathbf{y}} = 0$.
3) If there is only one communicating class, then, the average cost under policy f is,

$$J(f,\mathbf{y}) = \frac{1}{\theta}\ln\rho(\mathbf{L}^f), \quad \forall \mathbf{y}. \tag{14}$$

Proof: To show statement 1), it is sufficient to show that the state $\boldsymbol{\tau}$ is accessible from any state $\mathbf{x}\in\mathbb{Y}$. This is because, since $p_n < 1$ for each client n, by (8), when $Y(t)=\mathbf{x}$, no matter which value of control $U(t)$ is chosen, there is a positive probability that $Y(t+1) = (\mathbf{x}+1)\wedge\boldsymbol{\tau}$. We continually apply this argument for $\tau_{\max} := \max_{n=1}^{N}\tau_n$ time slots. Then, it can be shown that for any $\mathbf{x}\in\mathbb{Y}$, when $Y(t)=\mathbf{x}$, there is a positive probability that $Y(t+\tau_{\max}) = \boldsymbol{\tau}$, no matter which policy is applied. Thus, statement 1) follows.

For statement 2), by analyzing (8), it can be shown that if $(\mathbf{P}^f)_{\mathbf{y},\mathbf{y}} \neq 0$ holds, then one of the following holds:

(i) $\mathbf{y} = (\tau_1,\ldots,\tau_N)$;
(ii) $\mathbf{y}$ is such that $\exists n, y_n = 0, y_l = \tau_l, \forall l \neq n$, and $f(\mathbf{y}) = n$.

However, in case (i), $\mathbf{y}$ is non-transient by statement 1); in case (ii), client n is selected to transmit when the system state is $(\tau_1, \ldots, \tau_{n-1}, 0, \tau_{n+1}, \ldots, \tau_N)$, which violates the fact that policy f is an NE policy. Thus, statement 2) holds.

For statement 3), the proof is omitted due to space constraints; more details can be found in the online supplementary material. □

In the following, it is assumed that for any NE policy, only one communicating class exists. This is not restrictive for the following reasons: First, by the first statement in Lemma 2, there is only one non-transient communicating class. Also, the existence of a singleton transient communicating class has already been ruled out by the second statement of Lemma 2. Second, we can restrict the state space to the non-transient states, which form a single communicating class by the first statement in Lemma 2. Denote $p_{\max} := \max_{n=1}^{N} p_n$ and $\tau_{\max} := \max_{n=1}^{N} \tau_n$. Further, let $K := \lceil \tau_{\max}(1 - p_{\max})^{-\tau_{\max}} \rceil$.

Theorem 2: Let

$$\theta_{\text{th}} := \frac{\ln(K+1) - \ln(K)}{2N\,(K+1)}.$$

For the infinite-horizon MDP-2 with average cost criterion,

1) A stationary optimal policy exists when $\theta < \theta_{\text{th}}$;
2) This stationary optimal policy can be computed in a finite number of steps.

Proof: Let $T_{\text{Db}}(\mathbf{x})$ be the first passage time from state $\mathbf{x}$ to $(\tau_1, \ldots, \tau_N)$, i.e.,

$$T_{\text{Db}}(\mathbf{x}) := \min\left\{t > 0 \middle| Y(t) = (\tau_1, \ldots, \tau_N)\,, Y(0) = \mathbf{x}\right\}.$$

We first prove that for any stationary policy f,

$$\mathrm{E}_f[T_{\text{Db}}(\mathbf{x})] \leq K, \quad \forall \mathbf{x}, \tag{15}$$

which is the so-called simultaneous Doeblin condition [14].

In the proof of the first statement of Lemma 2, we have shown that for the system (8) evolving under the application of an arbitrary scheduling policy, and starting with an initial state $\mathbf{x} \in \mathbb{Y}$, if the system witnesses $\tau_{\max}$ consecutive transmission failures, it hits the state $\boldsymbol{\tau}$. That is, from any initial state $\mathbf{x}$, the system hits the state $\boldsymbol{\tau}$ within $\tau_{\max}$ time slots with a probability no less than $(1 - p_{\max})^{\tau_{\max}}$. Thus

$$\begin{aligned}\mathrm{E}_f\left[T_{\text{Db}}(\mathbf{x})\right] &\leq \sum_{j=1}^{+\infty} j\tau_{\max} \frac{(1-p_{\max})^{\tau_{\max}}}{\left[1-(1-p_{\max})^{\tau_{\max}}\right]^{-(j-1)}} \\ &= \frac{\tau_{\max}}{(1-p_{\max})^{\tau_{\max}}}. \end{aligned} \tag{16}$$

To prove the first inequality in (16), let us consider the event that the state $\boldsymbol{\tau}$ is not hit within $(j-1)\tau_{\max}$ slots. The probability of this event is at most $\left[1-(1-p_{\max})^{\tau_{\max}}\right]^{(j-1)}$. Further, if this event happens, the probability that the state $\boldsymbol{\tau}$ is hit within the next τ_{max} slots is at least $(1-p_{\max})^{\tau_{\max}}$, leading to (16). As a result, recalling the definition of K, we have (15). By combining (15) with the [14, Th. 3.1], statement 1) follows.

For statement 2), the average cost of any NE policy can be obtained by Lemma 2. Also, the average cost of any non-NE policy can be trivially obtained. Thus, the cost associated with any stationary policy can be computed. Since there are only a finite number of stationary policies, and because a stationary optimal policy exists by statement 1), statement 2) follows. □

VI. The High Reliability Asymptotic Approach

While the stationary optimal policy can be computed in a finite number of steps, as shown in Theorem 2, this procedure suffers from a significant computational complexity issue: If we denote by $|\mathbb{Y}|$ the cardinality of the state space in MDP-2, then the total number of possible stationary policies is $N^{|\mathbb{Y}|}$. Note that $|\mathbb{Y}| = \prod_{n=1}^{N}(\tau_n + 1)$, and so the cardinality of the state space increases exponentially in the number of clients N. To calculate the cost associated with each NE policy, the spectral radius of a $[0,1]^{|\mathbb{Y}|\times|\mathbb{Y}|}$ matrix needs to be calculated by Lemma 2. Policy iteration [27] cannot be applied to this non-irreducible risk-sensitive problem whose communicating class varies for different policies. Thus, to find the optimal policy, one needs to compare all the stationary policies with respect to their associated average costs. All these factors lead to great computational complexity of the approach.

We therefore propose a high-channel-reliability asymptotic approach to obtain near optimal policies, and show that this leads to a huge simplification with regards to computational complexity. In essence, we are interested in the scenarios when the channel reliability is relatively high, and we take advantage of this asymptotic approach to design policies of appealing structure and good performance. This is achieved by analyzing the cases when the channel reliabilities asymptotically approach 1, i.e. the high-reliability asymptotic regime.

In this section, we obtain some important preliminary results by applying the high-reliability asymptotic approach, which facilitates the design of asymptotically optimal policies in the next section. We begin by focusing on the two-client scenario for ease of exposition. The results are generalized to the multi-client scenario in Section VII.

A. Two-Client Scenario and the Modified-Least-Time-to-Go (MLG) Policy

We begin with some notations. Consider a two-client scenario with channel reliabilities $p_1 = 1 - b_1\epsilon$, $p_2 = 1 - b_2\epsilon$, where $\epsilon > 0$ is a small quantity and $b_1, b_2 > 0$. Without loss of generality, we suppose that $\tau_1 = \tau$ and $\tau_2 = \tau + \Delta$, where $\Delta \geq 0$.

The *modified-least-time-to-go (MLG)* policy is defined as

$$f^{\text{MLG}}(\mathbf{x}) = \begin{cases} 2 & \text{if } \mathbf{x} = (0, \Delta-1), \\ \max\left\{\arg\min_{n=1}^{N}(\tau_n - x_n)\right\} & \text{otherwise.} \end{cases}$$

It can be understood as follows: usually, the AP selects the client with the least gap between its corresponding threshold and its time elapsed since last delivery, i.e., the client with the least value of $\tau_n - x_n$, to transmit. That is the "least time to go". However, there is one exception: when the system state is $\mathbf{x} = (0, \Delta - 1)$, client 2 is selected to transmit instead of client 1 which has a smaller "gap" - $\tau_1 - 0 = \tau < \tau_2 - (\Delta - 1) = \tau + 1$.

It will be shown in Section VII that when $\epsilon \searrow 0$, the MLG policy is asymptotically optimal. We begin with some analyses.

B. Regeneration Cycle

We define the *cost incurred in time slots* $t_1, t_1+1, \ldots, t_2$ as:

$$\prod_{t=t_1}^{t_2} \exp\left(\theta \sum_{n=1}^{N} \mathbb{1}\left\{Y_n(t)=\tau_n\right\}\right). \tag{17}$$

In this system, we let the *regeneration point* of interest be the time slot when the state $(1,0)$ is hit, i.e., when $Y(t)=(1,0)$; and let the *regeneration cycle* be the time interval between two such successive regeneration points. Thus, under the application of a stationary policy f, MDP-2 evolves statistically the same during different regeneration cycles. Further, we denote by v_{cycle} the *cost incurred in a regeneration cycle* (17), and denote by l_{cycle} the *length of a regeneration cycle*. It directly follows that the random variables v_{cycle} and l_{cycle} are independent and identically distributed (i.i.d.) in different regeneration cycles. Consequently, the following holds almost surely

$$\begin{aligned} J(f,\mathbf{x}) &= \lim_{T\to\infty} \frac{1}{\theta}\cdot\frac{1}{T}\ln V_T^f(1,0) \\ &= \lim_{T\to\infty} \frac{1}{\theta}\frac{1}{T}\ln \mathrm{E}\left[\prod_{j=1}^{M_T^{(\text{cycle})}} v_{\text{cycle}}^{(j)}\right] \\ &= \lim_{T\to\infty} \frac{1}{\theta}\frac{M_T^{(\text{cycle})}}{T}\ln \mathrm{E}\left[v_{\text{cycle}}\right] \\ &= \frac{1}{\theta}\frac{\ln \mathrm{E}\left[v_{\text{cycle}}\right]}{\mathrm{E}\left[l_{\text{cycle}}\right]}, \quad \forall \mathbf{x}\in\mathbb{Y}. \end{aligned} \tag{18}$$

where, the first equality holds by (13) and (14); $v_{\text{cycle}}^{(j)}$ is the cost incurred during the j-th regeneration cycle, and $M_T^{(\text{cycle})}$ is the total number of regeneration cycles during T slots, so the second equality in (18) follows from the definition of $V_T^{\pi}(\mathbf{x})$ in (9); the third equality holds because $v_{\text{cycle}}^{(j)}, \forall j$ are i.i.d..

By (18), the analysis of the long-term average cost J is reduced to the analysis of the expected cost and expected length of a regeneration cycle, which facilitates the following discussion.

C. SS-Point and SS-Period

A time slot is called an *SS-point* if in the two successive time slots preceding it, the packet transmissions are both successful. We illustrate the SS-points in Fig. 2. More formally, we define the j-th SS-point as

$$\tau_j^{ss} := \begin{cases} \min\{t: t>0 \text{ and slots } t-1, t-2 \text{ have} \\ \quad \text{successful transmissions}\} \quad \text{for } j=1 \\ \min\{t: t>\tau_{j-1}^{ss} \text{ and slots } t-1, t-2 \text{ have} \\ \quad \text{successful transmissions}\} \quad \text{for } j=2,3,\ldots \end{cases} \tag{19}$$

The time interval between two successive SS-points is called an *SS-period*. The *cost incurred during an SS-period* is denoted by $v_{\text{ss}}(\mathbf{x})$ (recalling (17)), with $\mathbf{x}$ being the system state at the first time slot of the SS-period. With a stationary policy applied, for SS-periods with the identical starting state $\mathbf{x}$, the random variables $v_{\text{ss}}(\mathbf{x})$ are i.i.d.

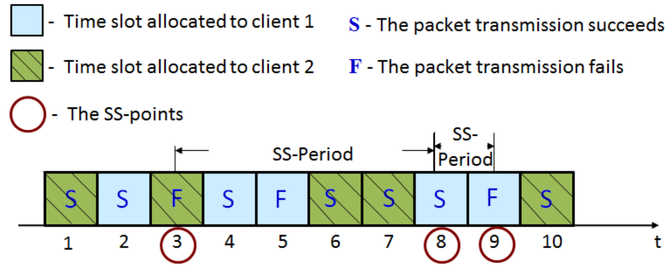


Fig. 2. SS-points and SS-periods illustrated in a two-client scenario. (We arbitrarily allocate the time slots here for illustrative purposes.)

Now we begin with some preliminary results for NE policies (recall NE policies in Section V).

First, with an arbitrary NE policy being applied, the following result holds for any value of the starting state $\mathbf{x}$ of an SS-period,

$$\mathrm{P}\left\{v_{\text{ss}}(\mathbf{x})>1\right\} = O(\epsilon). \tag{20}$$

This follows from the following facts:

i) $v_{\text{ss}}(\mathbf{x})>1$ only if $Y_n(t)=\tau_n$ holds for some client n and some time-slot t in this SS-period.
ii) For an SS-period, if there is no transmission failure, its length is exactly one time slot.
iii) With an NE policy being applied, the system state $\mathbf{x}$ at an SS-point satisfies $x_n \neq \tau_n, \forall n$.

The statement i) holds by the definition of $v_{\text{ss}}(\mathbf{x})$. The statement ii) holds by the definition of the SS-point and the SS-period. For statement iii), after successful delivery of a packet, at least one element of the system state is zero. Further, with an NE policy applied, if the system state is $(\tau_1, 0)$ then client 1 is selected to transmit; while if the state is $(0,\tau_2)$ then client 2 is selected to transmit. Thus, after another successful delivery of a packet, the resulting system state $\mathbf{x}$ must satisfy $x_n \neq \tau_n, \forall n$. By these facts, and by combining i) and iii), we conclude that if an SS-period incurs a cost greater than 1, it consists of at least two time slots. Then, it follows from ii) that if $v_{\text{ss}}(\mathbf{x})>1$, there is at least one transmission failure in this SS-period, leading to equation (20).

Here and in the sequel, we use the asymptotic big O and little o notations [29]: $z(\epsilon)=O(\epsilon^k)$ means that there exists a positive δ and a positive L such that $|z(\epsilon)| \leq L\epsilon^k$ when $0<\epsilon<\delta$, i.e., $\limsup_{\epsilon\to 0+} \frac{|z(\epsilon)|}{\epsilon^k} < \infty$, while $z(\epsilon)=o(\epsilon^k)$ means that for any positive ζ, there exists a positive δ such that $|z(\epsilon)| \leq \zeta\epsilon^k$ when $0<\epsilon<\delta$, i.e., $\lim_{\epsilon\to 0+}\frac{z(\epsilon)}{\epsilon^k}=0$.

Second, let $P_{\mathbf{x}}^{(\text{ss})}$ denote the probability that in a regeneration cycle, there is at least one SS-period whose starting state is $\mathbf{x}$. Then we have

$$\mathrm{E}\left[v_{\text{cycle}}\right] = 1 + \textstyle\sum_{\mathbf{x}} P_{\mathbf{x}}^{(\text{ss})}\left(\mathrm{E}\left[v_{\text{ss}}(\mathbf{x})\right]-1\right) + o(\epsilon^k), \tag{21}$$

where $k \geq 1$ is an integer such that there exists $d_1, d_2 > 0$ satisfying $d_1\epsilon^k \leq \mathrm{E}\left[v_{\text{cycle}}\right]-1 \leq d_2\epsilon^k$ when ϵ is sufficiently small. This result holds by noting the following facts:

1) Any regeneration cycle consists only of SS-periods.
2) If a regeneration cycle is associated with a cost $v_{\text{cycle}}>1$, then at least one of the SS-periods included in it has a cost $v_{\text{ss}}(\cdot)>1$.

3) Whenever ϵ is small enough, the probability that two or more SS-periods each incurs a cost $v_{\text{ss}}(\cdot) > 1$ is much less than the probability that only one of these SS-periods incurs a cost $v_{\text{ss}}(\cdot) > 1$.

For statements 1) and 2), the proof is direct and omitted. The statement 3) follows from (20). As a consequence, we have

$$\begin{aligned}\mathrm{E}[v_{\text{cycle}}] &= \mathrm{P}\{v_{\text{cycle}} = 1\}\cdot 1 + \mathrm{P}\{v_{\text{cycle}} > 1\}\mathrm{E}[v_{\text{cycle}}|v_{\text{cycle}} > 1] \\ &= 1 + \mathrm{P}\{v_{\text{cycle}} > 1\}\left(\mathrm{E}[v_{\text{cycle}}|v_{\text{cycle}} > 1] - 1\right) \\ &= 1 + \mathrm{P}\{\text{in a cycle}, \exists\text{SS-period such that } v_{\text{ss}}(\cdot) > 1\} \\ &\quad \cdot(\mathrm{E}[v_{\text{cycle}}|\ \exists\text{SS-period such that } v_{\text{ss}}(\cdot) > 1] - 1) \\ &= 1 + \mathrm{P}\{\text{a cycle has one SS-period with } v_{\text{ss}}(\cdot) > 1\} \\ &\quad \cdot(\mathrm{E}[v_{\text{cycle}}|\text{one SS-period with } v_{\text{ss}}(\cdot) > 1] - 1) \\ &\quad + o(\epsilon^k) = 1 + o(\epsilon^k) + \sum_{\mathbf{x}} P_{\mathbf{x}}^{(\text{ss})} \\ &\quad \times \mathrm{P}\{v_{\text{ss}}(\mathbf{x}) > 1\}(\mathrm{E}[v_{\text{ss}}(\mathbf{x})|v_{\text{ss}}(\mathbf{x}) > 1] - 1) \\ &= 1 + \sum_{\mathbf{x}} P_{\mathbf{x}}^{(\text{ss})}(\mathrm{E}[v_{\text{ss}}(\mathbf{x})] - 1) + o(\epsilon^k). \end{aligned} \tag{22}$$

In (22), the first equality follows from the law of total expectation. The second equality is obvious. The third equality follows from statements 1) and 2). The fourth equality follows from statement 3), where $k \geq 1$ is an integer such that there exists $d_1, d_2 > 0$ satisfying $d_1\epsilon^k \leq \mathrm{E}[v_{\text{cycle}}] - 1 \leq d_2\epsilon^k$ whenever ϵ is sufficiently small. The fifth equality holds by considering which SS-period in the given cycle causes a cost $v_{\text{ss}}(\cdot) > 1$, and recalling that $P_{\mathbf{x}}^{(\text{ss})}$ denotes the probability that in a regeneration cycle, there is at least one SS-period whose starting state is $\mathbf{x}$. The sixth equality holds by noting that

$$\begin{aligned}\mathrm{E}[v_{\text{ss}}(\mathbf{x})] &= P\{v_{\text{ss}}(\mathbf{x}) = 1\}\cdot 1 \\ &\quad + P\{v_{\text{ss}}(\mathbf{x}) > 1\}\mathrm{E}[v_{\text{ss}}(\mathbf{x})|v_{\text{ss}}(\mathbf{x}) > 1] \\ &= 1 + P\{v_{\text{ss}}(\mathbf{x}) > 1\}(\mathrm{E}[v_{\text{ss}}(\mathbf{x})|v_{\text{ss}}(\mathbf{x}) > 1] - 1).\end{aligned}$$

Thus, (21) directly follows from (22). Here, note that, by statement 3) above, we can repeatedly ignore events with relatively small probabilities for small enough ϵ. This "omission principle" technique captures the key aspect of the high-reliability asymptotic approach, and will be frequently applied in the following.

Now we consider a special *regeneration cycle which consists of only successful transmissions*. Let $\mathbb{X}_{\text{ss}}$ be the set of the system states hit in such a regeneration cycle.

Lemma 3: The following results hold:

1) Under the application of any NE policy,
$$\mathrm{E}\left[v_{\text{cycle}}\right] \geq 1 + \sum_{\mathbf{x}\in\mathbb{X}_{ss}} \left(\mathrm{E}\left[v_{\text{ss}}\left(\mathbf{x}\right)\right] - 1\right) + o(\epsilon^k), \tag{23}$$
where k is an integer such that there exists $d_1, d_2 > 0$ satisfying $d_1\epsilon^k \leq \mathrm{E}\left[v_{\text{cycle}}\right] - 1 \leq d_2\epsilon^k$ whenever ϵ is sufficiently small.
2) With the MLG policy applied
$$\mathrm{E}\left[v_{\text{cycle}}\right] = 1 + \sum_{\mathbf{x}\in\mathbb{X}_{ss}} \left(\mathrm{E}\left[v_{\text{ss}}\left(\mathbf{x}\right)\right] - 1\right) + o(\epsilon^k) \tag{24}$$
where k is similarly defined as in (23).
3) Further, with the MLG policy applied, if $\Delta \geq 2$
$$\mathbb{X}_{\text{ss}} = \{(1,0), (0,x_2), \quad \forall x_2 = 0, \ldots, \Delta - 1\} \tag{25}$$
$$\mathrm{E}[l_{\text{cycle}}] = \Delta + O(\epsilon); \tag{26}$$
$$\mathrm{E}[v_{\text{ss}}(1,0)] = 1 + b_1^{\tau-1}\epsilon^{\tau-1}\left(\exp(\theta)-1\right) + O(\epsilon^\tau), \tag{27}$$
$$\mathrm{E}[v_{\text{ss}}(0,x_2)] = 1 + O(\epsilon^\tau), \quad \forall x_2 = 0, \ldots, \Delta - 1. \tag{28}$$
Similar results can be obtained for $\Delta = 0, 1$.

Proof: The proof is omitted with more details in the online supplementary material. □

VII. Asymptotically Optimal Policies

We now determine asymptotically optimal policies and treat the general multi-client case in this section.

A. The MLG Policy in Two-Client Scenario

Consider a two-client scenario with channel reliabilities $p_1 = 1 - b_1\epsilon$, $p_2 = 1 - b_2\epsilon$ and inter-delivery thresholds $\tau_1 = \tau$, $\tau_2 = \tau + \Delta$, where $\Delta \geq 0$ is assumed without loss of generality.

Theorem 3: The following results hold:

1) With the MLG policy applied, for any initial state $\mathbf{x} \in \mathbb{Y}$, the risk-sensitive average cost is,
$$J(f^{\text{MLG}}, \mathbf{x}) = \begin{cases} A_0\epsilon^{\tau-1} + O(\epsilon^\tau) & \text{if } \Delta = 0, \\ \dfrac{\mathrm{e}^\theta - 1}{2\theta}\epsilon^{\tau-1}\sum_{j=0}^{\tau-1} b_1^j b_2^{\tau-1-j} + O(\epsilon^\tau) & \text{if } \Delta = 1, \\ \dfrac{\mathrm{e}^\theta - 1}{\theta\Delta} b_1^{\tau-1}\epsilon^{\tau-1} + O(\epsilon^\tau) & \text{if } \Delta \geq 2, \end{cases}$$
where
$$A_0 = \frac{\mathrm{e}^\theta - 1}{\theta}\sum_{j=1}^{\tau-2} b_1^j b_2^{\tau-1-j} + \frac{\mathrm{e}^{2\theta} - 1}{2\theta}\left(b_1^{\tau-1} + b_2^{\tau-1}\right).$$
2) The optimal cost over all the stationary policies, denoted $J^\star(\mathbf{x}) := \min_f J(f, \mathbf{x})$, satisfies the lower bound:
$$J^\star(\mathbf{x}) \geq \begin{cases} A_0\epsilon^{\tau-1} + o(\epsilon^{\tau-1}) & \text{if } \Delta = 0, \\ \dfrac{\mathrm{e}^\theta - 1}{2\theta}A_1\epsilon^{\tau-1} + o(\epsilon^{\tau-1}) & \text{if } \Delta = 1, \\ \dfrac{\mathrm{e}^\theta - 1}{\theta}A_2\epsilon^{\tau-1} + o(\epsilon^{\tau-1}) & \text{if } \Delta \geq 2, \end{cases}$$
where
A_0 is as in the statement above,
$$b_{\min} := \min\{b_1, b_2\},$$
$$A_1 := b_1^{\tau-1} + (\tau - 1)\, b_{\min}^{\tau-1},$$
$$A_2 := \min\left\{\frac{b_1^{\tau-1}}{\Delta}, \frac{b_1^{\tau-1} + (\tau-1)b_{\min}^{\tau-1}}{\Delta + 1}, \text{ and } \frac{b_1^{\tau-1} + (\tau-1)b_{\min}^{\tau-1} + \sum_{j=1}^{\tau-1} b_2^j b_1^{\tau-1-j}}{\Delta + 2}\right\}.$$
3) The MLG policy is asymptotically optimal in the high reliability asymptotic regime (i.e., when $\epsilon \searrow 0$) if any of the following conditions holds:
(i) $\Delta = 0$; (ii) $\Delta = 1$, and $b_1 \leq b_2$;
(iii) $\Delta \geq 2$, and $b_1^{\tau-1} \leq \Delta(\tau-1)b_2^{\tau-1}$.

Proof: In the following, we focus on the case when $\Delta \geq 2$. The proof for the cases when $\Delta = 1$ or 0 is similar and is therefore omitted.

For statement 1), with the MLG policy applied, the following result follows from the second and the third statements of Lemma 3,

$$\mathrm{E}[v_{\text{cycle}}] = 1 + b_1^{\tau-1}\epsilon^{\tau-1}\big(\exp(\theta) - 1\big) + O(\epsilon^{\tau}). \quad (29)$$

As a result, by combining (18), (26) and (29), statement 1) follows. Here, we note that, for any positive $d > 0$, $\ln\big(1 + d\epsilon^{\tau-1}\big) = d\epsilon^{\tau-1} + O(\epsilon^{2(\tau-1)})$ holds. This follows from the Taylor expansion of $\ln\big(1 + d\epsilon^{\tau-1}\big)$ and the definition of the asymptotic notation big $O(\cdot)$.

To prove statement 2), in the following, we first derive a lower bound on $\mathrm{E}[v_{\text{ss}}(\mathbf{x})]$ for all possible starting states $\mathbf{x}$ of an SS-period, and then analyze a lower bound for the average cost J by further combining (18) and (23).

For the lower bound on $\mathrm{E}[v_{\text{ss}}(\mathbf{x})]$, we begin with the case when the starting state $\mathbf{x} = (1, 0)$. That is, we focus on an SS-period starting with state $(1, 0)$ with an arbitrary stationary policy applied, and analyze how the system evolves during such a period. Depending on the policy applied, there are two possibilities,

(a) The policy serves client 2 before the earlier of these two events: $i)$ a successful packet delivery for client 1, $ii)$ the system hits the value $(\tau, \tau-1)$. Under such a policy, it can be shown that a cost $v_{\text{ss}}(1, 0) > 1$ is incurred with a probability strictly larger than $d\epsilon^{\tau-2}$ for some $d > 0$.

(b) The policy does not serve client 2 before the earlier of the following two events: $i)$ a successful packet delivery for client 1, $ii)$ the system hits the value $(\tau, \tau-1)$. Then, if failures occur in all of the first $\tau - 1$ time slots for the SS-period, the state $(\tau, \tau - 1)$ will be hit. Thus, a cost $v_{\text{ss}}(1, 0) \geq \exp(\theta)$ is incurred with a probability greater than or equal to $b_1^{\tau-1}\epsilon^{\tau-1}$.

The proof of (a) follows a similar argument as in the proof of Lemma 3, and is thus omitted. Consequently, combining (a) and (b) above, we have

$$\mathrm{E}[v_{\text{ss}}(1, 0)] \geq 1 + b_1^{\tau-1}\epsilon^{\tau-1}\left(\exp(\theta) - 1\right) + o(\epsilon^{\tau-1}).$$

A lower bound on $\mathrm{E}[v_{\text{ss}}(\mathbf{x})]$ for other possible starting states of an SS-period can be obtained in a similar way. Here, we summarize the results as follows,

i. $\forall\mathbf{x} \in \{(0, x_2)|x_2 \leq \Delta - 1\}$, $\mathrm{E}[v_{\text{ss}}(\mathbf{x})] \geq 1$.
ii. $\forall\mathbf{x} \in \{(0, x_2)|x_2 \geq \Delta + 2\} \bigcup \{(x_1, 0)|x_1 \geq 2\}$, $\mathrm{E}[v_{\text{ss}}(\mathbf{x})] \geq 1 + d\epsilon^{\tau-2} + o(\epsilon^{\tau-2})$, with some $d > 0$.
iii. $\mathrm{E}[v_{\text{ss}}(1, 0)] \geq 1 + b_1^{\tau-1}\epsilon^{\tau-1}\left(\exp(\theta) - 1\right) + o(\epsilon^{\tau-1})$.
iv. $\mathrm{E}[v_{\text{ss}}(0, \Delta{+}1)] \geq 1 + \sum_{j=1}^{\tau-1} b_2^{j} b_1^{\tau-1-j}\epsilon^{\tau-1}(\exp(\theta) - 1) + o(\epsilon^{\tau-1})$.
v. $\mathrm{E}[v_{\text{ss}}(0, \Delta)] \geq 1 + (\tau - 1)b_{\min}^{\tau-1}\epsilon^{\tau-1}(\exp(\theta) - 1) + o(\epsilon^{\tau-1})$.

Note that the possible starting states of an SS-period are of the form $(\cdot, 0)$ or $(0, \cdot)$. Thus, results i-v provide the lower bounds on $\mathrm{E}[v_{\text{ss}}(\mathbf{x})]$ for all possible starting states $\mathbf{x}$.

Now, note that by equality (18), the analysis of the average cost is reduced to the analysis of $\mathrm{E}[v_{\text{cycle}}]$ and $\mathrm{E}[l_{\text{cycle}}]$, and by the inequality (23), the lower bound on $\mathrm{E}[v_{\text{cycle}}]$ can be obtained by analyzing the SS-periods. Further, we note that $\mathrm{E}[l_{\text{cycle}}]$ is dominated by the length of a regeneration cycle consisting only of successful transmissions, as illustrated in (25) and (26). Combining these results, we obtain the statement 2).

Statement 3) directly follows from statements 1) and 2). □

By Theorem 3, the optimality of the MLG policy for the two-client scenario in the high-reliability asymptotic regime has been established. Note that the conditions in the third statement of Theorem 3 concern the difference of the inter-delivery thresholds, and the ratio of the relative failure probabilities.

B. Asymptotically Optimal Policy in the General Case

Now, we generalize the results obtained in Section VI and Section VII-A to the multi-client scenario. Consider a system with a single AP and N clients connected to it through wireless channels. The channel reliability of client n is $p_n = 1 - b_n\epsilon$, where $\epsilon > 0$ is a small quantity and parameter $b_n > 0, \forall n$. Without loss of generality, we assume that the inter-delivery thresholds are such that $N \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_N$. In the following, we focus exclusively on stationary policies, since we have already shown the existence of a stationary optimal policy in Theorem 2.

We begin with some notation. A *regeneration point* is defined as the time epoch when the system hits the state $(0, 1, \ldots, N{-}1)$, i.e., time t is a regeneration point if and only if $Y(t) = (0, 1, \ldots, N{-}1)$. (This is similar to the regeneration point in the 2-client scenario, which is the time epoch when the state $(1, 0)$ is hit.) Also, a *regeneration cycle* is defined as the time interval between two successive regeneration points. We further consider a special *regeneration cycle which consists only of successful transmissions*. We note that, with a given stationary policy, the system states which are hit during such a regeneration cycle turn out to be a deterministic sequence, which we denote by $\mathbb{X}_{\text{sN}}$. Denote by $|\mathbb{X}_{\text{sN}}|$ the length of this sequence, and by $\mathbb{X}_{\text{sN}}(j)$ the j-th state in this sequence.

In addition, similar to the definition of SS-point in (19), a time slot is called an *SN-point* if, in the N successive time slots preceding it, all packet transmissions are successful. More formally, the j-th SN-point is defined as

$$\tau_j^{sN} := \begin{cases} \min\{t : t > 0 \text{ and slots } t-1, \ldots, t-N \text{ are} \\ \qquad \text{successful transmissions}\} \text{ for } j = 1 \\ \min\{t : t > \tau_{j-1}^{sN} \text{ and slots } t-1, \ldots, t-N \text{ are} \\ \qquad \text{successful transmissions}\} \text{ for } j = 2, 3, \ldots \end{cases} \quad (30)$$

Then, the *SN-period* is defined as the time interval between two successive SN-points. Also, similar to the two-client case, we denote by $v_{\text{sN}}(\mathbf{x})$ the cost (17) incurred during such an SN-period when the starting state of the SN-period is $\mathbf{x}$.

Now we consider a time interval beginning with an arbitrary starting state $\mathbf{x} \in \mathbb{Y}$, and ending when the system hits the nearest SN-point that makes the length of this time interval no less than N. Let the cost (17) incurred during such a time interval be $\tilde{v}_{\text{sN}}(\mathbf{x})$. (The difference between $\tilde{v}_{\text{sN}}(\mathbf{x})$ and $v_{\text{sN}}(\mathbf{x})$ is that while the former is the cost incurred during a time interval of length greater than or equal to N, the latter denotes the cost incurred during an SN-period which may have a shorter length. We recall the example in Fig. 2 where an SN-period of length 1 is illustrated for a two-client scenario.)

Lemma 4: If a policy f is optimal, then it must satisfy,

$$J(f, \mathbf{x}) = O(\epsilon), \quad \forall\mathbf{x} \in \mathbb{Y}. \quad (31)$$

Moreover, for any policy f satisfying (31), the following results hold:

1) For any initial state $\mathbf{x} \in \mathbb{Y}$, the average cost satisfies,

$$J(f,\mathbf{x}) = \frac{1}{|\mathbb{X}_{\text{sN}}|}\sum_{j=1}^{|\mathbb{X}_{\text{sN}}|}\left(\mathrm{E}\left[v_{\text{sN}}\big(\mathbb{X}_{\text{sN}}(j)\big)\right]-1\right)+o(\epsilon^k) \quad (32)$$

where $k \geq 1$ is an integer such that there exist $d_1, d_2 > 0$ satisfying $d_1\epsilon^k \leq J(f,\mathbf{x}) \leq d_2\epsilon^k$ whenever ϵ is sufficiently small.

2) For any possible state $\mathbf{x}$ at an SN-point, we have,

$$\mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})] = 1 + \sum_{j=0}^{N-1}\left(\mathrm{E}\left[v_{\text{sN}}(\mathcal{S}^j(\mathbf{x}))\right]-1\right)+o(\epsilon^k) \quad (33)$$

where $\mathcal{S}^1(\mathbf{x})$ is the state that succeeds state $\mathbf{x}$ in the event of a successful transmission when policy f is applied, i.e.,

$$\mathcal{S}^1(\mathbf{x}) := (x_1+1,\ldots,x_{f(\mathbf{x})-1}+1,0, x_{f(\mathbf{x})+1}+1,\ldots,x_N+1)\wedge\boldsymbol{\tau};$$

also

$$\mathcal{S}^{j+1}(\mathbf{x}) := \mathcal{S}\left(\mathcal{S}^j(\mathbf{x})\right), \quad j=1,2,\cdots;\mathcal{S}^0(\mathbf{x}) := \mathbf{x},$$

and k is an integer such that there exist $d_1, d_2 > 0$ satisfying $d_1\epsilon^k \leq \mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})]-1 \leq d_2\epsilon^k$ whenever ϵ is sufficiently small.

Proof: The proof is by arguments similar to those in the two-client scenario (proof of (18), (20)–(24) in Section VI-B, VI-C and Lemma 3). □

By (32) and (33), one may note that the average cost J and $\mathrm{E}[\tilde{v}_{\text{sN}}(\cdot)]$ are closely related. This can be seen by noting that $\mathcal{S}^j\big(\mathbb{X}_{\text{sN}}(1)\big) = \mathbb{X}_{\text{sN}}(j+1), \forall j = 1,\ldots,|\mathbb{X}_{\text{sN}}|-1$ and $|\mathbb{X}_{\text{sN}}| \geq N$. Thus, the r.h.s. of (33) and the r.h.s. of (32) are closely related. By these observations, it is reasonable to propose the following assumption.

Assumption 1: A stationary policy that minimizes $\mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})]$ for each system state $\mathrm{x} \in \mathbb{Y}$, also minimizes $J(f,\mathbf{x})$.

Motivated by Assumption 1, Algorithm 1 shown alongside is proposed, which determines a stationary policy called an *"SN policy"*. Here,

$$\tilde{\mathcal{S}}_n(\mathbf{x}) := (x_1+1,\ldots,x_{n-1}+1,0, x_{n+1}+1,\cdots,x_N+1)\wedge\boldsymbol{\tau},$$

i.e., $\tilde{\mathcal{S}}_n(\mathbf{x})$ is the state that succeeds state $\mathbf{x}$ in the event of a successful transmission under control $U(t) = n$.

In words, Algorithm 1 tends to minimize $\mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})]$ for any system state $\mathbf{x} \in \mathbb{Y}$. This is accomplished in the following way: First, the algorithm classifies system states into sets, $\mathbb{Y}_0, \mathbb{Y}_1, \ldots, \mathbb{Y}_{\min_{n=1}^N\{\tau_n\}}$, where,

$$\mathbb{Y}_k := \{\mathbf{x} : \exists A > 0, \min_{\pi}\left(\mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})]\right) = 1 + A\epsilon^k + o(\epsilon^k)\}, \quad (34)$$

for $k = 0,1,\ldots,\min_{n=1}^N\{\tau_n\}$. Then, it determines or approximates the coefficient A in (34) for each system state $\mathbf{x}$ (denoted $A(\mathbf{x})$), and finally obtains the optimal policy based on $A(\mathbf{x})$.

Algorithm 1 SN Policy Algorithm

input : N, θ, $\tau_1,\ldots,\tau_N$, $b_1,\ldots,b_N$.
output: Policy $g(\mathbf{x}), \forall\mathbf{x} \in \mathbb{Y}$.
1 $\mathbb{Y}_0 = \{\mathbf{x} : \exists A > 0, \min_\pi\left(\mathrm{E}[\tilde{v}_{\text{sN}}(\mathbf{x})]\right) = 1 + A + o(1)\}$;
2 **foreach** $\mathbf{x} \in \mathbb{Y}_0$ **do**
3 $A(\mathbf{x})$ is as in Step 1;
4 $g(\mathbf{x}) \leftarrow \arg\min_{n=1}^N A(\tilde{\mathcal{S}}_n(\mathbf{x}))$;
5 $\mathbb{Z} \leftarrow \emptyset$; $\mathbb{Y}_{\text{remain}} \leftarrow \emptyset$
6 **foreach** $k = 1$ *to* $(\min_{n=1}^N \tau_n)$ **do**
7 $\mathbb{Z} \leftarrow \mathbb{Z} \cup \mathbb{Y}_{k-1}$;
8 $\mathbb{Y}_k \leftarrow \{\mathbf{x} : \mathbf{x}+\mathbf{1} \in \mathbb{Y}_{k-1}$ and $\mathbf{x} \notin \mathbb{Z}\}$;
9 **repeat** $\mathbb{Y}_k \leftarrow \mathbb{Y}_k \cup \{\mathbf{x} : \tilde{\mathcal{S}}_n(\mathbf{x}) \in \mathbb{Z} \cup \mathbb{Y}_k, \forall n$ and $\mathbf{x} \notin \mathbb{Z} \cup \mathbb{Y}_k\}$ **until** Y_k *not extend;*
10 **foreach** $k = 1$ *to* $(\min_{n=1}^N \tau_n)$ **do**
11 $\mathbb{Y}'_k \leftarrow \mathbb{Y}_k$;
12 **repeat**
13 $\mathbb{Y}''_k \leftarrow \mathbb{Y}'_k$; $\mathbb{Y}'_k \leftarrow \emptyset$;
14 **foreach** $\mathbf{x} \in \mathbb{Y}''_k$ **do**
15 $m \leftarrow \max\{j : \exists n, \tilde{\mathcal{S}}_n(\mathbf{x}) \in \mathbb{Y}_j\}$;
16 $U_{\text{set}} \leftarrow \{n : \tilde{\mathcal{S}}_n(\mathbf{x}) \in \mathbb{Y}_m\}$;
17 **if** $m > k$ **then**
18 $[A(\mathbf{x}), g(\mathbf{x})] \leftarrow \min_{n\in U_{\text{set}}} b_n A((\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau})$;
19 **else if** $\exists n \in U_{set}$, $A(\tilde{S}_n(\mathbf{X}))$ *not yet* **then**
20 $\mathbb{Y}'_k \leftarrow \mathbb{Y}'_k \cup \mathbf{x}$;
21 **else**
22 $[A(\mathbf{x}), g(\mathbf{x})] \leftarrow \min_{n\in U_{\text{set}}} A(\tilde{\mathcal{S}}_n(\mathbf{x})) + b_n A((\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau})\mathbb{1}\{(\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau} \in \mathbb{Y}_{k-1}\}$;
23 **until** $\mathbb{Y}'_k = \emptyset$ *or* $\mathbb{Y}'_k = \mathbb{Y}''_k$;
24 **if** $\mathbb{Y}'_k \neq \emptyset$ **then**
25 $\mathbb{Y}_{\text{remain}} \leftarrow \mathbb{Y}_{\text{remain}} \cup \mathbb{Y}'_k$; $B(\mathbf{x}) \leftarrow 0, \forall\mathbf{x} \in \mathbb{Y}_k$;
26 **foreach** $n = 1$ *to* $N-1$ **do**
27 **foreach** $\mathbf{x} \in \mathbb{Y}_k$ **do**
28 $U'_{\text{set}} \leftarrow g(\mathbf{x})$ or U_{set};
29 $B(\mathbf{x}) \leftarrow \min_{n\in U'_{\text{set}}} B(\tilde{\mathcal{S}}_n(\mathbf{x})) + b_n A((\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau})\mathbb{1}\{(\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau} \in \mathbb{Y}_{k-1}\}$
30 **foreach** $\mathbf{x} \in \mathbb{Y}'_k$ **do**
31 $[A(\mathbf{x}), g(\mathbf{x})] \leftarrow \min_{n\in U'_{\text{set}}} B(\tilde{\mathcal{S}}_n(\mathbf{x})) + b_n A((\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau})\mathbb{1}\{(\mathbf{x}+\mathbf{1})\wedge\boldsymbol{\tau} \in \mathbb{Y}_{k-1}\}$

Theorem 4: When Assumption 1 holds and $\mathbb{Y}_{\text{remain}}$ in Algorithm 1 is empty, the SN policy is asymptotically optimal in the high reliability asymptotic regime.

This directly follows from the design of the Algorithm. Also, later in Section IX, we provide an example, illustrated in Fig. 5 when these optimality conditions hold. [1]

VIII. Design of an Index-Based Policy

We now consider the scheduling problem in which $M < N$ clients are allowed to transmit packets simultaneously,

[1]Note that the Algorithm 1 can be further improved by using $\tilde{S}(\tilde{S}(\mathbf{x}))$ in Step 19–22.

using possibly orthogonal channels. We propose "index-based" policies, which are easily implementable and computationally simple.

Let $U(t) = (U_1(t), \ldots, U_N(t))$ be the control at time slot t, where $U_n(t) = 1$ if client n is selected to transmit in slot t, and $U_n(t) = 0$ otherwise. The constraint on the number of orthogonal channels is $\sum_{n=1}^{N} U_n(t) \leq M, \ \forall t$.

We propose an index-based policy similar to the Whittle Index policy, which is designed for the restless multi-armed bandit problem [17]. The key difference between Whittle's policy and our index-based policy is that unlike Whittle's policy, ours makes decisions in order to avert the risk. Another key difference is that while Whittle's Index policy draws inspiration from the solution to "relaxed problem" [17], the same is not true for the risk sensitive index-based policy proposed by us. In fact, it turns out that in the risk sensitive setting, even solving the relaxed problem is highly nontrivial, and no known methods exist for it.

The first step towards deriving our policy is to consider the following single-client "ω-subsidy" problem for each client n:

$$\min_{\pi} \lim_{T\to\infty} \frac{1}{\theta}\frac{1}{T} \ln \mathrm{E}_{\pi}\Big[\exp\Big(\theta \sum_{t=0}^{T-1} \mathbb{1}\left\{Y_n(t) = \tau_n\right\} - \omega \mathbb{1}\left\{U_n(t) = 0\right\} \Big)\Big], \quad (35)$$

where $Y_n(t) \in \{0, 1, \ldots, \tau_n\}$. That is, compared with the original problem, for each client, a subsidy ω is introduced to reduce the cost whenever client n is not selected to transmit. This is somewhat similar to the ω-subsidy problem in Whittle's policy design, and may be regarded as its risk-sensitive version.

Now, we focus on the ω-subsidy problem of a single client n. Thus, in the following, we omit the subscript n in p_n, τ_n, $U_n(t)$, and $Y_n(t)$ when no confusion is incurred. The following Lemmas show that the optimal policy for the ω-subsidy problem is of threshold type.

Lemma 5: Consider the single-client ω-subsidy problem defined in (35) with,

$$\omega < \frac{1}{\theta}\left[-\ln(1-p)\right]. \quad (36)$$

Then, the following results hold:

i) $\forall h \in \{1, 2, \ldots, \tau\}$, there exists a unique (ω, λ) pair satisfying the following equations,

$$\left[\mathrm{e}^{-\theta\omega} - (1-p)\right]\left(\lambda \mathrm{e}^{\theta\omega}\right)^{h} = p, \quad (37)$$

$$\left(\frac{\lambda}{1-p}\right)^{\tau-h}\left(\lambda \mathrm{e}^{\theta\omega}\right)^{h} - p\frac{1-\left(\frac{\lambda}{1-p}\right)^{\tau-h}}{1-p-\lambda} = \frac{p\,\mathrm{e}^{\theta}}{\lambda - \mathrm{e}^{\theta}(1-p)}, \quad (38)$$

Denote this pair of (ω, λ) by $W_1(h), \lambda_1(h)$.

ii) $\forall h \in \{0, 1, \ldots, \tau-1\}$, there exists a unique (ω, λ) pair satisfying (38) and

$$\left[\mathrm{e}^{-\theta\omega} - (1-p)\right]\frac{\lambda\left(\lambda \mathrm{e}^{\theta\omega}\right)^{h} - p}{1-p} = p. \quad (39)$$

Denote this pair of (ω, λ) by $W_2(h), \lambda_2(h)$.

Proof: The proof is omitted with more details in the online supplementary material. □

A threshold-type policy with a threshold $h \in \{0, 1, \ldots, \tau\}$ applies the following control,

$$U(t) = \begin{cases} 0, & \text{if } Y(t) < h; \\ 1, & \text{if } Y(t) \geq h. \end{cases}$$

Also, threshold value $h = 0$ refers to a policy which transmits irrespective of the state value, while a policy with threshold $h = \infty$ refers to one that never transmits. It can be shown that the problem (35) is solved by threshold-type policies.

Lemma 6: For the single-client ω-subsidy problem defined in (35) with inequality (36), an h-threshold policy is optimal if one of the following holds:

1) $h = 0$ and $\omega \leq W_2(0)$.
2) $h \in \{1, \ldots, \tau-1\}$ and $W_1(h) \leq \omega \leq W_2(h)$.
3) $h = \tau, \omega = W_1(\tau)$, and additionally

$$W_1(\tau) = \frac{\ln\left[p\mathrm{e}^{-\theta\tau} + (1-p)\right]}{-\theta}. \quad (40)$$

4) $h = \infty$ and $\omega \geq W_1(\tau)$.

Proof: The proof is omitted with more details in the online supplementary material. □

Now we define the notion of indexability for a risk sensitive MDP.

Definition 1: For the ω-subsidy problem (35), let $\mathbb{B}(\omega)$ denote the set of states where the optimal control is 0 (i.e., not to transmit). An ω-subsidy problem is indexable if $\omega_1 < \omega_2$ implies that $\mathbb{B}(\omega_1) \subset \mathbb{B}(\omega_2)$. The entire problem is indexable if each ω-subsidy problems associated with any client n is indexable.

If a problem is indexable, the *index* for a state i in the ω-subsidy problem is defined as the smallest value of subsidy ω under which the control 0 (not transmit) is optimal when the state value is i, i.e., the index is given by $\inf\{\omega : i \in \mathbb{B}(\omega)\}$. The *index-based policy* chooses the M clients which have the largest index values, and transmits their packets.

Theorem 5: For the ω-subsidy problem, the following results hold:

1) $W_1(h) = W_2(h-1), \ \forall h = 1, \ldots, \tau$;
2) $W_1(\tau)$ satisfies the inequality (36);
3) When the inequality (36) is violated, the h-threshold policy with $h = \infty$ is optimal.

As a result, the ω-subsidy problem is indexable, and the index for each state i is,

$$\begin{cases} W_2(i), & \text{for } i = 0, 1, \ldots, \tau-1; \\ W_1(\tau), & \text{for } i = \tau. \end{cases}$$

In addition, the MDP-2 is indexable.

Proof: For statement 1), first note that for $W_1(h)$ and $\lambda_1(h)$, equations (37) and (38) are satisfied. Also, for $W_2(h-1)$ and $\lambda_2(h-1)$, the following equations are satisfied,

$$\left[\mathrm{e}^{-\theta\omega} - (1-p)\right]\frac{\lambda\left(\lambda \mathrm{e}^{\theta\omega}\right)^{h-1} - p}{1-p} = p, \quad (41)$$

$$\left(\frac{\lambda}{1-p}\right)^{\tau-h-1}\left(\lambda \mathrm{e}^{\theta\omega}\right)^{h-1} - p\frac{1-\left(\frac{\lambda}{1-p}\right)^{\tau-h-1}}{1-p-\lambda} = \frac{p\,\mathrm{e}^{\theta}}{\lambda - \mathrm{e}^{\theta}(1-p)}. \quad (42)$$

Equations (41) and (42) are derived by replacing h by $h-1$ in (38) and (39). Recall that there are unique $W_1(h), \lambda_1(h)$ and unique $W_2(h-1), \lambda_2(h-1)$ by the statements i) and ii) in Lemma 6. Thus, in the following, we prove statement 1) by showing that (37) and (38) imply (41) and (42).

By (37),

$$\left(\lambda \mathrm{e}^{\omega\theta}\right)^h = \frac{p}{\mathrm{e}^{-\omega\theta} - (1-p)}. \quad (43)$$

Then, by substituting (43) into (41), we have,

$$\begin{aligned} \text{l.h.s. of (41)} &= \left[\mathrm{e}^{-\theta\omega} - (1-p)\right] \frac{\frac{p\mathrm{e}^{-\theta\omega}}{\mathrm{e}^{-\theta\omega}-(1-p)} - p}{1-p} \\ &= p = \text{r.h.s. of (41)}. \end{aligned} \quad (44)$$

Thus, (37) implies (41).

Further, in a similar way, it can be shown that (37), (38), and (41) imply (42). Thus, statement 1) holds.

Statement 2) simply follows from (40) and (36).

Statement 3) can be proved in a similar way as in the proof of Lemma 6, and we omit it.

The results for indexability and the value of the index simply follow from their definitions. □

Remark: Note that the numerical value of the index in Theorem 5 can be simply obtained by bisection search. Specifically, λ can be expressed as a function of ω by (37) or (39). Then, by substituting it into (38), we can obtain an equation for $W_1(h)$ or $W_2(h)$, which can be simply solved by bisection search.

Remark: In designing the index-based policy, we have generalized the Whittle Index policy for the risk-neutral case to a risk-sensitive restless multi-armed bandit problem. In the risk-neutral case, if a restless multi-armed bandit problem is indexable, the Whittle Index policy is proved to be asymptotic optimal when the total number of clients increases to infinity [30]. In this indexable risk-sensitive problem, from the numerical results, our index-based policy also performs well when the number of clients is sufficiently large, as shown in Fig. 7.

Online Learning: We briefly describe a method that can learn the indices online in case the equations (37)-(39) cannot be solved either because the parameters are unknown or time-varying, or because one wants to avoid the computational expense involved. As stated above, after having eliminated the λ by (37) or (39), we have a single equation involving only the parameter ω to determine the index. Solving this nonlinear equation with h set to i yields the index for state value i. Let us denote this equation by $G_{n,i}(\omega) = 0$, where the subscript is used to identify the client n and state value i. We can use a stochastic version of Newton's method [31] corresponding to solving the deterministic equation $G_{n,i}(\omega) = 0$, as follows,

$$\omega_{n,i}(k+1) = \Gamma\left\{\omega_{n,i}(k) - \alpha_k\left(\frac{G_{n,i}(\omega_{n,i}(k))}{G'_{n,i}(\omega_{n,i}(k))}\right)\right\}, \quad k = 1, 2, \ldots, \quad (45)$$

where $\alpha_k = 1/k$ is the step size of the k-th iteration, $G'_{n,i}(\omega)$ is the derivative of $G_{n,i}(\omega)$, and $\Gamma\{\cdot\}$ is the projection operator that maps the iterates onto the compact set $[0, -\frac{1}{\theta}\ln(1-p_n)]$. Projections are needed in order to keep the "noisy iterates" bounded, while the choice of step-sizes $\alpha_k = 1/k$ is standard in stochastic approximation algorithms and is necessary to smoothen out the noise. See [32] for further details. The scheduling is performed by choosing M clients having the largest indices $\omega_{n,Y_n(t)}$, where $Y_n(t)$ is the state of client n at time t. Thus, the method combines the updates of indices in (45) with the index-based policy. Let $\omega^\star$ be a root of the equation $G_{n,i}(\omega) = 0$, and ω_0 be the initial point for our iterations. The following conditions are needed in order to ensure the convergence of this online learning method,

1) There is a neighborhood of radius $\delta > 0$ around $\omega^\star$, denoted $\mathcal{N}_{\omega^\star}(\delta)$, such that $G'_{n,i}(\omega) \neq 0$, $\forall \omega \in \mathcal{N}_{\omega^\star}(\delta)$.
2) $G''_{n,i}(\omega)$ is continuous in $\mathcal{N}_{\omega^\star}(\delta)$.
3) The initial point ω_0 is sufficiently close to the root $\omega^\star$.

Then, the index values decided by the recursions (45) converge to the true value of the indices under the assumption that each client n is scheduled infinitely often when in state i. If this is not the case, one may artificially guarantee that such an assumption is true by choosing, with a small probability ϵ_t, a client uniformly at random, and with remaining $1-\epsilon_t$ probability according to the index rule. The resulting algorithm can be analyzed by the ODE method, as in [33] and [34].

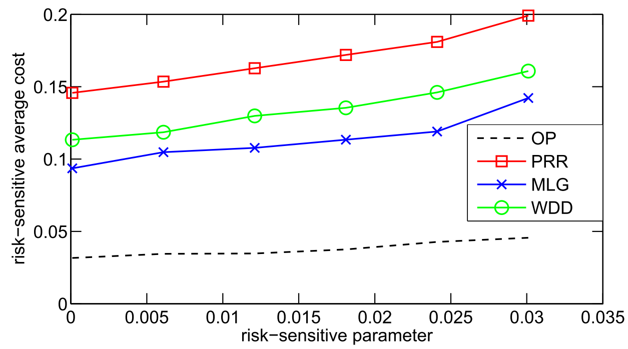


Fig. 3. Risk-sensitive average cost vs. risk-sensitive parameter θ for different wireless scheduling policies. (The parameters are $N = 2$, $p_1 = 0.4$, $p_2 = 0.1$, $\tau_1 = 20$, $\tau_2 = 40$).

IX. Simulations

We now present the results of a simulation study. In scenarios with heterogeneous inter-delivery thresholds and heterogeneous channel conditions, we implement the following scheduling policies and present their performance with respect to the risk-sensitive average cost:

- The optimal policy (OP) obtained by Theorem 2;
- The modified-least-time-to-go (MLG) policy (for the two-client scenario) proposed in Section VII-A;
- The SN policy proposed in Section VII-B;
- The index-based policy designed in Section VIII;
- The heuristic packet-level round-robin policy (PRR): a client keeps transmitting its packets until a successful transmission, then the next client takes its turn to transmit;
- The largest-weighted-delivery-debt (WDD) policy, which selects the client with the largest weighted *delivery debt* to transmit, where:

$$\text{Delivery Debt}_n = \frac{t}{p_n \tau_n} - \frac{M_t^{(n)}}{p_n}. \quad (46)$$

Note that, in (46), $M_t^{(n)}$ is the number of the packets delivered for client n by time t, as in (1). The WDD policy has been shown to be "timely-throughput" optimal [34].

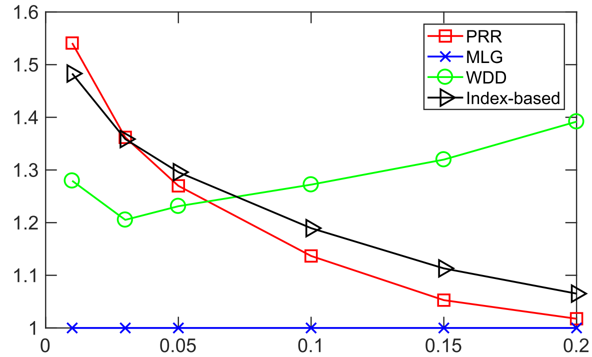


Fig. 4. Normalized risk-sensitive average cost (normalized by the cost of the optimal policy) vs. failure transmission parameter ϵ in a two-client scenario. ($p_1 = 1 - 2\epsilon$, $p_2 = 1 - \epsilon$, $\tau_1 = 3$, $\tau_2 = 5$, $\theta = 0.01$.)

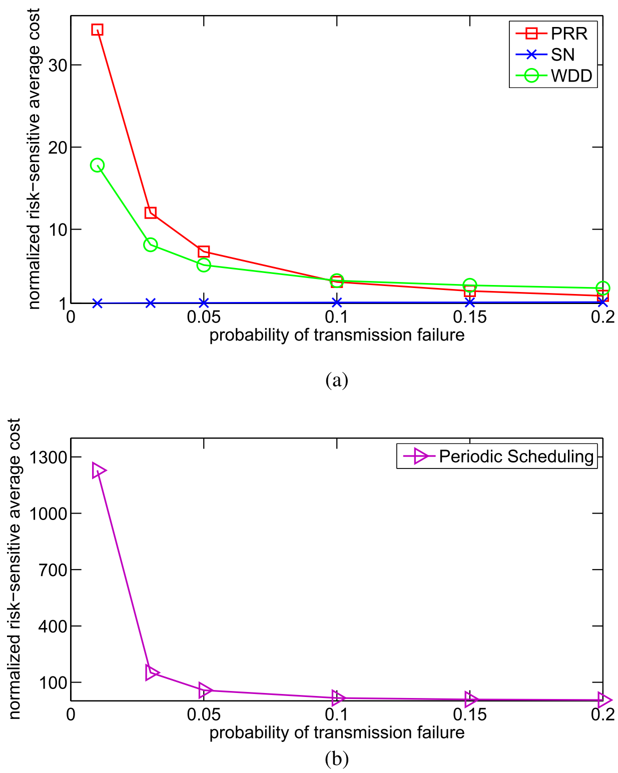


Fig. 5. Normalized risk-sensitive average cost (normalized by the cost of the optimal policy) vs. failure transmission parameter ϵ in a multi-client scenario. (The parameters are $N = 3$, $p_1 = p_2 = p_3 = 1 - \epsilon$, $\tau_1 = 4$, $\tau_2 = 6$, $\tau_3 = 8$, $\theta = 0.05$.) (a) Performance of PRR, WDD, and SN policies. (b) Performance of Periodic Scheduling.

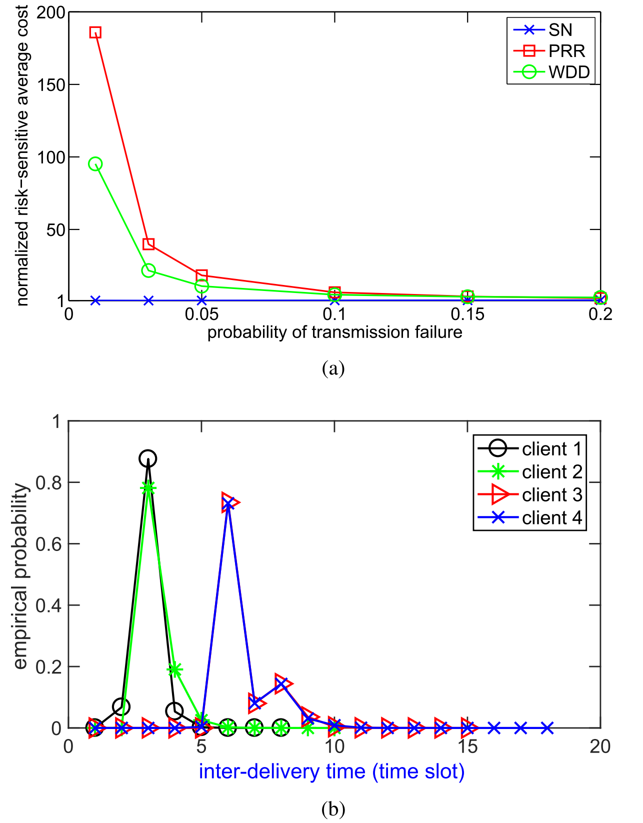


Fig. 6. Results for a multi-client scenario with parameters being $N = 4$, $p_1 = p_2 = p_3 = p_4 = 1 - \epsilon$, $\tau_1 = 4$, $\tau_2 = 5$, $\tau_3 = 9$, $\tau_4 = 10$, $\theta = 0.05$. (The risk-sensitive average cost is normalized by the cost of optimal policy.) (a) The normalized risk-sensitive average cost vs. ϵ. (b) Performance for the case of $\epsilon = 0.05$ with SN policy.

Fig. 3 demonstrates the costs of the scheduling policies in a two-client scenario for different values of the risk-sensitivity parameter θ. It can be seen that the optimal policy always outperforms all the other policies.

Fig. 4 demonstrates the risk-sensitive costs for different scheduling policies in a two-client scenario with different channel reliabilities. It can be seen that even for moderate channel reliabilities, e.g., $p_1 = 0.6$ and $p_2 = 0.8$, the MLG policy still achieves near-optimal performance, and has a smaller cost compared to all other heuristic policies. The performance of the index-based policy is also shown. As discussed in Section VIII and shown in Fig. 7, the index-based policy performs well only when the number of clients is sufficiently large. Therefore, it cannot be compared to the MLG policy in this two-client scenario.

Fig. 5 (a) demonstrates the risk-sensitive cost for different scheduling policies in a multi-client scenario with different channel reliabilities. It can be seen that even for moderate channel reliability probabilities, such as 0.8, the SN policy still achieves near-optimal performance, and has a smaller cost than all other heuristic policies. In Fig. 5 (b), the performance of the periodic scheduling (PS) policy [25] is also presented. The PS policy is known to be optimal when the failure probabilities are exactly zero. However, it can be seen that the PS policy has extremely poor performance even when the failure probability is very small, e.g., 0.01. This illustrates the importance of the asymptotic approach proposed: While there is no guarantee that a policy that is optimal for the case of zero channel failure probabilities also has a good performance when the failure probabilities are relatively small, in contrast, the policies resulting from the high-reliability asymptotic approach yield near optimal performance when the channel failure probabilities are sufficiently small.

In Fig. 6, we implement the policies in another multi-client scenario when the optimality condition for the SN policy stated in Theorem 4 is not satisfied. Fig. 6 (a) shows that in this case, the SN policy still has a near-optimal performance when the channel failure probability is small or even moderate. To be specific, the normalized average costs associated

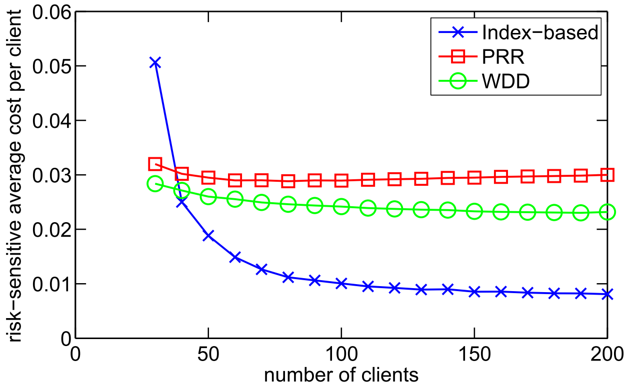


Fig. 7. The risk-sensitive average cost per client vs. the number of total clients for different policies.

with the SN policy are $1.031, 1.001, 1.046, 1.115, 1.167, 1.216$ for $\epsilon = 0.01, 0.03, 0.05, 0.10, 0.15, 0.20$, respectively. Also, by comparing with Fig. 5, we find that the advantage of the SN policy over other heuristic policies increases when the number of clients increases. Fig. 6 (b) demonstrates the empirical probabilities of the length of inter-delivery times for different clients when the SN policy is implemented, for the case of $p_1 = p_2 = p_3 = p_4 = 0.95$.

In Fig. 7, the performance of the index-based policy is compared with other heuristic policies under a scenario where at most 30% of the clients can transmit simultaneously. The clients in this scenario are classified into two classes: For any client n in the first class, $p_n = 0.6, \tau_n = 10$; while for any client n in the second class, $p_n = 0.8, \tau_n = 5$. Also, each class consists of exactly half of the clients. Further, θ is set to 0.5. Here, the performance of the optimal policy is not provided since its computational complexity is extremely high when the number of clients is large. However, it can be seen that the performance of the index-based policy is much better than that of other heuristic policies when the number of clients is sufficient large. In this case, when $N > 50$, the index-based policy already well outperforms other policies.

X. Conclusion

We have addressed the problem of designing scheduling policies to meet the packet inter-delivery time requirements of wirelessly connected clients in cyber-physical systems. We have proposed a novel risk-sensitive approach to severely penalize the large "exceedance" of the inter-delivery times over the desirable thresholds.

Although the resulting risk-sensitive MDP involves an infinite state space, it can be converted into an equivalent MDP that involves only a finite number of states. We thus establish the existence of a stationary optimal policy, and further determine an algorithm to obtain it in a finite number of steps.

To further simplify the computational complexity we have undertaken the following two approaches: i) When the channel reliabilities are close to 1 (high-reliability asymptotic approach) we have analyzed low complexity policies such as the modified-least-time-to-go (MLG) policy and the SN policy, and have proved that they are asymptotically optimal, ii) In the non-asymptotic case, we have extended risk-neutral index policies to the risk-sensitive scenario. We have also conducted a simulation study, and found that the proposed asymptotically optimal policies provide near-optimal performance even for moderate values of the failure probabilities, which justifies the high-reliability approach. We have also shown that the index-based policy well outperforms other heuristic policies when the number of clients is sufficiently large.

References

[1] X. Guo, R. Singh, P. R. Kumar, and Z. Niu, "A high reliability asymptotic approach for packet inter-delivery time optimization in cyber-physical systems," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, 2015, pp. 197–206.

[2] E. A. Lee, "Cyber physical systems: Design challenges," in *Proc. 11th IEEE Int. Symp. Object Compon.-Oriented Real-Time Distrib. Comput. (ISORC)*, May 2008, pp. 363–369.

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

[4] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proc. IEEE INFOCOM*, Mar. 2005, pp. 1723–1734.

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

[6] R. Singh, X. Guo, and P. R. Kumar, "Index policies for optimal mean-variance trade-off of inter-delivery times in real-time sensor networks," in *Proc. INFOCOM*, Apr./May 2015, pp. 505–512.

[7] R. Singh, I.-H. Hou, and P. R. Kumar, "Fluctuation analysis of debt based policies for wireless networks with hard delay constraints," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 2400–2408.

[8] R. Singh, I.-H. Hou, and P. R. Kumar, "Pathwise performance of debt based policies for wireless networks with hard delay constraints," in *Proc. IEEE 52nd Annu. Conf. Decis. Control (CDC)*, Dec. 2013, pp. 7838–7843.

[9] X. Guo, R. Singh, P. R. Kumar, and Z. Niu, "Optimal energy-efficient regular delivery of packets in cyber-physical systems," in *Proc. IEEE ICC*, Jun. 2015, pp. 3186–3191.

[10] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Manage. Sci.*, vol. 18, no. 7, pp. 356–369, 1972.

[11] K.-J. Chung and M. J. Sobel, "Discounted MDP's: Distribution functions and exponential utility maximization," *SIAM J. Control Optim.*, vol. 25, no. 1, pp. 49–62, 1987.

[12] W. H. Fleming and D. Hernández-Hernández, "Risk sensitive control of finite state machines on an infinite horizon. I," in *Proc. IEEE Conf. Decis. Control*, Dec. 1997, pp. 3407–3412.

[13] S. I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernández, S. Coraluppi, and P. Fard, "Risk sensitive Markov decision processes," in *Systems and Control in the Twenty-First Century*. Boston, MA, USA: Birkhäuser, 1997.

[14] R. Cavazos-Cadena and E. Fernández-Gaucherand, "Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions," *Math. Methods Oper. Res.*, vol. 49, no. 2, pp. 299–324, 1999.

[15] S. C. Jaquette, "A utility criterion for Markov decision processes," *Manage. Sci.*, vol. 23, no. 1, pp. 43–49, 1976.

[16] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[17] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, Jan. 1988.

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

[19] Y. Sadi and S. C. Ergen, "Optimal power control, rate adaptation, and scheduling for UWB-based intravehicular wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 219–234, Jan. 2013.

[20] R. Singh and A. Stolyar, "Maxweight scheduling: Asymptotic behavior of unscaled queue-differentials in heavy traffic," in *Proc. ACM SIGMETRICS*, Jun. 2015, pp. 431–432.

[21] R. Singh, X. Guo, and E. Modiano, "Risk-sensitive optimal control of queues," in *Proc. 56th IEEE Conf. Decis. Control*, Dec. 2017, pp. 3563–3568.

[22] E. Altman, V. Kavitha, F. De Pellegrini, V. Kamble, and V. Borkar, "Risk sensitive optimal control framework applied to delay tolerant networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3146–3154.

[23] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay-limited information," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 746–763, Feb. 2009.

[24] Y. Zhang and C. Tepedelenlioglu, "Applications of Tauberian theorem for high-SNR analysis of performance over fading channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 296–304, Jan. 2012.

[25] A. Bar-Noy, R. Bhatia, J. S. Naor, and B. Schieber, "Minimizing service and operation costs of periodic scheduling," in *Proc. ACM Symp. Discrete Algorithms*, 2002, pp. 11–20.

[26] D. Zhu, D. Mosse, and R. Melhem, "Multiple-resource periodic scheduling problem: How much fairness is necessary?" in *Proc. IEEE Real-Time Syst. Symp.*, Dec. 2003, pp. 142–151.

[27] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. Hoboken, NJ, USA: Wiley, 1994.

[28] S. M. Ross, *Applied Probability Models with Optimization Applications*. New York, NY, USA: Dover, 1992.

[29] D. E. Knuth, "Big omicron and big omega and big theta," *ACM SIGACT News*, vol. 8, no. 2, pp. 18–24, 1976.

[30] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *J. Appl. Probab.*, vol. 27, no. 3, pp. 637–648, Sep. 1990.

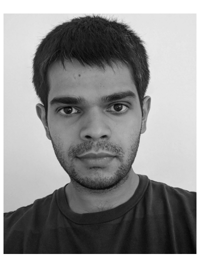
[31] K. E. Atkinson, *An Introduction to Numerical Analysis*. Hoboken, NJ, USA: Wiley, 2008.

[32] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

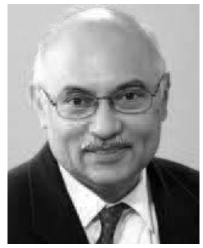
[33] V. S. Borkar and S. P. Meyn, "The O.D.E. Method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 447–469, 2000.

[34] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS for wireless," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 486–494.

Xueying Guo (S'14–M'17) received the B.E. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2011 and 2017, respectively. From 2013 to 2014, she was a Visiting Scholar with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. She is currently a Post-Doctoral Researcher with the Department of Computer Science, University of California at Davis, Davis, CA, USA. Her research interests include machine learning, reinforcement learning, and data-driven networking. She was a recipient of the Best Student Paper Award in the 25th IEEE International Teletraffic Congress in 2013.



Rahul Singh received the B.E. degree in electrical engineering from IIT Kanpur, Kanpur, India, in 2009, the M.S. degree in electrical engineering from the University of Notre Dame, South Bend, IN, USA, in 2011, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, in 2015. From 2015 to 2017, he was a Post-Doctoral Associate with the Laboratory for Information Decision Systems, Massachusetts Institute of Technology. He was then with Encoredtech as a Data Scientist. He is currently with Intel as a Deep Learning Engineer. His research interests include decentralized control of large-scale complex cyber-physical systems, operation of electricity markets with renewable energy, scheduling of networks serving real-time traffic, and deep learning.



P. R. Kumar (F'88) received the B.Tech. degree in electrical engineering (electronics) from IIT Madras, Chennai, India, in 1973, and the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, USA, in 1975 and 1977, respectively.

From 1977 to 1984, he was a Faculty Member with the Department of Mathematics, University of Maryland, Baltimore County. From 1985 to 2011, he was a Faculty Member with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois. He is currently with Texas A&M University, College Station, TX, USA, where he is the College of Engineering Chair in computer engineering. He was involved in problems in game theory, adaptive control, stochastic systems, simulated annealing, neural networks, machine learning, queuing networks, manufacturing systems, scheduling, wafer fabrication plants, and information theory. His current research interests are wireless networks, sensor networks, cyber-physical systems, and the convergence of control, communication, and computation.

He was a Guest Chair Professor with Tsinghua University, Beijing, China. He is an Honorary Professor at IIT Hyderabad. He is a member of the National Academy of Engineering of the USA and the Academy of Sciences of the Developing World. He received the IEEE Field Award for control systems, the Donald P. Eckman Award of the American Automatic Control Council, the Fred W. Ellersick Prize of the IEEE Communications Society, the Outstanding Contribution Award of ACM SIGMOBILE, and the Daniel C. Drucker Eminent Faculty Award from the College of Engineering, University of Illinois. He received the Honorary Doctorate from ETH Zurich.

Zhisheng Niu (F'12) graduated from Beijing Jiaotong University, China, in 1985, and received the M.E. and D.E. degrees from the Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. From 1992 to 1994, he was with Fujitsu Laboratories Ltd., Japan, and in 1994, he joined Tsinghua University, Beijing, China, where he is currently a Professor with the Department of Electronic Engineering. He is also a Guest Chair Professor with Shandong University, China. His major research interests include queuing theory, traffic engineering, mobile Internet, radio resource management of wireless networks, and green communication and networks.

Dr. Niu has been an active volunteer for various academic societies, including the Director for Conference Publications from 2010 to 2011 and the Director for the Asia-Pacific Board of the IEEE Communication Society from 2008 to 2009, a Membership Development Coordinator of the IEEE Region 10 from 2009 to 2010, the Councilor of IEICE-Japan from 2009 to 2011, and a Council Member of the Chinese Institute of Electronics from 2006 to 2011. He was a Distinguished Lecturer from 2012 to 2015 and the Chair of the Emerging Technology Committee from 2014 to 2015 of the IEEE Communication Society, a Distinguished Lecturer of the IEEE Vehicular Technologies Society from 2014 to 2016, a member of the Fellow Nomination Committee of the IEICE Communication Society from 2013 to 2014, a Standing Committee Member of the Chinese Institute of Communications from 2012 to 2016, and the Associate Editor-in-Chief of the IEEE/CIC joint publication *China Communications*.

Dr. Niu is a Fellow of the IEEE and the IEICE. He received the Outstanding Young Researcher Award from the Natural Science Foundation of China in 2009 and the Best Paper Award from the IEEE Communication Society Asia-Pacific Board in 2013. He was a co-recipient of the Best Paper Awards from the 13th, 15th, and 19th Asia-Pacific Conference on Communication in 2007, 2009, and 2013, respectively, and the International Conference on Wireless Communications and Signal Processing (WCSP'13), and the Best Student Paper Award from the 25th International Teletraffic Congress. He was the Chief Scientist of the National Basic Research Program (so called 973 Project) of China on the Fundamental Research on the Energy and Resource Optimized Hyper-Cellular Mobile Communication System from 2012 to 2016, which is the first national project on green communications in China.