Age Optimal Sampling Under Unknown Delay Statistics

Haoyue Tang, *Student Member, IEEE*, Yuchao Chen, Jintao Wang, *Senior Member, IEEE*, Pengkun Yang, and Leandros Tassiulas, *Fellow, IEEE*

***Abstract*—This paper revisits the problem of sampling and transmitting status updates through a channel with random delay under a sampling frequency constraint. We use the Age of Information (AoI) to characterize the status information freshness at the receiver. The goal is to design a sampling policy that can minimize the average AoI when the statistics of delay is unknown. We reformulate the problem as the optimization of a renewal-reward process, and propose an online sampling strategy based on the Robbins-Monro algorithm. We prove that the proposed algorithm satisfies the sampling frequency constraint. Moreover, when the transmission delay is bounded and its distribution is absolutely continuous, the average AoI obtained by the proposed algorithm converges to the minimum AoI when the number of samples K goes to infinity with probability 1. We show that the optimality gap decays with rate $\mathcal{O}(\ln K/K)$, and the proposed algorithm is minimax rate optimal. Simulation results validate the performance of our proposed algorithm.**

***Index Terms*—Age of information, minimax optimality, online learning, renewal-reward process.**

I. Introduction

With the proliferation of autonomous vehicles and intelligent manufacturing, status updates are becoming a larger part of communications [3]. Status updates are crucial to the efficient control and monitoring in such applications, and therefore should be delivered to the destination as timely as possible. To measure the timeliness of status update information at the receiver, the Age of Information (AoI), or simply Age is proposed [4]. Since then, the design of Age optimal transmission and sampling strategies under communication constraints has received wide attention.

When the transmission statistics (e.g., delay distribution, packet-loss probabilities) are known in advance, designing AoI minimum transmission strategies can be formulated into a Markov decision process (MDP) [2], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. When the generation of status updates are controlled by external sources, AoI minimum cross-layer scheduling and transmission have been studied in [5], [6], and [7]; When the generation process can be controlled at will, the joint sampling and transmission of status update packets have been studied in [9], [10], [11], and [8]. In continuous time scenarios, by modeling the external status update generation as a random process, the expected AoI performance under different service disciplines are analyzed in [12] and [13].

Designing AoI minimum sampling strategies in an unknown environment can be formulated as a sequential decision making problem, where online and reinforcement learning algorithms can be employed [16], [17], [18], [19], [20]. When the generation of status update packets is controlled by external sources, AoI minimum adaptive packet scheduling and link selection algorithms have been proposed [16], [17], and [18]. Tripathi et al. model the timeliness of status updates to be a time-varying function of the AoI [19], and a robust online learning algorithm is proposed. When the status update packets can be generated at will, [20] models the data freshness requirement as a minimum AoI constraint, and proposes scheduling algorithms that can achieve a sub-linear utility regret while satisfying the AoI constraint. However, the ultimate goal in [20] is to optimize the total utility over the entire network, rather than the AoI performance. Designing Age optimal sampling and transmission strategies have been studied in [10], [21], [22], [23], and [24], where various deep reinforcement learning algorithms (e.g., SARSA, Actor-Critic, Q-Learning) have been employed. However, the convergence rate of those algorithms are not well understood. Although the online sampling strategies proposed in [25] and [26] is shown to converge to the optimum strategy almost surely, the optimality of the algorithm is not known.

Manuscript received 31 March 2022; revised 27 September 2022; accepted 15 December 2022. Date of publication 2 January 2023; date of current version 20 January 2023. The work of Haoyue Tang was supported by NSF, AI Institute for Edge Computing Leveraging Next Generation Networks (Athena), under Grant CNS-2112562. The work of Yuchao Chen and Jintao Wang was supported in part by the Tsinghua University-China Mobile Research Institute Joint Innovation Center. The work of Pengkun Yang was supported by NSFC under Grant 12101353 and in part by the Tsinghua University Initiative Scientific Research Program. The work of Leandros Tassiulas was supported in part by NSF under Grant CNS-2112562 and in part by the Office of Naval Research NSF CNS-2112562. An earlier version of this paper was presented in part at the IEEE Infocom2022 [DOI: 10.1109/INFOCOM48880.2022.9796970]. *(Corresponding author: Jintao Wang.)*

Haoyue Tang was with the Department of Electronic Engineering, Tsinghua University, Beijing 100190, China. She is now with the Institute of Network Science, Yale University, New Haven, CT 06520 USA (e-mail: haoyue.tang@yale.edu).

Yuchao Chen is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with the Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China (e-mail: cyc20@mails.tsinghua.edu.cn).

Jintao Wang is with the Department of Electronic Engineering, Research Institute, Tsinghua University, Beijing 100084, China, and also with the Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China (e-mail: wangjintao@mail.tsinghua.edu.cn).

Pengkun Yang is with the Center for Statistical Science, Tsinghua University, Beijing 100190, China (e-mail: yangpengkun@tsinghua.edu.cn).

Leandros Tassiulas is with the Department of Electrical Engineering, Institute for Network Science, Yale University, New Haven, CT 06520 USA (e-mail: leandros.tassiulas@yale.edu).

Communicated by A. Eryilmaz, Associate Editor At Large for Networking and Computation.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3233782.

Digital Object Identifier 10.1109/TIT.2022.3233782

In general, although there is a growing number of literature on Age optimal transmission in unknown environment, how to design effective generate-at-will sampling strategies with theoretical guarantees is not well understood. To answer this question, we revisited the point-to-point status update system (Fig. 1) in [2] and [27], where a sensor samples and transmits update packets to the destination through a channel with a random delay. The goal is to design an online sampling strategy that minimizes the average AoI at the destination when the delay statistics is unknown. The contributions of the paper are as follows:

- Our work is the first to design a Robbins-Monro based online policy to minimize the average AoI when the delay statistics is unknown. Moreover, by using the Lyapunov-Drift-Plus-Penalty approach, our algorithm can satisfy the sampling frequency constraint concurrently (Theorem 5).
- When there is no sampling constraint, we show that the time-averaged AoI of the proposed algorithm converges to the limit point of an ordinary differential equation (ODE) almost surely. By showing that the limit point of the ODE is unique and stationary, we prove that the time-averaged AoI obtained by the proposed algorithm converges to the minimum AoI with probability 1 (Theorem 2). The optimality gap of the proposed online learning algorithm decays with rate $\mathcal{O}(\ln K/K)$, where K is the total number of samples (Theorem 3).
- By using the Le Cam's two point method from non-paramatric statistics, we show that under the worst case delay distribution, the gap between the average AoI of any online learning algorithm and the minimum AoI with known delay statistics decays with rate larger than $\Omega(\ln K/K)$, where K is the total number of samples (Theorem 4). Both the mathematics tool and the converse result are novel in the field of stochastic approximation. We show that the convergence rate of the proposed algorithm (Theorem 3) is minimax order optimum.

Independent of this work, [26] proposes a similar Robbins-Monro algorithm to minimize the average AoI penalty for a two-way delay communication system. It is worth noting that, by using the sampling frequency debt as a dual optimizer, our modified Robbins-Monro algorithm satisfies the sampling frequency constraint at the transmitter side. Our algorithm can be extended to the problem of minimizing the average AoI penalty with a sampling frequency constraint, because computing the optimal updating threshold is equivalent to solving an equation. Moreover, the proof techniques for almost sure convergence are different, with ours using the ODE method. We further establish the minimax lower bound of the average AoI gap of any online algorithm.

II. Problem Formulation

A. System Model

Similar to [2] and [27], we consider a status update system depicted in Fig. 1, where a sensor observes a time sensitive process, samples status updates and sends them to the destination through a channel. The channel transmits update packets based on a First-Come-First-Serve (FCFS) principle, and each update packet experiences a random transmission delay. Due to the transmission delay, update packets may have to wait in the queue before the last transmission finishes. Once the packet is received by the destination, an acknowledgement (ACK) will be received by the sensor immediately.

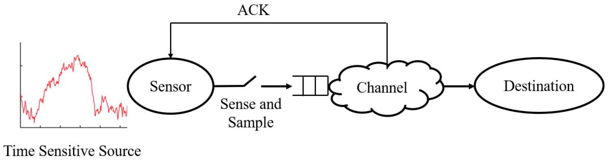

Fig. 1. A point-to-point status update system.

Similar to [15], suppose the sensor can sample update packets at any time $t \in \mathbb{R}^+$ at his own will. The sampling time-stamp and channel transmission delay of the k-th sampled packet are denoted by S_k and D_k, respectively. We assume each transmission delay $D_k, k \in \{1, 2, \cdots\}$ is identically and independently distributed (i.i.d.) following the probability measure $\mathbb{P}_D$.

Assumption 1: The probability measure $\mathbb{P}_D$ is absolutely continuous on $[0, \infty)$. Its expectation and second order moment is bounded, i.e.,

$$0 < \overline{D}_{\mathsf{lb}} \leq \overline{D} \triangleq \mathbb{E}_{\mathbb{P}_D}[D] \leq \overline{D}_{\mathsf{ub}} < \infty, \tag{1a}$$

$$0 < M_{\mathsf{lb}} \leq \mathbb{E}_{\mathbb{P}_D}[D^2] \leq M_{\mathsf{ub}} < \infty. \tag{1b}$$

Let R_k be the reception time-stamp of the k-th update packet. Notice that the service of the k-th packets starts at $\max\{R_{k-1}, S_k\}$, therefore, R_k can be computed recursively through equation $R_k = \max\{R_{k-1}, S_k\} + D_k$. If the transmission of the $(k-1)$-th update packet has not finished before the k-th update packet has been sampled, i.e., $R_{k-1} > S_k$, the k-th packet has to wait in the queue and then becomes stale. Therefore, to keep information at the destination fresh, it is better to *wait* for the ACK of the $(k-1)$-th update packet before sampling the k-th packet, i.e., $S_k \geq R_{k-1}$. By using such a *waiting* policy, the reception time-stamp of the k-th update packet can be simplified to $R_k = S_k + D_k$. We denote $W_k := S_{k+1} - R_k$ to be the *waiting* time after receiving the k-th sample.

B. Age of Information

AoI measures the time elapsed since the freshest information stored at the destination is generated [4]. Let $i(t) := \arg\max\{k \in \mathbb{N}^+ | R_k \leq t\}$ be the index of the latest sample received by the destination before time t. The AoI at time t, denoted by $A(t)$ is:

$$A(t) := t - S_{i(t)}. \tag{2}$$

A sample path of AoI evolution is depicted in Fig. 2.

C. Optimization Problem Formulation

We aim at minimizing the average AoI by designing a sampling strategy $\pi \triangleq \{W_1, W_2, \cdots\}$. Specifically, we only focus on the class of "*causal*" policies Π, where the waiting time W_k

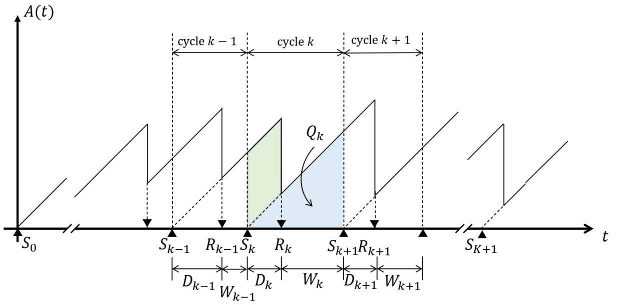

Fig. 2. Illustration of AoI evolution.

is selected based on the past delay and sampling time-stamps denoted by $\mathcal{H}_{k-1} := \{(S_i, D_i)\}_{i=1}^{k-1}$. No future information $\{D_i\}_{i>k}$ can be used for decision making. To facilitate further analysis, assume that each waiting time is upper bounded by W_{ub}, and denote Π as the class of causal policies whose waiting time $W_k \in [0, W_{\mathsf{ub}}]$.[1] Let K be the total number of sampling times. The expected time average AoI using policy π is defined by[2]:

$$\overline{A}_\pi \triangleq \limsup_{K\to\infty} \frac{\mathbb{E}\left[\int_{t=0}^{S_{K+1}} A(t)\mathrm{d}t\right]}{\mathbb{E}[S_{K+1}]}, \tag{3}$$

where the AoI $A(t)$ is determined by both the transmission delay $\{D_k\}$ and sampling strategy π.

To facilitate further computation and analysis, we define "*cycle*" k to be the time interval between the k-th and the $(k+1)$-th sampling time-stamps. Since the transmission delay D_k in each cycle k is i.i.d., we have $\mathbb{E}[S_{K+1}] = \mathbb{E}\left[\sum_{k=1}^{K}(D_k + W_k)\right]$. Similarly, let $X_k := \int_{t=S_k}^{S_{k+1}} A(t)\mathrm{d}t$ be the cumulative AoI in cycle k, which is the sum of the area of a parallelogram and a triangle, i.e.,

$$X_k = (D_{k-1} + W_{k-1})D_k + \frac{1}{2}(D_k + W_k)^2.$$

Then the cumulative AoI over interval $[0, S_{K+1})$ can be rewritten as a sum of X_k, i.e.,

$$\begin{aligned}\mathbb{E}\left[\int_{t=0}^{S_{K+1}} A(t)\mathrm{d}t\right] &= \mathbb{E}\left[\sum_{k=1}^{K} X_k\right]\\ =&\mathbb{E}\left[\sum_{k=1}^{K} q(D_{k-1}, W_{k-1}, D_k, W_k)\right],\end{aligned} \tag{4}$$

where function q is defined as follows:

$$q(d', w', d, w) := (d' + w')d + \frac{1}{2}(d + w)^2.$$

Designing the optimum strategy π that minimizes the expected average AoI can be formulated as the following optimization problem:

Problem 1:

$$\mathsf{AoI}_{\mathsf{opt}} \triangleq \inf_{\pi\in\Pi} \limsup_{K\to\infty} \frac{\mathbb{E}\left[\sum_{k=1}^{K} q(D_{k-1}, W_{k-1}, D_k, W_k)\right]}{\mathbb{E}\left[\sum_{k=1}^{K}(D_k + W_k)\right]}, \tag{5a}$$

$$\text{s.t. } \liminf_{K\to\infty} \frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K}(D_k + W_k)\right] \geq \frac{1}{f_{\mathsf{max}}}, \tag{5b}$$

where f_{max} is the maximum time average sampling frequency the status update system can afford due to various resource constraints (i.e., energy or system operation frequency).

Let $\pi^\star$ be the optimum policy that achieves $\mathsf{AoI}_{\mathsf{opt}}$. According to [2], policy $\pi^\star$ has a threshold structure. When the delay distribution $\mathbb{P}_D$ is known, Sun et al. [2] proposed to compute the optimum threshold through a bi-section search. In this paper, we assume only the lower and upper bounds of the average delay and second order moment $\overline{D}_{\mathsf{lb}}, \overline{D}_{\mathsf{ub}}, M_{\mathsf{lb}}$ and M_{ub} can be used at the transmitter.[3] The closed form expression of distribution $\mathbb{P}_D$ is not accessible to the transmitter and hence cannot be used for decision making.

III. Problem Resolution

In this section, we will first reformulate Problem 1 into a renewal-reward process. In Section III-B, we then propose an adaptive sampling strategy that can learn the optimum policy $\pi^\star$ when the number of samples goes to infinity. The theoretical performance of the algorithm is analyzed in Section III-C.

A. A Renewal-Reward Process Reformulation

A policy $\pi \in \Pi$ is *stationary deterministic* if the waiting time W_k is a stationary mapping from the transmission delay D_k, i.e., $W_k = w(D_k)$ and function $w : [0, \infty) \mapsto [0, W_{\mathsf{ub}}]$ is a deterministic function that specifies the waiting time. Let $\Pi_{\mathsf{SD}} \subseteq \Pi$ be the set of stationary deterministic policy such that:

$$\Pi_{\mathsf{SD}} \triangleq \{\pi \in \Pi : W_k = w(D_k), \forall k\}.$$

When $\mathbb{P}_D$ is known, we then have the following theorem according to [2]:

Theorem 1: [2, Theorem 2 Restated] There is a stationary deterministic policy $\pi^\star \in \Pi_{\mathsf{SD}}$ that is optimal to Problem 1.

With slight abuse of notations, we denote $\pi(d)$ to be the waiting time selection function of a stationary deterministic policy by observing transmission delay d. With Theorem 1, denote L_2 to be the Lebesgue space. Searching for the optimum stationary deterministic policy $\pi^\star$ that achieves $\mathsf{AoI}_{\mathsf{opt}}$ in Problem 1 can be reformulated into Problem 2 as follows:

[1] The assumption is reasonable and will not hurt the optimality in policy design when the upper bound W_{ub} is selected to be large. This is because waiting for an infinitely long time is not beneficial to AoI minimization.

[2] Another definition of the time average AoI can be the limits of expected total AoI over an observation window $[0, T)$ divide the length of the window T, i.e., $\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T A(t)\mathrm{d}t\right]$. The two definitions are both reasonable. Specifically, when π is a stationary randomized policy such that the Markov chain $\{(D_k, W_k)\}$ has only one ergodic class, the two definitions are equal [28].

[3] This assumption is reasonable since $\overline{D}_{\mathsf{lb}}$ and M_{lb} can be computed using the header time, and $\overline{D}_{\mathsf{ub}}$, M_{ub} can be computed using the maximum Round Trip Time (RTT).

Problem 2 (Renewal-Reward Process Optimization Reformulation):

$$\mathsf{AoI}_{\mathsf{opt}} = \inf_{\pi \in L_2} \left(\frac{\mathbb{E}[\frac{1}{2}\left(D+\pi(D)\right)^2]}{\mathbb{E}[D+\pi(D)]} + \overline{D} \right), \tag{6a}$$

$$\text{s.t. } \mathbb{E}\left[D+\pi(D)\right] \geq \frac{1}{f_{\mathsf{max}}}. \tag{6b}$$

The detailed derivation is the same as [2] and is hence omitted. Problem 2 can be viewed as the optimization of a renewal-reward process in the sense that:

- The delay D_k observed in each cycle k is i.i.d. following distribution $\mathbb{P}_D$.
- Let $L_k := D_k + \pi(D_k)$ be the length of the k-th cycle. Since D_k is i.i.d. and $\pi(\cdot)$ is a deterministic function, L_k is an i.i.d. random variable.
- Denote $Q_k := \frac{1}{2}\left(D_k + \pi(D_k)\right)^2$, which can be viewed as the reward received in cycle k. Due to the i.i.d. assumption of D_k, the reward Q_k is also an i.i.d. random variable.

As a result, the length and reward (L_k, Q_k) in frame k is independent of $(L_{k'}, Q_{k'})$ in other frames $k' \neq k$. Moreover, the expectation $\mathbb{E}[L_k] \leq \mathbb{E}[D + W_{\mathsf{ub}}] < \infty$ and $\mathbb{E}[Q_k] \leq \mathbb{E}[\frac{1}{2}(D+W_{\mathsf{ub}})^2] < \infty$ are both bounded. Problem 2 cast into the renewal-reward process optimization framework.

B. Proposed Online Algorithm

We will first review the computation of $\pi^\star$ when the delay statistics $\mathbb{P}_D$ is known, and then propose an online algorithm that learns policy $\pi^\star$ adaptively. For simplicity, let Π_{cons} be the set of stationary deterministic policies whose sampling frequency is below f_{max}, i.e.,

$$\Pi_{\mathsf{cons}} \triangleq \{\pi \in \Pi_{\mathsf{SD}} | \mathbb{E}\left[D+\pi(D)\right] \geq \frac{1}{f_{\mathsf{max}}}\}.$$

1) Design $\pi^\star$ With Known $\mathbb{P}_D$: Recall that $\overline{A}_{\pi^\star}$ is the minimum time average AoI any policy $\pi \in \Pi_{\mathsf{cons}}$ can achieve, i.e.,

$$\overline{A}_\pi = \frac{\mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2\right]}{\mathbb{E}[D+\pi(D)]} + \overline{D} \geq \overline{A}_{\pi^\star}. \tag{7}$$

Deducting $\overline{D}$ on both sides of inequality (7), we have:

$$\frac{\mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2\right]}{\mathbb{E}[D+\pi(D)]} \geq \overline{A}_{\pi^\star} - \overline{D}. \forall \pi \in \Pi_{\mathsf{cons}}. \tag{8}$$

For simplicity, denote $\gamma^\star = \overline{A}_{\pi^\star} - \overline{D}$ and then then multiplying $\mathbb{E}[D+\pi(D)]$ on both sides of inequality (8), we then have the following inequality:

$$\frac{1}{2}\mathbb{E}[(D+\pi(D))^2] - \gamma^\star \mathbb{E}[D+\pi(D)] \geq 0, \forall \pi \in \Pi_{\mathsf{cons}}. \tag{9}$$

Notice that (9) takes equality if and only if policy π is AoI minimum. Therefore, when $\gamma^\star$ is known, $\pi^\star$ can be obtained by solving the following functional optimization problem:

Problem 3 (Functional Optimization Problem):

$$\theta_{\mathsf{opt}} \triangleq \min_{\pi \in \Pi_{\mathsf{SD}}} \mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2 - \gamma^\star(D+\pi(D))\right], \tag{10a}$$

$$\text{s.t. } \mathbb{E}[D+\pi(D)] \geq \frac{1}{f_{\mathsf{max}}}. \tag{10b}$$

Inequality (9) shows $\theta_{\mathsf{opt}} = 0$. To find the optimum policy that achieves θ_{opt}, we place the sampling frequency constraint (10b) into the objective function (10a) using a dual optimizer $\underline{\nu \geq 0}$, we can formulate the Lagrange function as follows:

$$\begin{aligned}\mathcal{L}(\gamma,\nu,\pi) :=& \mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2 - (\gamma+\nu)(D+\pi(D))\right] \\ &+ \nu\frac{1}{f_{\mathsf{max}}}.\end{aligned} \tag{11}$$

As is shown in [2, Theorem 4], for fixed γ and ν, the optimum policy $\pi^\star_{\gamma,\nu}$ that minimizes the Lagrange function (11) specifies the waiting time through:

$$\pi^\star_{\gamma,\nu}(d) = (\gamma + \nu - d)^+. \tag{12}$$

Plugging the optimum policy into the Lagrange function (11), we have:

$$\begin{aligned}&\inf_\pi \mathcal{L}(\gamma,\nu,\pi) \\ =&\mathbb{E}\left[\frac{1}{2}\max\{(\gamma+\nu), D\}^2 - \gamma\max\{\gamma+\nu, D\}\right] \\ &+ \nu\left(\frac{1}{f_{\mathsf{max}}} - \mathbb{E}\left[\max\{(\gamma+\nu), D\}\right]\right).\end{aligned} \tag{13}$$

Let $\nu^\star := \arg\sup_{\nu \geq 0} \inf_{\pi \in \Pi_{\mathsf{SD}}} \mathcal{L}(\gamma^\star, \nu, \pi)$ be the dual optimizer that resolves the Lagrange function when $\gamma = \gamma^\star$. Notice that when $\pi^\star = \pi^\star_{\gamma^\star,\nu^\star}$ is used,

$$\begin{aligned}&\theta_{\mathsf{opt}} \\ =&\mathbb{E}\left[\frac{1}{2}\max\{(\gamma^\star+\nu^\star), D\}^2 - \gamma^\star\max\{(\gamma^\star+\nu^\star), D\}\right] \\ =&0.\end{aligned} \tag{14}$$

We then have the necessary condition on $\gamma^\star$:

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{2}\max\{(\gamma^\star+\nu^\star), D\}^2 - \gamma^\star\max\{(\gamma^\star+\nu^\star), D\}\right] \\ =&0.\end{aligned} \tag{15}$$

The following lemma characterizes the upper and lower bound of $\gamma^\star$, the proof will be provided in Appendix A:

Lemma 1: The optimum ratio $\gamma^\star$ can be upper and lower bounded by:

$$\gamma_{\mathsf{lb}} \leq \gamma^\star \leq \gamma_{\mathsf{ub}},$$

where

$$\begin{aligned}\gamma_{\mathsf{lb}} &:= \frac{1}{2}\overline{D}_{\mathsf{lb}}, \\ \gamma_{\mathsf{ub}} &:= \frac{\frac{1}{2}M_{\mathsf{ub}} + \overline{D}_{\mathsf{ub}}\frac{1}{f_{\mathsf{max}}} + \frac{1}{2}\frac{1}{f^2_{\mathsf{max}}}}{\overline{D}_{\mathsf{lb}} + \frac{1}{f_{\mathsf{max}}}}.\end{aligned}$$

2) An Online Learning Algorithm π_{online} Through the Robbins-Monro Algorithm: When the delay statistics $\mathbb{P}_D$ is known, $(\gamma^\star + \nu^\star)$ can be computed directly using a bi-section method [2]. When $\mathbb{P}_D$ is unknown, such computation is impossible because equation (15) is unknown. As an alternative, we approximate $\gamma^\star$ and $\nu^\star$ respectively. To meet the frequency constraint, we use sequence $\{U_k\}$ to track the sampling frequency constraint violation up to time S_k. Notice that the use of dual optimizer ν is to guarantee the sampling frequency constraint is satisfied, we use $\nu_k = \frac{1}{V}U_k$ as the dual optimizer in cycle k, where $V > 0$ is fixed as a constant. Then to find the root $\gamma^\star$ of equation (15) assuming that $\nu^\star = \nu_k$ is the dual optimizer, we use a sequence $\{\gamma_k\}$ to approximate $\gamma^\star$ in cycle k using the Robbins-Monro algorithm [29]. We start by initializing $\gamma_1 \in \text{Uni}\left([\gamma_{\mathsf{lb}}, \gamma_{\mathsf{ub}}]\right)$. The algorithm operates in cycle k as follows:

- After the transmission delay D_k of the k-th update packet is observed, we choose a waiting time W_k based on the current estimation γ_k and violation U_k:
$$W_k = \left(\gamma_k + \frac{1}{V}U_k - D_k\right)^+, \tag{16a}$$
where $V > 0$ is fixed as a constant. We then wait for W_k to take the next sample and then compute the cycle length $L_k = D_k + W_k$ as well as reward $Q_k = \frac{1}{2}(D_k + W_k)^2$.
- We then update γ_k via the Robbins-Monro algorithm [29] as follows:
$$\gamma_{k+1} = \left[\gamma_k + \eta_k\left(Q_k - \gamma_k L_k\right)\right]_{\gamma_{\mathsf{lb}}}^{\gamma_{\mathsf{ub}}}, \tag{16b}$$
where $[\gamma]_a^b = \min\{b, \max\{\gamma, a\}\}$ and $\{\eta_k\}$ is a set of diminishing step sizes that is selected to be:
$$\eta_k = \begin{cases} \frac{1}{2\overline{D}_{\mathsf{lb}}}, & k = 1; \\ \frac{1}{(k+2)\overline{D}_{\mathsf{lb}}}, & k \geq 2. \end{cases} \tag{16c}$$
- To guarantee that the sampling frequency constraint is not violated, we update the violation U_k up to the end of cycle k using:
$$U_{k+1} = \left(U_k + \left(\frac{1}{f_{\mathsf{max}}} - L_k\right)\right)^+. \tag{16d}$$

C. Theoretic Analysis

The evolution of the time average AoI optimality gap as a function of time t is hard to analyze in general. As an alternative, define ratio
$$\tilde{A}_K := \frac{\mathbb{E}\left[\int_{t=0}^{S_{K+1}} A(t)\mathrm{d}t\right]}{\mathbb{E}[S_{K+1}]}. \tag{17}$$

This metric is reasonable in the sense that $\tilde{A}_K$ is the ratio between the expected cumulative AoI up to the K-th cycle and the running length up to cycle K. Let π_K be the waiting time specification rule in cycle K. According to equation (16a), function $\pi_K(d) = (\gamma_K + \frac{1}{V}U_K - d)^+$. We measure the performance of the proposed algorithm via the convergence rate of difference $\tilde{A}_K - \overline{A}_{\pi^\star}$ and the expected average AoI difference between using policy π_K and $\pi^\star$, i.e., $\overline{A}_{\pi_K} - \overline{A}_{\pi^\star}$. The main results are as follows:

Theorem 2: When there is no transmission constraint, i.e., $f_{\mathsf{max}} = \infty$ and the transmission delay $D < B < \infty$ is upper bounded by B, by using the proposed online sampling algorithm π_{online}, the threshold $\{\gamma_k\}$ converges to the optimum threshold $\gamma^\star$ with probability 1, i.e.,
$$\lim_{K\to\infty} \gamma_K \overset{\text{a.s.}}{=} \gamma^\star. \tag{18a}$$

As a result, the average AoI of the proposed policy converges to the minimum $\overline{A}_{\pi^\star}$ with probability 1, i.e.,
$$\lim_{K\to\infty} \frac{\int_0^{S_{K+1}} A(t)\mathrm{d}t}{S_{K+1}} \overset{\text{a.s.}}{=} \overline{A}_{\pi^\star}, \tag{18b}$$
Proof for Theorem 2 is provided in Appendix C.

The next theorem characterizes the convergence rate of the proposed algorithm, whose proof is provided in Appendix B:

Theorem 3: [4] Up to frame K, the difference $\mathbb{E}[(\gamma_K - \gamma^\star)^2]$ can be bounded by:
$$\mathbb{E}[(\gamma_K - \gamma^\star)^2] \leq \frac{1}{K}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}. \tag{19a}$$

The difference between the expected time-averaged AoI by using policy π_K and $\pi^\star$ can be upper bounded by:
$$\overline{A}_{\pi_K} - \overline{A}_{\pi^\star} \leq \frac{L_{\mathsf{ub}}^4}{\overline{D}\,\overline{D}_{\mathsf{lb}}^2}\frac{1}{K} = \mathcal{O}\left(\frac{1}{K}\right). \tag{19b}$$
and the difference $\tilde{A}_K - \overline{A}_{\pi^\star}$ can be upper bounded by:
$$\tilde{A}_K - \overline{A}_{\pi^\star} \leq \frac{L_{\mathsf{ub}}^4}{\overline{D}\,\overline{D}_{\mathsf{lb}}^2} \times \frac{1 + \ln K}{K} = \mathcal{O}\left(\frac{\ln K}{K}\right), \tag{19c}$$
where $L_{\mathsf{ub}} = B + \gamma_{\mathsf{ub}}$.

Remark 1: When there is no sampling constraint, the proposed online algorithm learns the optimum policy adaptively, since both $\overline{A}_{\pi_K} - \overline{A}_{\pi^\star}$ and $\tilde{A}_{\pi_K} - \overline{A}_{\pi^\star}$ goes to 0 as K goes to infinity.

Remark 2: As is shown in equation (19a)-(19c), if the estimated average transmission lower bound $\overline{D}_{\mathsf{lb}}$ is closer to $\overline{D}$ and the upper bound L_{ub} is closer to $\overline{L}^\star$, the upper bound of both the estimation error $\mathbb{E}[(\gamma_K - \gamma^\star)^2]$ and the average AoI difference $\overline{A}_K - \overline{A}_{\pi^\star}$ are be smaller. This implies a good estimation on the upper and lower bound of $\overline{D}$ help minimize the average AoI.

Theorem 4: Let $\pi_{\mathbb{P}}^\star$ denote the AoI minimum sampling policy when the delay distribution is $\mathbb{P}$ and let $\gamma_{\mathbb{P}}^\star$ be the optimum updating threshold. At the end of cycle k, let $\hat{\gamma} : \mathbb{R}^k \mapsto \mathbb{R}^+$ be an estimator of ratio $\gamma_{\mathbb{P}}^\star$ using historical transmission delays $\mathcal{H}_k$. The minimax estimation error of $\gamma_{\mathbb{P}}^\star$ satisfies:
$$\min_{\hat{\gamma}} \max_{\mathbb{P}} \mathbb{E}\left[(\hat{\gamma}(\mathcal{H}_k) - \gamma_{\mathbb{P}}^\star)^2\right] \geq \Omega(1/k). \tag{20}$$

For any δ satisfies $0 < \delta < \left(\sqrt[3]{\frac{1}{2} + \sqrt{\frac{5}{4}}} + \sqrt[3]{\frac{1}{2} - \sqrt{\frac{5}{4}}}\right)/2$, let $\mathcal{P}_w(\delta)$ be the set of delay distributions that: (i) is absolutely continuous and upper bounded by B; (ii) when delay $D \sim \mathbb{P}$, by using the AoI optimum policy $\pi_{\mathbb{P}}^\star$, the probability of waiting to take the next sample is larger than δ, i.e., $p_w(\mathbb{P}) := \mathbb{E}_{D\sim\mathbb{P}}\left[\Pr(D \geq \gamma_{\mathbb{P}}^\star)\right] \geq \delta$. Then the time average AoI using any causal sampling algorithm π has the following lower bound:
$$\inf_{\pi\in\Pi} \sup_{\mathbb{P}\in\mathcal{P}_w(\delta)} \left(\frac{\mathbb{E}\left[\int_0^{S_{K+1}} A(t)\mathrm{d}t\right]}{\mathbb{E}[S_{K+1}]} - \overline{A}_{\pi_{\mathbb{P}}^\star}\right) \geq \delta \cdot \Omega\left(\frac{\ln K}{K}\right). \tag{21}$$

[4] By selecting proper stepsizes, the results still holds if the upper and lower bound on $\gamma^\star$ is unknown [30].

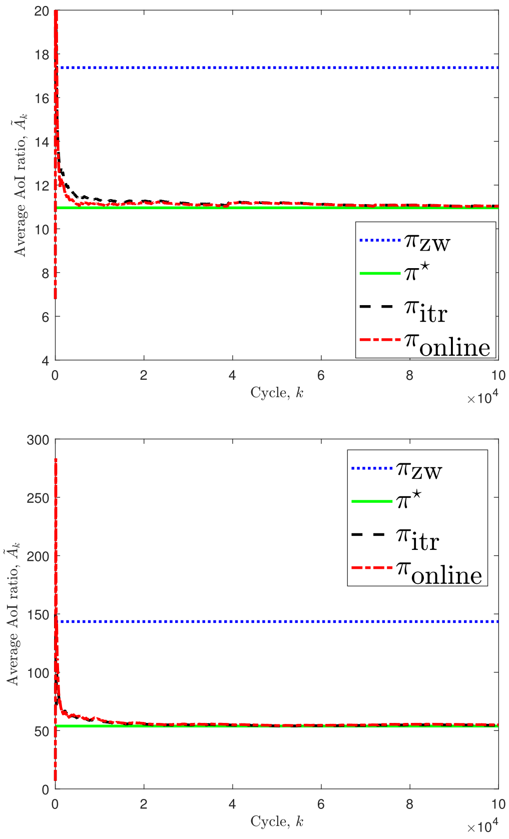

Fig. 3. The average AoI ratio evolution as a function of cycle k. Left $\mathsf{lognormal}(1, 1.3)$; Right $\mathsf{Weibul}(1, 0.3)$.

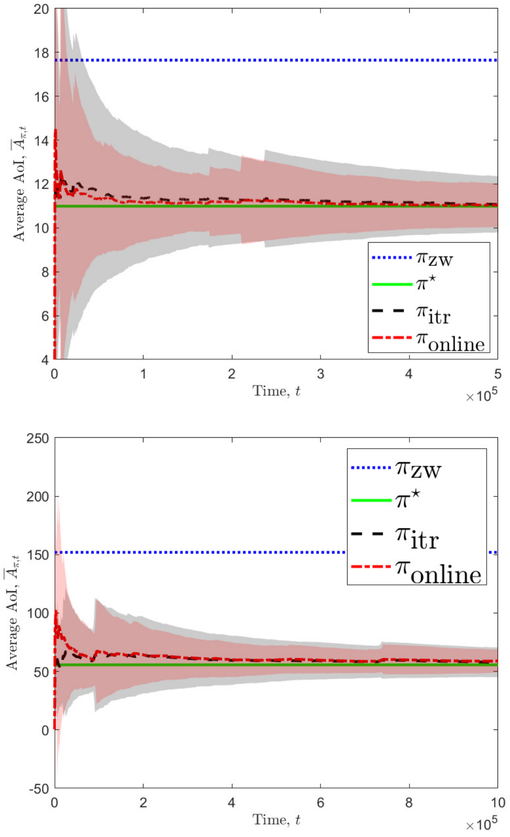

Fig. 4. The average time AoI evolution. Left $\mathsf{lognormal}(1, 1.3)$; Right $\mathsf{Weibul}(1, 0.3)$.

The proof is provided in Appendix F.

Remark 3: The order of the convergence rate of the $\mathbb{E}[(\gamma_k - \gamma^\star)^2]$ and $\mathbb{E}[\tilde{A}_k - \overline{A}_{\pi^\star}]$ (Theorem 3) match the converse bounds in Theorem 4. Therefore, the proposed algorithm is minimax order optimal.

Next, we analyze the sampling frequency violation behaviour of the proposed online policy. We have the following assumptions:

Assumption 2: Problem 1 can be strictly feasible. There exists $\epsilon > 0$ and a $\pi_\epsilon \in \Pi_{\mathsf{SD}}$ so that, by using policy π_ϵ, we have the following inequality,

$$\mathbb{E}\left[D + \pi(D)\right] \geq \frac{1}{f_{\mathsf{max}}} + \epsilon. \tag{22}$$

Under Assumption 2, we have the following result:

Theorem 5: The sampling constraint can be satisfied in the sense that:

$$\liminf_{K\to\infty} \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}(W_k + D_k)\right] \geq \frac{1}{f_{\mathsf{max}}}. \tag{23}$$

The proof is provided in Appendix I.

IV. Simulation Results

We validate the performance of the proposed algorithms via numerical simulations. We consider two sets of heavy tailed distribution that characterize the heavy traffic characteristics:

(a) $\mathsf{lognormal}(\mu, \sigma)$: log-normal distribution parameterized by μ and σ, i.e., the density function of the transmission delay distribution is $p(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$.

(b) $\mathsf{Weilbur}(a, b)$: Weilbur distribution parameterized by scale parameter a and shape parameter b, i.e., the density function $p(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1}\exp\left(-\left(\frac{x}{a}\right)^b\right)$.

A. Updating Without a Sampling Frequency Constraint

We first verify the asymptotic performance of π_{online} when there is no sampling frequency constraint, i.e., $f_{\mathsf{max}} = \infty$. We study and compare the following three strategies: (1) zero-wait policy that specifies $\pi_{\mathsf{zw}}(d) = 0, \forall d$; (2) the optimum policy $\pi^\star$ computed by [2]; (3) the iterative threshold computation method π_{itr} proposed by [25]. We compute the empirical mean and second-order moment of the first 100 transmission delays, i.e., $\hat{D} = \frac{1}{100}\sum_{k=1}^{100} D_k$, $\hat{M} = \frac{1}{100}\sum_{k=1}^{100} D_k^2$. We then set $D_{\mathsf{lb}} = \hat{D}/10, D_{\mathsf{ub}} = 10\hat{D}\ M_{\mathsf{lb}} = \hat{M}/10, M_{\mathsf{ub}} = 10\hat{M}$. Simulations are carried out when the transmission delay follows the log-normal distribution with parameters $\mu = 1$ and $\sigma = 1.3$. We plotted the AoI ratio up to cycle k, i.e., $\tilde{A}_k = \frac{\mathbb{E}\left[\int_0^{S_{k+1}} A(t)\mathrm{d}t\right]}{\mathbb{E}[S_{K+1}]}$ in Fig. 3. The mean of the time average AoI $\overline{A}_{\pi,t} = \frac{1}{t}\int_{t'=0}^{t} A(t')\mathrm{d}t'$ as well as its confidence interval are illustrated in Fig. 4. All the expectations are computed by taking the average of 100 runs. According to Fig. 3, the AoI ratio $\tilde{A}_k$ converges to the optimum AoI obtained by the

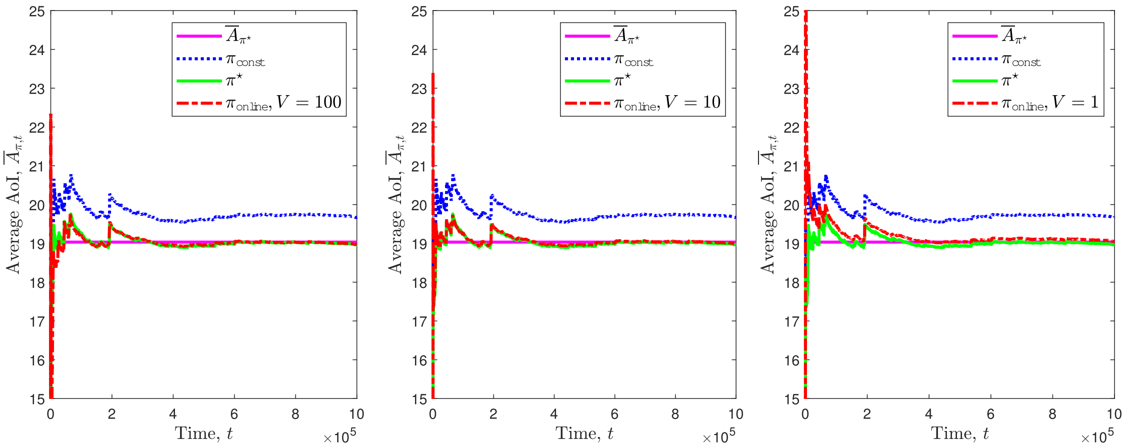

Fig. 5. The average AoI ratio evolution of a single sample path under sampling constraint.

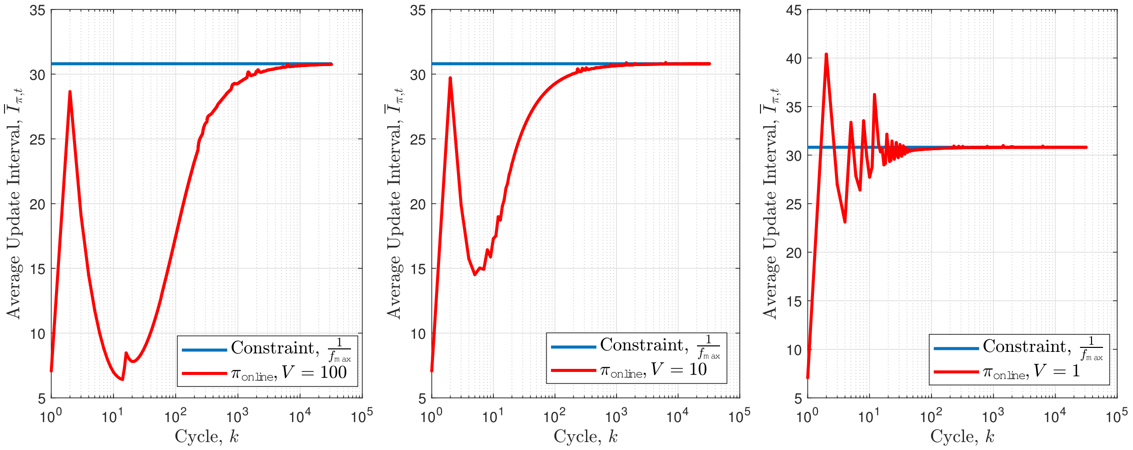

Fig. 6. The average sampling interval of a single sample path using different V.

optimum policy $\pi^\star$, which has been proved theoretically in Theorem 3. Moreover, when the proposed online learning policy π_{online} is used, the optimality gap between $\overline{A}_{\pi,t}$ AoI and the minimum AoI $\overline{A}_{\pi^\star}$ diminishes when time t goes to infinity. Compared with policy π_{itr}, the average AoI ratio of our proposed algorithm converges faster to $\overline{A}_{\pi^\star}$ and the variance is smaller.

B. Updating Under a Sampling Frequency Constraint

Next we study the performance of the proposed algorithm when the sampling constraint exists. Since the zero-wait sampling policy and the iterative threshold computing policy [25] may not satisfy the sampling frequency constraint, we compare the proposed algorithm with (1) a constant wait policy π_{const} that specifies waiting time by $\pi_{\text{const}}(d) = \frac{1}{f_{\text{max}}} - \overline{D}, \forall d$; (2) the optimum policy $\pi^\star$ computed by [2]. Simulations are carried out when the transmission delay follows the log-normal distribution with parameter $\mu = 1$, $\sigma = 1.5$, and the sampling frequency constraint is selected to be $f_{\text{max}} = \frac{1}{10\overline{D}}$. We plot the average AoI performance of a single sample path in Fig. 5 and the corresponding average sampling interval $\overline{I}_{\pi,K} \triangleq \frac{S_{K+1}}{K}$ in Fig. 6. From Fig. 5, it can be observed that the constant wait policy incurs a larger AoI, which is harmful to the data freshness performance. As expected, the average AoI of the proposed online algorithm converges to the average AoI of the optimum policy $\pi^\star$ when time t goes to infinity. Moreover, when time t increases, the average sampling interval converges to $\frac{1}{f_{\text{max}}}$, which means the sampling frequency is not violated. Similar to the queueing length-utility trade-off in network utility maximization [31], we found that choosing a smaller V (i.e., $V = 1$ in Fig. 6) guarantees that the sampling frequency constraint can be satisfied at a earlier stage, while choosing a larger V (i.e., $V = 100$ or $V = 10$ in Fig. 5 shows that the average AoI converges to the minimum AoI faster.

C. Addressing Practical Issues in Communication Networks–Timeout

Preemption, i.e., stop the previous transmission and restart a new on when the transmission delay is larger than a threshold can effectively minimize the average AoI. As is revealed by [27, Lemma 1], for pre-emption strategies with threshold τ, i.e., take a new sample and transmit it when the previous delay is larger than τ, the optimum sampling strategy $\pi_{\text{pre}}^{\tau,\star}$ still has a threshold structure. Let n_k be the number of retransmissions before the ACK of the $(k-1)$-th received sample and let D_k be the transmission delay of the $(k-1)$-th received sample,

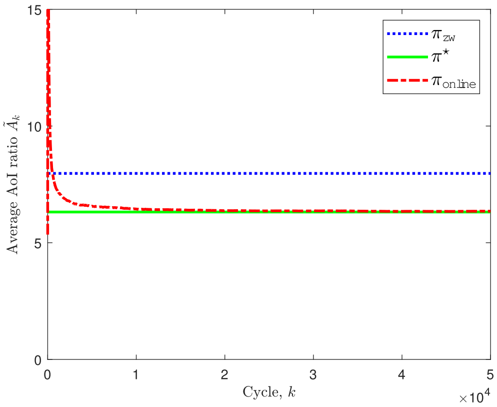

Fig. 7. The average AoI performance with time-out.

after the ACK of the $(k-1)$-th sample is received, policy $\pi_{\mathsf{pre}}^{\tau,\star}$ selects waiting time W_k as follows:

$$W_k = (\gamma^\star_{\mathsf{pre}} + \nu^\star_{\mathsf{pre}} - \tilde{D}_k)^+, \tag{24}$$

where $\tilde{D}_k := n_k\tau + D_k$ and the coefficient $\gamma^\star_{\mathsf{pre}} = \overline{A}_{\pi_{\mathsf{pre}}^{\tau,\star}} - \mathbb{E}[D|D \leq \tau]$ is defined similar to $\gamma^\star$, $\nu^\star_{\mathsf{pre}}$ is the dual optimizer for satisfying the sampling frequency constraint. For threshold policies with transmission preemption, the length of frame k now becomes L_k and the reward becomes $Q_k = \frac{1}{2}L_k^2$. Plugging the computation of L_k and Q_k back into algorithm (16a)-(16d) yields the online algorithm with transmission preemption.

In Fig. 7, we plotted the average AoI of different algorithms when a timeout threshold of $\tau = 10$ is used. The transmission delay follows $\mathsf{lognormal}(1, 1.3)$. From Fig. 7, the average AoI of our proposed online learning algorithm achieves a smaller AoI compared with the zero-wait policy, and approaches the optimum when the number of samples approaches infinity.

V. Conclusion

In this paper, we considered a sensor sampling and transmitting status updates to the receiver over a channel with random delay. We addressed the problem of minimizing the expected time average AoI under a sampling frequency constraint when the delay distribution is unknown. We reformulated the AoI minimization problem into a renewal-reward process optimization, and we propose an online sampling strategy based on the Robbins-Monro algorithm. We proved that the proposed algorithm can learn the optimum sampling policy almost surely when the number of samples K goes to infinity, and the average sampling frequency constraint can be satisfied. We prove that the convergence rate of the proposed algorithm is minimax optimum under certain conditions. Simulation results validate the adaptive performance of the proposed algorithm. Interesting extensions with piece-wise stationary delay distribution will be our future work.

Appendix A
Proof of Lemma 1

Proof: The lower bound of $\gamma^\star$ can be computed as follows:

$$\begin{aligned}\gamma^\star &= \frac{\mathbb{E}\left[\frac{1}{2}(D+\pi^\star(D))^2\right]}{\mathbb{E}[D+\pi^\star(D)]}\\ &\overset{(a)}{\geq} \frac{1}{2}\frac{\mathbb{E}[D+\pi^\star(D)]^2}{\mathbb{E}[D+\pi^\star(D)]}\\ &= \frac{1}{2}\mathbb{E}[D+\pi^\star(D)]\\ &\overset{(b)}{\geq} \frac{1}{2}\mathbb{E}[D] \overset{(c)}{\geq} \frac{1}{2}\overline{D}_{\mathsf{lb}},\end{aligned} \tag{25}$$

where inequality (a) is obtained by Jensen's inequality $\mathbb{E}\left[(D+\pi^\star(D))^2\right] \geq \mathbb{E}[(D+\pi^\star(D))]^2$; inequality (b) is because $0 \leq \pi(D) \leq W_{\mathsf{ub}}$ and the non-negativity of D; inequality (c) obtained due to Assumption 1.

To establish the upper bound of $\gamma^\star$, we consider the constant wait policy π_{const}, namely the waiting interval is fixed as a constant $W_k \equiv \frac{1}{f_{\mathsf{max}}}$ for any cycle k. According to (6a), the expected average AoI of policy π_{const} can be computed by:

$$\begin{aligned}\overline{A}_{\pi_{\mathsf{const}}} &= \frac{\mathbb{E}\left[\frac{1}{2}(D+\pi_{\mathsf{const}}(D))^2\right]}{\mathbb{E}\left[D+\pi_{\mathsf{const}}(D)\right]} + \overline{D}\\ &\leq \frac{\frac{1}{2}M_{\mathsf{ub}} + \overline{D}\frac{1}{f_{\mathsf{max}}} + \frac{1}{2}\frac{1}{f_{\mathsf{max}}^2}}{\overline{D}_{\mathsf{lb}} + \frac{1}{f_{\mathsf{max}}}} + \overline{D}.\end{aligned} \tag{26}$$

Notice that policy π_{const} may not be the AoI optimum strategy, i.e., $\overline{A}_{\pi^\star} \leq \overline{A}_{\pi_{\mathsf{const}}}$. Recall that the optimum ratio is computed by $\gamma^\star = \overline{A}_{\pi^\star} - \overline{D}$, we have:

$$\gamma^\star \leq \overline{A}_{\pi_{\mathsf{const}}} - \overline{D} \leq \frac{\frac{1}{2}M_{\mathsf{ub}} + \overline{D}_{\mathsf{ub}}\frac{1}{f_{\mathsf{max}}} + \frac{1}{2}\frac{1}{f_{\mathsf{max}}^2}}{\overline{D}_{\mathsf{lb}} + \frac{1}{f_{\mathsf{max}}}} =: \gamma_{\mathsf{ub}}. \tag{27}$$

■

Appendix B
Proof of Theorem 3

Proof: First, recall that the ratio γ_k in any cycle k is upper bounded by γ_{ub}, since the transmission delay is bounded $D \leq B$, the length L_k and reward Q_k in cycle k can be upper bounded by:

$$\begin{aligned}L &\leq D + (\gamma - D)^+ \leq B + \gamma_{\mathsf{ub}} =: L_{\mathsf{ub}},\\ Q &= \frac{1}{2}L^2 \leq L_{\mathsf{ub}}^2.\end{aligned} \tag{28}$$

Let $\overline{L}^\star := \mathbb{E}[D+\pi^\star(D)]$ and $\overline{Q}^\star := \mathbb{E}[\frac{1}{2}(D+\pi^\star(D))^2]$ be the expected average cycle length and the expected average reward if the optimum policy $\pi^\star$ is used. We will first provide the following lemmas:

Lemma 2: The expected cycle length $\mathbb{E}[L_k|\gamma_k]$ and the expected reward $\mathbb{E}[Q_k|\gamma_k]$ received in cycle k satisfies:

$$\mathbb{E}\left[Q_k - \gamma_k L_k|\gamma_k\right] \leq (\gamma^\star - \gamma_k)\overline{L}^\star, \tag{29a}$$

$$\mathbb{E}\left[Q_k - \gamma^\star L_k|\gamma_k\right] \leq -(\gamma^\star - \gamma_k)\left(\mathbb{E}[L_k|\gamma_k] - \overline{L}^\star\right). \tag{29b}$$

Lemma 3: Recall from equation (4), the cumulative AoI in cycle k is $X_k = Q_k + L_{k-1}D_k$. The cumulative AoI up to the end of cycle K, i.e., $\mathbb{E}\left[\int_0^{S_{K+1}} A(t)\mathrm{d}t\right] = \mathbb{E}\left[\sum_{k=1}^K X_k\right]$, satisfies the following inequality:

$$\mathbb{E}\left[\sum_{k=1}^K (X_k - (\gamma^\star + \overline{D})L_k)\right] \leq \mathbb{E}\left[\sum_{k=1}^K (\gamma^\star - \gamma_k)^2\right]. \tag{30}$$

Proofs for Lemma 2 and 3 are provided in Appendix D and E. Through (30), the average cost deviation can be upper bounded by:

$$\begin{aligned}
&\tilde{A}_K - \overline{A}_{\pi^\star}\\
=&\frac{\mathbb{E}\left[\sum_{k=1}^K X_k\right]}{\mathbb{E}\left[\sum_{k=1}^K L_k\right]} - (\gamma^\star + \overline{D})\\
=&\frac{\mathbb{E}\left[\sum_{k=1}^K (X_k - (\gamma^\star + \overline{D})L_k)\right]}{\mathbb{E}\left[\sum_{k=1}^K L_k\right]}\\
\le&\frac{\mathbb{E}\left[\sum_{k=1}^K (\gamma^\star - \gamma_k)^2\right]}{\mathbb{E}\left[\sum_{k=1}^K L_k\right]}.
\end{aligned} \tag{31}$$

We then prove inequalities in Theorem 3 as follows:

A. *Proof of* (19a)

For simplicity, denote

$$z_{k+1} := \gamma_k + \eta_k(Q_k - \gamma_k L_k). \tag{32}$$

Since $\gamma_{k+1} = [z_{k+1}]_{\gamma_{\mathsf{lb}}}^{\gamma_{\mathsf{ub}}}$ and $\gamma^\star \in [\gamma_{\mathsf{lb}}, \gamma_{\mathsf{ub}}]$, we can bound the stepsize deviation $(\gamma_{k+1} - \gamma^\star)^2$ using $(z_{k+1} - \gamma^\star)^2$:

$$(\gamma_{k+1} - \gamma^\star)^2 = ([z_{k+1}]_{\gamma_{\mathsf{lb}}}^{\gamma_{\mathsf{ub}}} - [\gamma^\star]_{\gamma_{\mathsf{lb}}}^{\gamma_{\mathsf{ub}}}) \le (z_{k+1} - \gamma^\star)^2. \tag{33}$$

We proceed to upper bound $(z_{k+1} - \gamma^\star)^2$ as follows:

$$\begin{aligned}
&\frac{1}{2}(z_{k+1} - \gamma^\star)^2\\
\overset{(a)}{=}&\frac{1}{2}\left(\gamma_k - \gamma^\star + \eta_k\left(Q_k - \gamma_k L_k\right)\right)^2\\
=&\frac{1}{2}(\gamma_k - \gamma^\star)^2 + \frac{1}{2}\eta_k^2\left(Q_k - \gamma_k L_k\right)^2\\
&+ \eta_k(\gamma_k - \gamma^\star)\left(Q_k - \gamma_k L_k\right)\\
\overset{(b)}{\le}&\frac{1}{2}(\gamma_k - \gamma^\star)^2 + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4 + \eta_k(\gamma_k - \gamma^\star)\left(Q_k - \gamma_k L_k\right),
\end{aligned} \tag{34}$$

where equality (a) is obtained from the definition of z_k in (32); inequality (b) is obtained because $Q_k = \frac{1}{2}L_k^2 \le L_{\mathsf{ub}}^2$ and $\gamma_k L_k \le L_{\mathsf{ub}}^2$. Then, taking the conditional expectation on both sides of (34), we have:

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[(z_{k+1} - \gamma^\star)^2 | \gamma_k\right]\\
\le&\frac{1}{2}(\gamma_k - \gamma^\star)^2 + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4\\
&+ \eta_k(\gamma_k - \gamma^\star)\mathbb{E}\left[Q_k - \gamma_k L_k | \gamma_k\right].
\end{aligned} \tag{35}$$

We then proceed to bound the last term in (35), i.e.,

$$(\gamma_k - \gamma^\star)\mathbb{E}\left[Q_k - \gamma_k L_k | \gamma_k\right] \tag{36}$$

- If the current $\gamma_k - \gamma^\star \ge 0$, by plugging (29a) into (36), we have:

$$\begin{aligned}
&(\gamma_k - \gamma^\star)\mathbb{E}[Q_k - \gamma_k L_k | \gamma_k]\\
\le& -(\gamma_k - \gamma^\star)^2 \overline{L}^\star \le -(\gamma_k - \gamma^\star)^2 \overline{D},
\end{aligned} \tag{37}$$

  where the last inequality is obtained because $\overline{L}^\star \ge \overline{D}$.

- If the current $\gamma_k - \gamma^\star \le 0$, we can upper the last term in inequality (35) as follows:

$$\begin{aligned}
&(\gamma_k - \gamma^\star)\mathbb{E}[Q_k - \gamma_k L_k | \gamma_k]\\
=&(\gamma_k - \gamma^\star)\mathbb{E}[Q_k - \gamma^\star L_k | \gamma_k]\\
&- (\gamma_k - \gamma^\star)^2 \mathbb{E}[L_k | \gamma_k]\\
\overset{(c)}{\le}&(\gamma_k - \gamma^\star)(\overline{Q}^\star - \gamma^\star \overline{L}^\star) - (\gamma_k - \gamma^\star)^2 \mathbb{E}[L_k | \gamma_k]\\
=& -(\gamma_k - \gamma^\star)^2 \mathbb{E}[L_k | \gamma_k]\\
\overset{(d)}{\le}& -(\gamma_k - \gamma^\star)^2 \overline{D},
\end{aligned} \tag{38}$$

  where inequality (c) is because $\mathbb{E}[Q_k - \gamma^\star L_k | \gamma_k] \ge \overline{Q}^\star - \gamma^\star \overline{L}^\star = 0$ and inequality (d) is because $\mathbb{E}[L_k | \gamma_k] \ge \overline{D}$.

Plugging (37) and (38) into (35), then taking the expectation with respect to γ_k yields:

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[(z_{k+1} - \gamma^\star)^2 | \gamma_k\right]\\
=&\left(\frac{1}{2} - \eta_k \overline{D}\right)\mathbb{E}\left[(\gamma_k - \gamma^\star)^2\right] + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4\\
\le&\left(\frac{1}{2} - \eta_k \overline{D}_{\mathsf{lb}}\right)\mathbb{E}\left[(\gamma_k - \gamma^\star)^2\right] + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4.
\end{aligned} \tag{39}$$

By taking the expectation of inequality (39) with respect to ratio γ_k and plugging in it into (33), we can upper bound $\mathbb{E}[(\gamma_{k+1} - \gamma^\star)^2]$ by:

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[(\gamma_{k+1} - \gamma^\star)^2\right]\\
\le&\frac{1}{2}\left(1 - 2\eta_k \overline{D}_{\mathsf{lb}}\right)\mathbb{E}\left[(\gamma_k - \gamma^\star)^2\right] + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4.
\end{aligned} \tag{40}$$

Next, by choosing stepsizes $\eta_1 = \frac{1}{2\overline{D}_{\mathsf{lb}}}$ and $\eta_k = \frac{1}{(k+2)\overline{D}_{\mathsf{lb}}}, \forall k > 1$, we can then show by induction that

$$\frac{1}{2}\mathbb{E}[(\gamma_k - \gamma^\star)^2] \le \frac{1}{2k}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}. \tag{41}$$

The proof is as follows:

- When $k = 2$, plugging the stepsize $\eta_1 = \frac{1}{2\overline{D}_{\mathsf{lb}}}$ into (40) yields:

$$\frac{1}{2}\mathbb{E}[(\gamma_2 - \gamma^\star)^2] \le \frac{1}{8}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2} \le \frac{1}{4}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}.$$

- When $k > 2$, assuming that $\frac{1}{2}\mathbb{E}[(\gamma_k - \gamma^\star)^2] \le \frac{1}{2k}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}$, recall that the stepsize $\eta_k = \frac{1}{(k+2)\overline{D}_{\mathsf{lb}}}$, we have

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[(\gamma_{k+1} - \gamma^\star)^2\right]\\
\le&\left(\frac{1}{2} - \eta_k \overline{D}_{\mathsf{lb}}\right)\mathbb{E}\left[(\gamma_k - \gamma^\star)^2\right] + \frac{1}{2}\eta_k^2 L_{\mathsf{ub}}^4\\
\le&\left(1 - \frac{2}{k+2}\right)\frac{1}{2k}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2} + \frac{1}{2}\frac{1}{(k+2)^2}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\\
=&\frac{1}{2}\left(\frac{1}{k+2} + \frac{1}{(k+2)^2}\right)\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\\
=&\frac{1}{2}\frac{k+3}{(k+2)^2}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\\
\overset{(f)}{\le}&\frac{1}{2}\frac{1}{(k+1)}\frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2},
\end{aligned} \tag{42}$$

where inequality (f) is obtained because $(k+1)(k+3) \leq (k+2)^2$.

B. Proof of (19c)

Summing up the inequality (19a) from cycle $k = 1$ to K we have:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=1}^{K}(\gamma^\star - \gamma_k)^2\right] \\
&\leq \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\left(\sum_{k=1}^{K}\frac{1}{k}\right) \\
&\overset{(a)}{\leq} \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\left(1 + \int_{k=1}^{K}\frac{1}{k}\mathrm{d}k\right) \\
&\overset{(b)}{=} \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\left(1 + \ln K\right),
\end{aligned} \tag{43}$$

where inequality (a) is obtained because $\frac{1}{k} \leq \int_{k'=k-1}^{k}\frac{1}{k'}\mathrm{d}k', \forall k > 1$ and equality (b) is obtained because $\int_a^b \frac{1}{x}\mathrm{d}x = \ln b - \ln a$. Plugging inequality (43) into (31) we have:

$$\begin{aligned}
\tilde{A}_K - \overline{A}_{\pi^\star} &= \frac{\mathbb{E}\left[\sum_{k=1}^{K}(\gamma^\star - \gamma_k)^2\right]}{\mathbb{E}\left[\sum_{k=1}^{K}L_k\right]} \\
&\leq \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}\left(1 + \ln K\right)\frac{1}{\mathbb{E}\left[\sum_{k=1}^{K}L_k\right]} \\
&\overset{(c)}{\leq} \frac{L_{\mathsf{ub}}^4}{\overline{D}\,\overline{D}_{\mathsf{lb}}^2} \times \frac{1 + \ln K}{K},
\end{aligned} \tag{44}$$

where inequality (c) is because $\mathbb{E}\left[\sum_{k=1}^{K}L_k\right] \geq \mathbb{E}\left[\sum_{k=1}^{K}D_k\right] = K\overline{D}$. This finishes the proof of (19c).

C. Proof of (19b)

Recall that the expected time average AoI using stationary policy π_K with ratio γ_K can be computed by

$$\overline{A}_{\pi_K} = \frac{\mathbb{E}[\frac{1}{2}((\gamma_K - D)^+ + D)^2]}{\mathbb{E}[(\gamma_K - D)^+ + D]} + \overline{D}.$$

Since $\pi^\star$ is the optimum stationary policy that achieves the smallest AoI, therefore for any stationary policy π_K, we have $\overline{A}_{\pi_K} \geq \overline{A}_{\pi^\star}$ and the optimality gap can be upper bounded by:

$$\begin{aligned}
&\overline{A}_{\pi_K} - \overline{A}_{\pi^\star} \\
&= \frac{\mathbb{E}\left[\frac{1}{2}((\gamma_K - D)^+ + D)^2\right]}{\mathbb{E}\left[(\gamma_K - D)^+ + D\right]} - \gamma^\star \\
&= \frac{\mathbb{E}\left[\frac{1}{2}((\gamma_K - D)^+ + D)^2 - \gamma_K((\gamma_K - D)^+ + D)\right]}{\mathbb{E}[(\gamma_K - D)^+ + D]} \\
&\quad + (\gamma_K - \gamma^\star) \\
&\overset{(d)}{=} \frac{\mathbb{E}\left[Q_K - \gamma_K L_K\right]}{\mathbb{E}[L_K]} + (\gamma_K - \gamma^\star) \\
&\overset{(e)}{\leq} \frac{(\gamma^\star - \gamma_K)\overline{L}^\star}{\mathbb{E}[(\gamma_K - D)^+ + D]} + (\gamma_K - \gamma^\star) \\
&= (\gamma_K - \gamma^\star) \\
&\quad \times \left(\frac{\mathbb{E}\left[(\gamma_K - D)^+ + D\right] - \mathbb{E}\left[(\gamma^\star - D)^+ + D\right]}{\mathbb{E}\left[(\gamma_K - D)^+ + D\right]}\right) \\
&= \frac{(\gamma_K - \gamma^\star)}{\mathbb{E}[(\gamma_K - D)^+ + D]}\mathbb{E}[(\gamma_K - D)^+ - (\gamma^\star - D)^+] \\
&\leq \frac{1}{\overline{D}}(\gamma_K - \gamma^\star)^2,
\end{aligned} \tag{45}$$

where equality (d) is by definition that $Q_K = \frac{1}{2}((\gamma_K - D_K)^+ + D_K)^2$, $L_K = (\gamma_K - D_K)^+ + D_K$ and the transmission delay D_K is i.i.d.; inequality (e) is obtained by taking the expectation with respect to γ_k of inequality $\mathbb{E}[Q_K - \gamma_K L_K|\gamma_k] \leq (\gamma^\star - \gamma_K)\overline{L}^\star$ from Lemma 2.

Plugging (34) into inequality (45), we can then complete the proof of Theorem 3:

$$\mathbb{E}\left[\overline{A}_{\pi_k} - \overline{A}_{\pi^\star}\right] \leq \frac{L_{\mathsf{ub}}^4}{\overline{D}\,\overline{D}_{\mathsf{lb}}^2}\frac{1}{k}.$$

■

Appendix C
Proof of Theorem 2

A. Proof of (18a)

The proof is divided into two steps, first we will show that $\{\gamma_k\}$ converges to the limit points of an Ordinary Differential Equation (ODE) with probability 1, and then we will show that the $\gamma^\star$ is the unique stationary point of the ODE.

Notice that when there is no sampling frequency constraint, $\nu_k \equiv 0$. For each $D \sim \mathbb{P}_D$, define function

$$g(\gamma; D) := \frac{1}{2}((\gamma - D)^+ + D)^2 - \gamma((\gamma - D)^+ + D), \tag{46}$$

and the expectation over $\mathbb{P}_D$ is denoted by:

$$\overline{g}(\gamma) := \mathbb{E}\left[g(\gamma; D)\right]. \tag{47}$$

With function g, the update rule in equation (16b) can be rewritten as follows:

$$\gamma_{k+1} = \left[\gamma_k + \eta_k Y_k\right]_{\gamma_{\mathsf{lb}}}^{\gamma_{\mathsf{ub}}}, \tag{48}$$

where $Y_k := g(\gamma_k; D_k)$.

Next, we will show that the update step-size $\{\eta_k\}$ and Y_k satisfy the following properties:

(1.1) Since γ_k is bounded, the second order moment of Y_k is bounded, i.e.,

$$\begin{aligned}
\mathbb{E}\left[|Y_k|^2\right] &= \mathbb{E}\left[(Q_k - \gamma_k L_k)^2\right] \\
&\leq \mathbb{E}\left[\left(\frac{1}{2}((\gamma_k - D_k)^+ + D_k)^2\right)^2\right] \\
&\quad + \mathbb{E}\left[\gamma_k^2\left((\gamma_k - D_k)^+ + D_k\right)^2\right] < \infty.
\end{aligned} \tag{49}$$

(1.2) Since D_k appears i.i.d. and γ_k is determined by historical $\{Y_i\}_{i \leq k-1}$, we have

$$\begin{aligned}
\mathbb{E}\left[Y_k\right] &= \mathbb{E}[Y_k|\gamma_1, \{Y_i\}_{i \leq k-1}] \\
&= \mathbb{E}\left[g(\gamma_k, D_k)|\gamma_k\right] = \overline{g}(\gamma_k).
\end{aligned} \tag{50}$$

(1.3) Function $\overline{g}(\cdot)$ is continuous.

Notice that the step-sizes $\{\eta_k\}$ are chosen such that $\sum_{k=1}^{\infty}\eta_k = \infty$ and $\sum_{k=1}^{\infty}\eta_k^2 < \infty$. The ratio in the k-th cycle on sample path ω is denoted by $\gamma_k(\omega)$, according to [32, p.126, Theorem 2.1], with probability 1, the limits $\gamma_k(\omega)$ are trajectories of the following ordinary differential equation:

$$\dot{\gamma} = \overline{g}(\gamma). \tag{51}$$

We will then show that $\gamma^\star$ is the unique stationary point of ODE (51). The derivative $\overline{g}(\gamma)$ can be computed by:

$$\overline{g}'(\gamma) = -\gamma \cdot \text{Pr}\left(D \leq \gamma\right), \tag{52}$$

Therefore, function $\overline{g}(\gamma)$ is monotonically non-increasing over $\mathbb{R}^+$, and is monotonically decreasing for γ that satisfies $\text{Pr}(\gamma > D) > 0$. Therefore, if zero-wait policy is not optimum, i.e., $\text{Pr}(\gamma^\star > 0) > 0$, then $\overline{g}(\gamma^\star) = 0$ and $\gamma^\star$ is the unique solution to the following equation

$$\overline{g}(\gamma) = 0. \tag{53}$$

We will then show $\gamma^\star$ is the unique stationary point of ODE (51) through Lyapunov stability analysis, where the Lyapunov function is denoted by $V(\gamma) := \frac{1}{2}(\gamma - \gamma^\star)$. Then we have:

$$\dot{V}(\gamma) = (\gamma - \gamma^\star)\overline{g}(\gamma). \tag{54}$$

According to the monotonic characteristic from (52), we have $\dot{V}(\gamma) < 0, \forall \gamma \neq \gamma^\star$ and the global stability of $\gamma^\star$ is verified from Lyapunov theorem. Since $\{\gamma_k\}$ almost surely to the limit point of the ODE (51) and $\gamma^\star$ is the unique stationary point of (51), we conclude that γ_k converges to $\gamma^\star$ almost surely.

B. Proof of (18b)

Let a_k be the average AoI up to frame k, which can be computed by:

$$a_k := \frac{\int_{t=0}^{S_{k+1}} A(t)\mathrm{d}t}{S_{k+1}} = \frac{\frac{1}{k}\int_{t=0}^{S_{k+1}} A(t)\mathrm{d}t}{\frac{1}{k}S_{k+1}}. \tag{55}$$

To show that sequence $\{a_k\}$ converges to $\overline{A}_{\pi^\star}$ almost surely, we will first show that the denominator in (55) is strictly positive with probability 1. Notice that $\frac{1}{k}S_{k+1}$ can be computed by:

$$\frac{1}{k}S_{k+1} = \frac{1}{k}\sum_{k'=1}^{k}(D_{k'} + W_{k'}) \geq \frac{1}{k}\sum_{k'=1}^{k} D_{k'}. \tag{56}$$

Since the transmission delays $\{D_{k'}\}$ are i.i.d., taking the limit on both sides inequality (56), the law of large number shows:

$$\liminf_{k\to\infty} \frac{1}{k}S_{k+1} \geq \liminf \frac{1}{k}\sum_{k'=1}^{k} D_{k'} \overset{\text{a.s.}}{=} \overline{D} > 0. \tag{57}$$

Inequality (57) implies, sequence $\frac{1}{k}S_{k+1}$ is strictly larger than a positive constant with probability 1. To prove sequence $\{a_k\} = \frac{\int_{t=0}^{S_{k+1}} A(t)\mathrm{d}t}{S_{k+1}}$ converges to $\overline{A}_{\pi^\star}$, it is equivalent to show that

$$\lim_{k\to\infty} \theta_k \overset{\text{a.s.}}{=} 0, \tag{58}$$

$$\text{where } \theta_k := \frac{1}{k}\int_{t=0}^{S_{k+1}} A(t)\mathrm{d}t - \overline{A}_{\pi^\star} \cdot \left(\frac{1}{k}S_{k+1}\right).$$

The proof will proceed in two steps: (i) we will show that with probability 1, $\{\theta_k\}$ converges to the limit points of an ODE; (ii) we will show that 0 is the unique stationary point of the ODE. The first step is to rewrite the evolution of $\{\theta_k\}$ into a recursive form. Recall that the cumulative AoI in frame k is $\int_{S_k}^{S_{k+1}} A(t)\mathrm{d}t = Q_k + L_{k-1}D_k$ and the optimum AoI $\overline{A}_{\pi^\star} = \gamma^\star + \overline{D}$, θ_k can be rewritten as follows:

$$\begin{aligned}
\theta_k =& \frac{1}{k}\sum_{k'=1}^{k}(Q_{k'} + L_{k'-1}D_{k'}) \\
& - \left(\gamma^\star + \overline{D}\right) \cdot \left(\frac{1}{k}\sum_{k'=1}^{k} L_{k'}\right) \\
=& \frac{1}{k}\left((k-1)\theta_{k-1} + Q_k + L_{k-1}D_k - (\gamma^\star + \overline{D})L_k\right) \\
=& \theta_{k-1} + \frac{1}{k}\underbrace{\left(Q_k - (\gamma^\star + \overline{D})L_k - \theta_{k-1} + L_{k-1}D_k\right)}_{=:Y_k} \\
=& \theta_{k-1} + \frac{1}{k}\left(\mathbb{E}[Y_k|\mathcal{H}_{k-1}] + (Y_k - \mathbb{E}[Y_k|\mathcal{H}_{k-1}])\right)
\end{aligned} \tag{59}$$

To further simply the evolution of θ_k, we make the following definitions on function $f(\theta, \gamma; d)$:

$$f(\theta, \gamma; d) := \frac{1}{2}\left((\gamma - d)^+ + d\right)^2 - \gamma \cdot \left((\gamma - d)^+ + d\right) - \theta. \tag{60}$$

Let $f(\theta, \gamma) := \mathbb{E}_D[f(\theta, \gamma; D)]$ be the expectation over D. Specifically, denote function $\overline{f}(\theta)$ as the value of $f(\theta, \gamma)$ when $(\gamma = \gamma^\star)$. By definition $\overline{f}(\theta)$ can be simplified as follows:

$$\begin{aligned}
\overline{f}(\theta) :=& f(\theta, \gamma^\star) \\
=& \mathbb{E}_D\Big[\frac{1}{2}\left((\gamma^\star - D)^+ + D\right)^2 \\
& - \gamma^\star \cdot \Big((\gamma^\star - D)^+ + D\Big)\Big] - \theta \\
\overset{(a)}{=}& -\theta,
\end{aligned} \tag{61}$$

where equality (a) is because $\mathbb{E}\left[\frac{1}{2}\left((\gamma^\star - D^+ + D)\right)^2\right] - \gamma^\star\mathbb{E}\left[(\gamma^\star - D)^+ + D\right] = 0$.

Then given historical transmission $\mathcal{H}_{k-1}$, the conditional expectation of Y_k can be computed by:

$$\begin{aligned}
& \mathbb{E}\left[Y_k|\mathcal{H}_{k-1}\right] \\
=& \mathbb{E}[f(\theta_{k-1}, \gamma_k; D)] - \overline{D}\mathbb{E}[L_k|\gamma_k] + L_{k-1}\overline{D} \\
=& f(\theta_{k-1}, \gamma_k) \\
& - \underbrace{\overline{D} \cdot \left(\mathbb{E}\left[(\gamma_k - D)^+ + D\right] - \mathbb{E}\left[(\gamma^\star - D)^+ + D\right]\right)}_{=:\beta_{k,1}} \\
& + \underbrace{\overline{D} \cdot \left(\left(\gamma_{k-1} - D_{k-1}\right)^+ + D_{k-1}\right) - \mathbb{E}\left[(\gamma_{k-1} - D)^+ + D\right]\right)}_{=:\beta_{k,2}}
\end{aligned} \tag{62}$$

Finally, denote $\delta M_k := Y_k - \mathbb{E}[Y_k|\mathcal{H}_{k-1}]$ and plugging equality (62) into equation (59), we have:

$$\theta_k = \theta_{k-1} + \frac{1}{k} \cdot \left(f(\theta_{k-1}, \gamma_k) + \delta M_k - \beta_{k,1} + \beta_{k,2})\right), \quad (63)$$

Denote $\epsilon_k := \frac{1}{k}$, which can be viewed as the step-size for updating θ_k. Term $\beta_{k,1}$ and $\beta_{k,2}$ can be viewed as two bias terms. Define $t_0 = 0$ and the cumulative step-sizes up to cycle k is denoted by $t_k = \sum_{i=0}^{k-1} \epsilon_i$. Therefore, $\ln k \leq t_k \leq 1 + \ln(k-1)$. For $t \geq 0$, let $m(t)$ be the unique value such that $t_{m(t)} \leq t < t_{m(t)+1}$. We have

$$m(t) = \lfloor \exp(t) \rfloor. \quad (64)$$

We then present the following properties about the recursive equation (63):

(2.1) Notice that in each frame k, Q_k, L_k are bounded. Therefore, θ_k is bounded and hence $\sup_k \mathbb{E}[|Y_k|]$ is bounded.

(2.2) Function $f(\theta, \gamma)$ is continuous in θ by definition.

(2.3) For each $\theta < \infty$, function $|f(\theta,\gamma)| \leq \mathbb{E}\left[\frac{1}{2}((\gamma - D)^+ + D)^2\right] + \gamma\mathbb{E}\left[(\gamma - D)^+ + D\right] < \infty$ is bounded. The difference between $f(\theta, \gamma)$ and $\overline{f}(\theta)$ can be computed by

$$\begin{aligned}&\left|f(\theta,\gamma) - \overline{f}(\theta)\right| = \left|\mathbb{E}\left[(\gamma - D)^+ - (\gamma^\star - D)^+\right]\right| \\ \leq &|\gamma - \gamma^\star|. \end{aligned} \quad (65)$$

Therefore, for each k we have:

$$\begin{aligned}
&\Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i (f(\theta,\gamma_i) - \overline{f}(\theta))\right| \geq \mu\right) \\
\leq &\frac{\mathbb{E}\left[\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i (f(\theta,\gamma_k) - \overline{f}(\theta))\right|\right]}{\mu} \\
\leq &\frac{1}{\mu}\mathbb{E}\left[\sum_{i=k}^{\infty} \epsilon_i \cdot \left|f(\theta,\gamma_i) - \overline{f}(\theta)\right|\right] \\
\overset{(a)}{\leq} &\frac{1}{\mu}\mathbb{E}\left[\sum_{i=k}^{\infty} \frac{1}{i^{3/4}} \cdot \left(\frac{1}{i^{1/4}} \cdot |\gamma_i - \gamma^\star|\right)\right] \\
\overset{(b)}{\leq} &\frac{1}{\mu}\sqrt{\left(\sum_{i=k}^{\infty} \left(i^{-3/4}\right)^2\right) \cdot \mathbb{E}\left[\sum_{i=k}^{\infty} \left(i^{-1/2} \cdot (\gamma_i - \gamma^\star)^2\right)\right]} \\
\overset{(c)}{\leq} &\frac{1}{\mu}\sqrt{\left(\sum_{i=k}^{\infty} i^{-3/2}\right) \cdot \left(\sum_{i=k}^{\infty} i^{-1/2} \cdot \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2} i^{-1}\right)} \\
\leq &\frac{2}{\mu} \cdot \frac{1}{\sqrt{k-1}} \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}.
\end{aligned} \quad (66)$$

where inequality (a) is because (65); inequality (b) is from Cauchy-Schwarz; inequality (c) is because (19a) from Theorem 3. Taking the limit on both sides of inequality (66), and recall $m(k) = \lfloor \exp(k) \rfloor$ from equation (64), we have:

$$\begin{aligned}
&\lim_{k\to\infty} \Pr\left(\sup_{j\geq m(k)}\left|\sum_{i=m(k)}^{j} \epsilon_i \cdot (g(\theta,\gamma_i) - \overline{g}(\theta))\right| \geq \mu\right) \\
\leq &\lim_{k\to\infty} \frac{2}{\mu} \cdot \frac{1}{\sqrt{\exp(k) - 1}} = 0.
\end{aligned} \quad (67)$$

(2.4) Given historical transmission $\mathcal{H}_{k-1}$, the difference δM_k only depends on D_k and has mean zero. Since γ_k is upper bounded in each frame and the delay D_k is second order bounded, the expectation Q_k, L_k are both upper bounded and the difference sequence δM_k is second order bounded. Therefore sequence $M_k := \sum_{k'=1}^{k} \epsilon_{k'} \delta M_{k'}$ is also a martingale sequence. According to [32, Chapter 5, Eq. (2.6)], for each $\mu > 0$, we have

$$\begin{aligned}
&\lim_{k\to\infty} \Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i \delta M_i\right| \geq \mu\right) \\
= &\lim_{k\to\infty} \Pr\left(\sup_{j\geq k} |M_j - M_k| \geq \mu\right) = 0.
\end{aligned} \quad (68)$$

(2.5) $\beta_{k,1}$ and $\beta_{k,2}$ can be viewed as two bias terms in the recursive form. Next we will show:

$$\lim_{k\to\infty} \Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i (\beta_{i,1} + \beta_{i,2})\right| \geq \mu\right) = 0. \quad (69)$$

The proof is as follows: through the union bound we have

$$\begin{aligned}
&\lim_{k\to\infty} \Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i (\beta_{i,1} + \beta_{i,2})\right| \geq \mu\right) \\
\leq &\lim_{k\to\infty} \Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i \beta_{i,1}\right| \geq \mu/2\right) \\
&+ \lim_{k\to\infty} \Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i \beta_{i,2}\right| \geq \mu/2\right).
\end{aligned} \quad (70)$$

For given k, we can upper bound the first term in inequality (70) as follows:

$$\begin{aligned}
&\Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i \beta_{i,1}\right| \geq \mu/2\right) \\
\overset{(d)}{\leq} &\frac{\mathbb{E}\left[\sup_{j\geq k}\left|\sum_{i=k}^{j} \epsilon_i \beta_{i,1}\right|\right]}{\mu/2} \\
\leq &\frac{2}{\mu}\mathbb{E}\left[\sum_{i=k}^{\infty} \frac{1}{i}|\beta_{i,1}|\right] \\
\overset{(e)}{\leq} &\frac{2}{\mu}\sqrt{\left(\sum_{i=k}^{\infty} \left(\frac{1}{i^{3/4}}\right)^2\right) \cdot \mathbb{E}\left[\sum_{i=k}^{\infty} \left(\frac{1}{i^{1/4}}\beta_{i,1}\right)^2\right]} \\
\overset{(f)}{\leq} &\frac{2}{\mu}\sqrt{\left(\sum_{i=k}^{\infty} i^{-3/2}\right) \cdot \mathbb{E}\left[\sum_{i=k}^{\infty} i^{-1/2}(\gamma_i - \gamma^\star)^2\right]} \\
\leq &\frac{2}{\mu}\sqrt{\left(\sum_{i=k}^{\infty} i^{-3/2}\right) \cdot \left(\sum_{i=k} i^{-3/2}\right) \frac{L_{\mathsf{ub}}^4}{\overline{D}_{\mathsf{lb}}^2}} \\
\leq &\frac{4L_{\mathsf{ub}}^2}{\overline{D}_{\mathsf{lb}}} \frac{1}{\sqrt{k}}.
\end{aligned} \quad (71)$$

where inequality (d) is from Markov inequality; inequality (e) is from Cauchy-Schwarz; inequality (f)

comes from (19a) in Theorem 3. Taking the limit with respect to k on both sides of inequality (71), we have:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\left|\sum_{i=m(k)}^{j}\epsilon_i\beta_{i,1}\right|\geq \mu/2\right)=0. \tag{72}$$

Notice that the second part $\beta_{k,2}$ is predicable given historical transmission $\mathcal{H}_{k-1}$. It is also a martingale sequence given $\mathcal{H}_{k-2}$. Therefore, $b_k := \sum_{k'=1}^{k}\epsilon_k\beta_{k,2}$ is also a martingale sequence. Through [32, Chapter 5, Eq. (2.6)] we can obtain:

$$\begin{aligned}&\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j}\epsilon_i\beta_{i,2}\right|\geq \mu/2\right)\\&=\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}|b_j-b_k|\geq \mu/2\right)=0.\end{aligned} \tag{73}$$

Plugging (71) and (73) into (70) verifies (69).

(2.6) Function f is uniformly bounded for $\theta \in \left[0, 2L_{\mathsf{ub}}^2\right], \gamma\in[\gamma_{\mathsf{lb}},\gamma_{\mathsf{ub}}]$.

(2.7) For each γ we have:

$$|f(\theta_1,\gamma)-f(\theta_2,\gamma)|=|\theta_1-\theta_2|, \tag{74}$$

and $\lim_{\theta\to\infty}|\theta|=0$.

(2.8) Sequence $\frac{1}{k}$ satisfies $\sum_{k'=1}^{\infty}\frac{1}{k'}=\infty$.

Therefore, according to [32, p.166, Theorem 1.1],[5] with probability 1, sequence θ_k converges to the limit point of the following ODE:

$$\dot{\theta}=\overline{f}(\theta)=-\theta. \tag{75}$$

Notice that $\theta=0$ is the unique stationary point of the ODE (75). Therefore,

$$\lim_{k\to\infty}\theta_k=\lim_{k\to\infty}\frac{1}{k}\left(\int_{t=0}^{S_{k+1}}A(t)\mathrm{d}t-(\gamma^\star+\overline{D})S_{k+1}\right)=0, \text{ w.p.1.} \tag{76}$$

Finally, plugging (76) into (58) implies:

$$\lim_{k\to\infty}\frac{\int_{t=0}^{S_{k+1}}A(t)\mathrm{d}t}{S_{k+1}}\overset{\text{a.s.}}{=}\gamma^\star+\overline{D}=\overline{A}_{\pi^\star}. \tag{77}$$

Appendix D
Proof of Lemma 2

Proof: Notice that in each cycle k, the waiting time W_k is chosen to minimize the objective function (10a), therefore we have:

$$\begin{aligned}&\mathbb{E}\left[Q_k-\gamma_k L_k|\gamma_k\right]\overset{(a)}{\leq}(\overline{Q}^\star-\gamma_k\overline{L}^\star)\\&=(\overline{Q}^\star-\gamma^\star\overline{L}^\star)+(\gamma^\star-\gamma_k)\overline{L}^\star\overset{(b)}{=}(\gamma^\star-\gamma_k)\overline{L}^\star,\end{aligned} \tag{78}$$

where equality (a) is because policy π_k used in cycle k minimizes the Lagrange function. Equality (b) is obtained because on the stationary point $\gamma^\star$ we have $\overline{Q}^\star=\gamma^\star\overline{L}^\star$. This verifies the first inequality in Lemma 2.

Then, adding $(\gamma_k-\gamma^\star)\mathbb{E}[L_k|\gamma_k]$ to both sides of (78) leads to:

$$\mathbb{E}\left[Q_k-\gamma^\star L_k|\gamma_k\right]\leq(\gamma_k-\gamma^\star)\mathbb{E}\left[L_k-\overline{L}^\star|\gamma_k\right]. \tag{79}$$

which verifies the second inequality. ■

Appendix E
Proof of Lemma 3

Proof: To find the upper bound of $\mathbb{E}\left[\sum_{k'=1}^{k}((Q_{k'}+L_{k'-1}D_{k'})-(\gamma^\star+\overline{D})L_{k'})\right]$, first we add $\mathbb{E}[L_{k-1}D_k|\gamma_k]$ on both sides on (29b) and obtain:

$$\begin{aligned}&\mathbb{E}[(Q_k+L_{k-1}D_k)-\gamma^\star L_k|\gamma_k]\\&\leq-(\gamma^\star-\gamma_k)\left(\mathbb{E}[L_k|\gamma_k]-\overline{L}^\star\right)+\mathbb{E}[L_{k-1}D_k|\gamma_k].\end{aligned} \tag{80}$$

Next, we can proceed to simplify (80) by:

$$\begin{aligned}&\mathbb{E}[(Q_k+L_{k-1}D_k)-\gamma^\star L_k|\gamma_k]\\&\overset{(a)}{\leq}-(\gamma^\star-\gamma_k)\left(\mathbb{E}[L_k|\gamma_k]-\overline{L}^\star\right)+L_{k-1}\overline{D}\\&\overset{(b)}{\leq}(\gamma^\star-\gamma_k)^2+L_{k-1}\overline{D},\end{aligned} \tag{81}$$

where inequality (a) is because D_k is independent of L_{k-1} and thus $\mathbb{E}[L_{k-1}D_k|\gamma_k]=\mathbb{E}[L_{k-1}]\overline{D}$; inequality (b) is because

$$\begin{aligned}\mathbb{E}[L_k-\overline{L}^\star|\gamma_k]&=\mathbb{E}\left[(\gamma_k-D)^+-(\gamma^\star-D)^+\right]\\&\leq|\gamma_k-\gamma^\star|.\end{aligned} \tag{82}$$

Summing up inequality (81) from cycle $k=1$ to K and take the expectation with respect to γ_k, we have:

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=1}^{K}((Q_k+L_{k-1}D_k)-\gamma^\star L_k)\right]\\&\leq\mathbb{E}\left[\sum_{k=1}^{K}(\gamma^\star-\gamma_k)^2\right]-\mathbb{E}\left[\sum_{k=1}^{K}L_k\right]\overline{D}.\end{aligned} \tag{83}$$

Deducting $\mathbb{E}\left[\sum_{k=1}^{K}L_k\right]\overline{D}+\mathbb{E}[L_K]\gamma^\star$ on both sides of inequality (83) yields:

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=1}^{K}((Q_k+L_{k-1}D_k)-(\gamma^\star+\overline{D})L_k)\right]\\&\leq\mathbb{E}\left[\sum_{k=1}^{K}(\gamma^\star-\gamma_k)^2\right]-\mathbb{E}[L_K]\overline{D}\leq\mathbb{E}\left[\sum_{k=1}^{K}(\gamma^\star-\gamma_k)^2\right].\end{aligned} \tag{84}$$

This completes the proof of Lemma 3. ■

Appendix F
Proof of Theorem 4

A. *Proof of Inequality* (20)

Proof: For each distribution $\mathbb{P}$, the optimum ratio $\gamma_{\mathbb{P}}^\star$ satisfies the following equation:

$$\frac{1}{2}\mathbb{E}\left[((\gamma_{\mathbb{P}}^\star-D)^++D)^2\right]-\gamma_{\mathbb{P}}^\star\mathbb{E}\left[(\gamma_{\mathbb{P}}^\star-D)^++D\right]=0. \tag{85}$$

The minimax estimation error bound on $\hat{\gamma}$ is established through the Le Cam's two point method [33], [34]. Let $\mathbb{P}_1$ and

[5] As mentioned on [32, p. 166, Eq. (1.10)], assumption (A1.6) in [32, p. 165] becomes: function g is uniformly bounded, [32, p. 166, Theorem 1.1] is still true.

$\mathbb{P}_2$ be two probability distributions and denote $\gamma_1 := \gamma^\star_{\mathbb{P}_1}$, $\gamma_2 := \gamma^\star_{\mathbb{P}_2}$ for simplicity. Through Le Cam's inequality, we have:

$$\inf_{\hat{\gamma}} \sup_{\mathbb{P}} \mathbb{E}[(\hat{\gamma}(\mathcal{H}_k) - \gamma^\star_{\mathbb{P}})^2] \geq (\gamma_1 - \gamma_2)^2 \cdot \mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}, \tag{86}$$

where $\mathbb{P} \wedge \mathbb{Q} = \int \min\{\mathrm{d}\mathbb{P}, \mathrm{d}\mathbb{Q}\}$.

To use the Le Cam's method, the first step is to find two distribution $\mathbb{P}_1, \mathbb{P}_2$ such that the difference $(\gamma_1 - \gamma_2)^2$ is large but $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$ can be lower bounded. We consider $\mathbb{P}_1 = \text{Uni}([0,1])$ be the uniform distribution. When $D \sim \mathbb{P}_1$, equation (85) can be simplified into:

$$-\frac{1}{6}\gamma_1^3 - \frac{1}{2}\gamma_1 + \frac{1}{6} = 0. \tag{87}$$

Since γ is a real number, according to the solution of cubic equation, we have:

$$\gamma_1 = \left(\sqrt[3]{\frac{1}{2} + \sqrt{\frac{5}{4}}} + \sqrt[3]{\frac{1}{2} - \sqrt{\frac{5}{4}}}\right). \tag{88}$$

Recall that $\mathbb{P}_1$ is a uniform distribution, therefore the probability of waiting by using the optimum policy $\pi^\star_{\mathbb{P}_1}$ is:

$$p_{\text{w, uni}} := \Pr\left(D \leq \gamma_1 | D \sim \mathbb{P}_1\right) = \gamma_1. \tag{89}$$

Distribution $\mathbb{P}_2$ is defined through the density function $p_2(x) = \frac{\mathsf{P}_2(\mathrm{d}x)}{\mathrm{d}x}$:

$$p_2(x) = \begin{cases} 1 - c\sqrt{1/k}, & 0 \leq x \leq \delta/2; \\ 1, & \delta/2 < x < 1 - \delta/2; \\ 1 + c\sqrt{1/k}, & 1 - \delta/2 \leq x \leq 1; \\ 0, & \text{otherwise}. \end{cases} \tag{90}$$

where $\delta = \min\{1/3, p_{\text{w, uni}}/2\}$ and $c < 1/2$ is fixed as a constant.

Lower bounding $(\gamma_2 - \gamma_1)^2$ is divided into two steps: first we will prove $\gamma_2 \geq \gamma_1$; then we will obtain the lower bound of γ_2 through Taylor expansion. For simplicity, let function $h_1(\cdot)$ and $h_2(\cdot)$ be:

$$h_1(\gamma) := \mathbb{E}_{D \sim \mathbb{P}_1}\left[\frac{1}{2}((\gamma - D)^+ + D)^2 - \gamma\left((\gamma - D)^+ + D\right)\right], \tag{91a}$$

$$h_2(\gamma) := \mathbb{E}_{D \sim \mathbb{P}_2}\left[\frac{1}{2}((\gamma - D)^+ + D)^2 - \gamma\left((\gamma - D)^+ + D\right)\right]. \tag{91b}$$

Then γ_1 and γ_2 satisfy $h_1(\gamma_1) = 0$ and $h_2(\gamma_2) = 0$.

Step 1 (Showing $\gamma_2 > \gamma_1$): The derivative of function $h_2(\gamma)$ can be computed by:

$$h_2(\gamma)' = -\mathbb{E}_{D \sim \mathbb{P}_2}\left[(\gamma - D)^+ + D\right] < 0. \tag{92}$$

Therefore, function $h_2(\gamma)$ is monotonically decreasing.

We will then show $h_2(\gamma_1) > 0$. Since $h_1(\gamma_1) = 0$, it is sufficient to show that $h_2(\gamma_1) > h_1(\gamma_1)$. The difference $h_2(\gamma) - h_1(\gamma)$ can be computed as follows:

$$\begin{aligned} &h_2(\gamma) - h_1(\gamma) \\ =& \mathbb{E}_{\mathbb{P}_2}\left[\frac{1}{2}\left((\gamma - D)^+ + D\right)^2 - \gamma\left((\gamma - D)^+ + D\right)\right] \\ &- \mathbb{E}_{\mathbb{P}_1}\left[\frac{1}{2}\left((\gamma - D)^+ + D\right)^2 - \gamma\left((\gamma - D)^+ + D\right)\right] \\ \overset{(a)}{=}& \int_{1-\delta/2}^{1} \frac{c}{\sqrt{k}}\left(\frac{1}{2}\max\{\gamma, x\}^2 - \gamma\max\{\gamma, x\}\right)\mathrm{d}x \\ &- \int_0^{\delta/2} \frac{c}{\sqrt{k}}\left(\frac{1}{2}\max\{\gamma, x\}^2 - \gamma\max\{\gamma, x\}\right)\mathrm{d}x \\ =& \int_0^{\delta/2} \frac{c}{\sqrt{k}}\Big(\frac{1}{2}\left(\max\{\gamma, x + (1-\delta)\}^2 - \max\{\gamma, x\}^2\right) \\ &- \gamma\left(\max\{\gamma, x + (1-\delta)\} - \max\{\gamma, x\}\right)\Big)\mathrm{d}x \\ =& \int_0^{\delta/2} \frac{c}{\sqrt{k}}\left(\frac{1}{2}\left(\max\{\gamma, x + (1-\delta)\} + \max\{\gamma, x\}\right) - \gamma\right) \\ &\times \left(\max\{\gamma, x + (1-\delta)\} - \max\{\gamma, x\}\right)\mathrm{d}x \\ \overset{(b)}{\geq}& 0, \end{aligned} \tag{93}$$

where equality (a) is because $(\gamma - D)^+ + D = \max\{\gamma, D\}$; inequality (b) is because $\max\{\gamma, x + (1-\delta)\} - \max\{\gamma, x\} \geq 0$ for $\delta < 1$, and $\max\{\gamma, x + (1-\delta)\} + \max\{\gamma, x\} \geq 2\gamma$. Therefore, $h_2(\gamma_1) \geq h_1(\gamma_1) = 0$. Since $h_2(\gamma_2) = 0$ and function $h_2(\cdot)$ is monotonically decreasing, we have $\gamma_2 \geq \gamma_1$.

Step 2 (Taylor Expansion to Lower Bound $h_2(\gamma_1)$): Through Taylor expansion, we have:

$$(\gamma_2 - \gamma_1) = \frac{h_2(\gamma_2) - h_2(\gamma_1)}{h_2'(\gamma)} = \frac{h_2(\gamma_1) - h_2(\gamma_2)}{-h_2'(\gamma)}, \tag{94}$$

where $\gamma \in [\gamma_1, \gamma_2]$. To lower bound $(\gamma_2 - \gamma_1)$, it is suffice to lower bound $h_2(\gamma_1) - h_2(\gamma_2)$ and upper bound $h_2'(\gamma)$. By Corollary 1, since $c \leq 1/2$ and $\delta < 1$, we can upper bound γ_2 by:

$$\gamma_2 \leq \frac{\frac{1}{2}\mathbb{E}_{\mathbb{P}_2}[D^2]}{\overline{D}} \leq \frac{\frac{1}{2}\left(\frac{1}{3} + \frac{\delta}{2} \times \frac{c}{\sqrt{k}}\right)}{1/2} \leq 1. \tag{95}$$

Therefore, according to (92), for any $\gamma \in [\gamma_1, \gamma_2]$, the derivative $h_2'(\gamma)$ can be upper bounded by:

$$|h_2'(\gamma)| = \mathbb{E}_{\mathbb{P}_2}[(\gamma_2 - D)^+ + D] \leq \gamma_2 + \mathbb{E}_{\mathbb{P}_2}[D] \leq \frac{3}{2}. \tag{96}$$

Notice that $h_2(\gamma_2) = 0$ and $h_1(\gamma_1) = 0$, lower bounding $h_2(\gamma_1) - h_2(\gamma_2)$ is equivalent to lower bounding $h_2(\gamma_1) - h_1(\gamma_1)$, which is as follows:

$$\begin{aligned} &h_2(\gamma_1) - h_1(\gamma_1) \\ =& \mathbb{E}_{\mathbb{P}_2}\left[\frac{1}{2}\left((\gamma_1 - D)^+ + D\right)^2 - \gamma_1\left((\gamma_1 - D)^+ + D\right)\right] \\ &- \mathbb{E}_{\mathbb{P}_1}\left[\frac{1}{2}\left((\gamma_1 - D)^+ + D\right)^2 - \gamma_1\left((\gamma_1 - D)^+ + D\right)\right] \\ =& \frac{c}{\sqrt{k}} \times \int_0^{\delta/2}\Big(\frac{1}{2}\left(\max\{\gamma_1, x + (1-\delta)\}^2 - \max\{\gamma_1, x\}^2\right) \\ &- \gamma_1\left(\max\{\gamma_1, x + (1-\delta)\} - \max\{\gamma_1, x\}\right)\Big)\mathrm{d}x \end{aligned}$$

$$=:\frac{c}{\sqrt{k}}N_1. \tag{97}$$

Plugging (97) and (96) into (94), the lower bound on $(\gamma_2-\gamma_1)$ can be obtained by:

$$(\gamma_2-\gamma_1)\geq\frac{2N_1c}{3}\frac{1}{\sqrt{k}}. \tag{98}$$

Next, we proceed to lower bound $\mathbb{P}_1^{\otimes k}\wedge\mathbb{P}_2^{\otimes k}$. Notice that:

$$\mathbb{P}_1^{\otimes k}\wedge\mathbb{P}_2^{\otimes k}=1-\frac{1}{2}|\mathbb{P}_1^{\otimes k}-\mathbb{P}_2^{\otimes k}|_1, \tag{99}$$

where $|\mathbb{P}-\mathbb{Q}|_1=\int|\mathrm{d}\mathbb{P}-\mathrm{d}\mathbb{Q}|_1$ is the ℓ_1 distance between probability distribution $\mathbb{P}$ and $\mathbb{Q}$. To lower bound $\mathbb{P}_1^{\otimes k}\wedge\mathbb{P}_2^{\otimes k}$, it is sufficient to upper bound $|\mathbb{P}_1^{\otimes k}-\mathbb{P}_2^{\otimes k}|_1$ as follows:

$$\begin{aligned}&\frac{1}{2}\left|\mathbb{P}_1^{\otimes k}-\mathbb{P}_2^{\otimes k}\right|_1\\&\overset{(c)}{\leq}\sqrt{\frac{1}{2}D_{\mathsf{KL}}(\mathbb{P}_2^{\otimes k}||\mathbb{P}_1^{\otimes k})}\\&=\sqrt{\frac{1}{2}kD_{\mathsf{KL}}(\mathbb{P}_2||\mathbb{P}_1)}\\&\overset{(d)}{\leq}\sqrt{\frac{1}{2}k\int_0^1\left(p_2(x)-1+\frac{1}{\min\{p_2(x),1\}}(p_2(x)-1)^2\right)\mathrm{d}x}\\&\overset{(e)}{\leq}\sqrt{\frac{1}{2}k\frac{1}{\inf_{0\leq d\leq 1}p_2(d)}\int_0^1(p_2(x)-1)^2\mathrm{d}x}\\&\leq\sqrt{\frac{1}{2}k\frac{1}{1-c\sqrt{1/k}}\delta\frac{c^2}{k}}\leq\sqrt{\delta c^2},\end{aligned} \tag{100}$$

where inequality (c) is from Pinsker's inequality; where inequality (d) is because the density function $p_1(x)=1$ for uniform distribution, therefore $D_{\mathsf{KL}}(\mathbb{P}_2||\mathbb{P}_1)=\int_0^1p_2(x)\ln p_2(x)\mathrm{d}x$, where $p_2(x)$ is the density function defined in (90); inequality (e) is because function $g(t):=(t\ln t)$ is convex, its derivative $g(t)''=1/t$, therefore, through Taylor expansion we have $g(t)\leq g(1)+(t-1)+\frac{1}{2}\frac{1}{\min\{t,1\}}(t-1)^2=(t-1)+\frac{1}{2}\frac{1}{\min\{t,1\}}(t-1)^2$. By choosing $c=1/2$ and recall that $\delta<1$, inequality (100) can be upper bounded by:

$$\frac{1}{2}|\mathbb{P}_1^{\otimes k}-\mathbb{P}_2^{\otimes k}|_1\leq\frac{1}{2}, \tag{101}$$

Plugging (101) into (99), we have:

$$\mathbb{P}_1^{\otimes k}\wedge\mathbb{P}_2^{\otimes k}\geq 1/2. \tag{102}$$

Finally, plugging (102) and (98) into the Le Cam's inequality (86) yields:

$$\inf_{\hat{\gamma}}\sup_{\mathbb{P}}\mathbb{E}\left[(\hat{\gamma}(\mathcal{H}_k)-\gamma_{\mathbb{P}}^{\star})^2\right]\geq\frac{2N_1^2c^2}{9}\cdot\frac{1}{k}, \tag{103}$$

which verifies (20).

B. Proof of Inequality (21)

We begin the proof of Theorem 2 by introducing the following Lemma:

Lemma 4: Suppose $\gamma^\star$ is the optimum threshold policy $\pi^\star$ selects and let $p_w:=\Pr(D\leq\gamma^\star)$ be the probability of waiting to before taking the next sample. For any stationary policy π, denote $q_\pi:=\mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2\right]$ and $l_\pi:=\mathbb{E}[D+\pi(D)]$ be the expected average reward and length of each cycle, which satisfy the following inequality:

$$q_\pi\geq\gamma^\star l_\pi+\frac{1}{2}p_w\left(l_\pi-\overline{L}^\star\right)^2. \tag{104}$$

Inequality (104) implies, for any causal policy π, the expected reward and frame length satisfy:

$$\begin{aligned}\mathbb{E}\left[Q_k|\mathcal{H}_{k-1}\right]\geq&\gamma^\star\mathbb{E}\left[L_k|\mathcal{H}_{k-1}\right]\\&+\frac{1}{2}p_w\left(\mathbb{E}[L_k|\mathcal{H}_{k-1}]-\overline{L}^\star\right)^2.\end{aligned} \tag{105}$$

Notice that the delay D_k is independent of $\mathcal{H}_{k-1}$ and L_{k-1}. Therefore, $\mathbb{E}[D_kL_{k-1}|\mathcal{H}_{k-1}]=L_{k-1}\overline{D}$. Adding $\mathbb{E}[D_kL_{k-1}|\mathcal{H}_{k-1}]$ on both sides of inequality (105) yields:

$$\begin{aligned}&\mathbb{E}\left[Q_k+D_kL_{k-1}|\mathcal{H}_{k-1}\right]\\&\geq\gamma^\star\mathbb{E}[L_k|\mathcal{H}_{k-1}]+\overline{D}L_{k-1}\\&\quad+\frac{1}{2}p_w\left(\mathbb{E}[L_k|\mathcal{H}_{k-1}]-\overline{L}^\star\right)^2.\end{aligned} \tag{106}$$

For any policy π, denote $z_k(h_k):=\mathbb{E}[L_k|\mathcal{H}_{k-1}=h_{k-1}]$ to be the expected frame-length obtained by π when the historical transmission delay $\mathcal{H}_{k-1}=h_{k-1}=\{d_1,\cdots,d_{k-1}\}$. Summing up (106) from cycle 1 to K and take the expectation with respect to $\mathcal{H}_K$, we have:

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=1}^K(Q_k+D_kL_{k-1})\right]\\&\geq(\gamma^\star+\overline{D})\mathbb{E}\left[\sum_{k=1}^KL_k\right]-\overline{D}(B+W_{\mathsf{ub}})\\&\quad+\frac{1}{2}p_w\mathbb{E}\left[\sum_{k=1}^K(z_k(\mathcal{H}_{k-1})-\overline{L}^\star)^2\right].\end{aligned} \tag{107}$$

Dividing $\mathbb{E}\left[\sum_{k=1}^KL_k\right]$ on both sides of inequality (107) and recall that $\overline{A}_{\pi^\star}=\gamma^\star+\overline{D}$, for any causal policy π, we have:

$$\begin{aligned}&\frac{\mathbb{E}\left[\sum_{k=1}^K(Q_k+D_kL_{k-1})\right]}{\mathbb{E}\left[\sum_{k=1}^KL_k\right]}-\overline{A}_{\pi^\star}\\&\geq-\frac{B+W_{\mathsf{ub}}}{K}\\&\quad+\frac{1}{KL_{\mathsf{ub}}}\times\frac{1}{2}p_w\mathbb{E}\left[\sum_{k=1}^K(z_k(\mathcal{H}_{k-1})-\overline{L}^\star)^2\right].\end{aligned} \tag{108}$$

For any delay distribution $\mathbb{P}\in\mathcal{P}_w(\delta)$, the waiting probability satisfies $p_w\geq\delta$ by definition. Then to establish the lower bound of $\frac{\mathbb{E}\left[\sum_{k=1}^K(Q_k+D_kL_{k-1})\right]}{\mathbb{E}\left[\sum_{k=1}^KL_k\right]}-\overline{A}_{\pi^\star}$, it remains to lower bound $\mathbb{E}\left[(z_k(\mathcal{H}_{k-1})-\gamma^\star)^2\right]$, which is provided in the following lemma:

Lemma 5: For any mapping rule $z_k:\mathbb{R}^k\mapsto\mathbb{R}$, we have the following minimax bound:

$$\begin{aligned}&\inf_{z_{k+1}}\sup_{\mathbb{P}_w(\delta)}\mathbb{E}\left[(z_{k+1}(h_k)-\overline{L}^\star(\mathbb{P}))^2\right]\geq\Omega\left(\frac{1}{k}\right),\\&\forall 0<\delta\leq\left(\sqrt[3]{\frac{1}{2}+\sqrt{\frac{5}{4}}}+\sqrt[3]{\frac{1}{2}-\sqrt{\frac{5}{4}}}\right)/2.\end{aligned} \tag{109}$$

Proof for Lemma 5 is provided in Appendix H.

Therefore, taking the minimax on both sides of inequality (108) and then plugging (109) from Theorem 5 in to the inequality, for any causal policy π, we have:

$$\begin{aligned}&\inf_{\pi\in\Pi}\sup_{\mathbb{P}\in\mathcal{P}_w(\delta)}\left(\frac{\mathbb{E}\left[\int_0^{S_{K+1}}A(t)\mathrm{d}t\right]}{\mathbb{E}[S_{K+1}]}-\overline{A}_{\pi^\star(\mathbb{P})}\right)\\ \ge&\frac{B+W_{\mathsf{ub}}}{K}+\frac{1}{KL_{\mathsf{ub}}}\times\sum_{k=1}^{K}\inf_{z_{k+1}}\sup_{\mathbb{P}\in\mathcal{P}_w(\delta)}\frac{1}{2}p_w(\mathbb{P})\\ &\times\mathbb{E}\left[(z_k(\mathcal{H}_{k-1})-\overline{L}^\star)^2\right]\\ \ge&\delta\cdot\Omega\left(\frac{\sum_{k=2}^{K}\frac{1}{k-1}}{K}\right)=\delta\cdot\Omega\left(\frac{\ln K}{K}\right).\end{aligned}\tag{110}$$

■

Appendix G
Proof of Lemma 4

Proof: Denote $\Pi_l\triangleq\{\pi|\mathbb{E}[D+\pi(D)]=l,\forall\pi\in\Pi\}$ to be the set of stationary policies whose expected cycle length equals l. If l satisfies $\overline{D}\le l\le\overline{D}+W_{\mathsf{ub}}$, set $\Pi_l\neq\emptyset$ because choosing a constant waiting time $\pi(d)\equiv l-\overline{D}$ will lead to an average cycle length of l directly. Next, we establish the lower bound of the expected average reward q_π for any policy $\pi\in\Pi_l$, which can be formulated into an optimization problem:

Problem 4:

$$q_{l,\mathsf{opt}}\triangleq\inf_{\pi}\mathbb{E}\left[\frac{1}{2}(D+\pi(D))^2\right],\tag{111}$$

$$\text{s.t. }\mathbb{E}\left[D+\pi(D)\right]=l.\tag{112}$$

This optimization problem can be solved through a Lagrange multiplier approach. The Lagrange function is as follows:

$$\begin{aligned}\mathcal{L}_1(\pi,\lambda,\mu)\triangleq&\frac{1}{2}\mathbb{E}\left[(D+\pi(D))^2\right]+\lambda(\mathbb{E}[D+\pi(D)]-l)\\&+\mathbb{E}[\pi(D)\mu(D)],\end{aligned}\tag{113}$$

where λ and $\mu(d)\ge 0,\forall d$ are dual variables. For function $\omega(\cdot)\in L_2$, the Gâteaux derivative of the Lagrange function L_1 is denoted by $\delta\mathcal{L}_1(\pi;\lambda,\mu,\omega)$:

$$\begin{aligned}\delta\mathcal{L}_1(\pi,\lambda,\mu;\omega):=&\lim_{\epsilon\to 0}\frac{\mathcal{L}_1(\pi+\epsilon\omega,\lambda,\mu)-\mathcal{L}(\pi,\lambda,\mu)}{\epsilon}\\=&\mathbb{E}\left[(D+\pi(D)+\lambda+\mu(D))\omega(D)\right].\end{aligned}\tag{114}$$

The primal feasibility of the KKT conditions require:

$$\delta\mathcal{L}_1(\pi,\lambda,\mu;\omega)=0,\forall\omega\in L_2,\tag{115a}$$

and the Complete Slackness conditions require:

$$\lambda\left(\mathbb{E}[D+\pi(D)]-l\right)=0,\tag{115b}$$

$$\pi(d)\mu(d)=0,\forall d.\tag{115c}$$

Plugging the expression of the Gâteaux derivative (114) into the KKT condition (115a) and considering the CS conditions in (115b) and (115c), the optimum policy $\pi_l^\star$ to Problem 4 is as follows:

$$\pi_l^\star(d)=(\gamma_l-d)^+,\tag{116}$$

where the selection of γ_l satisfies:

$$\mathbb{E}[(\gamma_l-D)^+]=l-\overline{D}.\tag{117}$$

Before we proceed to lower bound $\mathbb{E}_{\pi_l^\star}\left[\frac{1}{2}(D+\pi(D))^2\right]$, we provide the following statement: recall that $\gamma^\star$ is the optimum updating threshold and leads to an average framelength of $\overline{L}^\star=\mathbb{E}[D+(\gamma^\star-D)^+]$, the difference between γ_l and $\gamma^\star$ can be upper bounded by

$$|\gamma_l-\gamma^\star|\ge|l-\overline{L}^\star|.\tag{118}$$

This is because for any threshold $\gamma_1\ge\gamma_2$, $(\gamma_1-d)^+\ge(\gamma_2-d)^+$ and therefore

$$\begin{aligned}0\le&\mathbb{E}\left[(\gamma_1-D)^++D\right]-\mathbb{E}\left[(\gamma_2-D)^++D\right]\\=&\mathbb{E}\left[(\gamma_1-\gamma_2)\mathbb{I}(D\le\gamma_1)\right]\\&+\mathbb{E}\left[(\gamma_1-D)\mathbb{I}(\gamma_2\le D\le\gamma_1)\right]\\\le&\gamma_1-\gamma_2.\end{aligned}\tag{119}$$

We then lower bound $\mathbb{E}\left[\frac{1}{2}((\gamma_l-D)^++D)^2\right]$ by dividing into the following two cases:

- Case 1: $l\ge\overline{L}^\star$, it can be easily verify that $\gamma_l\ge\gamma^\star$. Therefore, we have:

$$\begin{aligned}&\frac{1}{2}\mathbb{E}\left[((\gamma_l-D)^++D)^2\right]\\=&\frac{1}{2}\mathbb{E}\left[\gamma_l^2\mathbb{I}(D\le\gamma_l)\right]+\frac{1}{2}\mathbb{E}\left[D^2\mathbb{I}(D>\gamma_l)\right]\\=&\frac{1}{2}\mathbb{E}\left[(\gamma^\star)^2\mathbb{I}(D\le\gamma^\star)\right]+\frac{1}{2}\mathbb{E}\left[D^2\mathbb{I}(D>\gamma^\star)\right]\\&+\frac{1}{2}\mathbb{E}\left[(\gamma_l^2-(\gamma^\star)^2)\mathbb{I}(D\le\gamma^\star)\right]\\&+\frac{1}{2}\mathbb{E}\left[(\gamma_l^2-D^2)\mathbb{I}(\gamma^\star\le D\le\gamma_l)\right]\\\overset{(a)}{\ge}&\overline{Q}^\star+\frac{1}{2}\mathbb{E}\left[(\gamma_l-\gamma^\star)^2\mathbb{I}(D\le\gamma^\star)\right]\\&+\mathbb{E}\left[\gamma^\star(\gamma_l-\gamma^\star)\mathbb{I}(D\le\gamma^\star)\right]\\&+\mathbb{E}\left[\gamma^\star(\gamma_l-D)\mathbb{I}(\gamma^\star\le D\le\gamma_l)\right]\\\overset{(b)}{\ge}&\gamma^\star\overline{L}^\star+\frac{1}{2}p_w(\gamma_l-\gamma^\star)^2+\gamma^\star(l-\overline{L}^\star)\\\overset{(c)}{\ge}&\gamma^\star l+\frac{1}{2}p_w(l-\overline{L}^\star)^2,\end{aligned}\tag{120}$$

where inequality (a) is obtained because $\gamma_l^2-(\gamma^\star)^2\ge(\gamma_l-\gamma^\star)^2+2\gamma^\star(\gamma_l-\gamma^\star)$ and for delay d that satisfies $\gamma^\star\le d\le\gamma_l^\star$, $(\gamma_l^\star)^2-d^2=d(\gamma_l^\star-d)\ge\gamma^\star(\gamma_l^\star-d)$; inequality (b) is because $l-\overline{L}^\star=\mathbb{E}\left[(\gamma_l-\gamma^\star)\mathbb{I}(D\le\gamma^\star)\right]+\mathbb{E}\left[(\gamma_l-D)\mathbb{I}(\gamma^\star\le D\le\gamma_l)\right]$ and inequality (c) is obtained because of (119).

- Case 2: $l \leq \overline{L}^{\star}$, similarly, it can be verified that $\gamma_l \leq \gamma^{\star}$. As a result:

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\left[((\gamma_l - D)^{+} + D)^2\right] \\
=&\frac{1}{2}\mathbb{E}\left[\gamma_l^2\mathbb{I}(D \leq \gamma_l)\right] + \frac{1}{2}\mathbb{E}\left[D^2\mathbb{I}(D > \gamma_l)\right] \\
=&\frac{1}{2}\mathbb{E}\left[(\gamma^{\star})^2\mathbb{I}(D \leq \gamma^{\star})\right] + \frac{1}{2}\mathbb{E}\left[D^2\mathbb{I}(D > \gamma^{\star})\right] \\
&-\frac{1}{2}\mathbb{E}\left[((\gamma^{\star})^2 - \gamma_l^2)\mathbb{I}(D \leq \gamma^{\star})\right] \\
&-\frac{1}{2}\mathbb{E}\left[(D^2 - \gamma_l^2)\mathbb{I}(\gamma_l \leq D \leq \gamma^{\star})\right] \\
=&\overline{Q}^{\star} + \frac{1}{2}\mathbb{E}\left[(\gamma_l - \gamma^{\star})^2\mathbb{I}(D \leq \gamma^{\star})\right] \\
&+\mathbb{E}\left[\gamma^{\star}(\gamma_l - \gamma^{\star})\mathbb{I}(D \leq \gamma^{\star})\right] \\
&-\mathbb{E}\left[\gamma^{\star}(\gamma^{\star} - D)\mathbb{I}(\gamma_l \leq D \leq \gamma^{\star})\right] \\
\overset{(d)}{\geq}&\gamma^{\star}\overline{L}^{\star} + \frac{1}{2}p_w(l - \overline{L}^{\star})^2 - \gamma^{\star}(\overline{L}^{\star} - l) \\
=&\gamma^{\star} l + \frac{1}{2}p_w(l - \overline{L}^{\star})^2,
\end{aligned} \tag{121}$$

where inequality (d) is obtained similarly as inequality (a)-(c).

∎

Appendix H
Proof of Lemma 5

Proof: The minimax risk bound on $\hat{l} - \overline{L}^{\star}$ is established similarly using the Le Cam's two point method. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two delay distribution from $\mathcal{P}_w(\delta)$ and denote $l_1 := \mathbb{E}_{\mathbb{P}_1}[(\gamma_1 - D)^{+} + D]$, $l_2 := \mathbb{E}_{\mathbb{P}_2}[(\gamma_2 - D)^{+} + D]$ be the optimum frame length by using AoI minimum policies $\pi^{\star}_{\mathbb{P}_1}$ and $\pi^{\star}_{\mathbb{P}_2}$. By Le Cam's inequality, we have:

$$\inf_{\hat{l}} \sup_{\mathbb{P}\in\mathcal{P}_w(\delta)} \mathbb{E}[(\hat{l}(\mathcal{H}_k) - \overline{L}^{\star}(\mathbb{P}))^2] \geq (l_1 - l_2)^2 \cdot \mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}. \tag{122}$$

Similar to the proof of (108) in Appendix F-A, we choose $\mathbb{P}_1$ to be the uniform distribution and $\mathbb{P}_2$ is defined through (90). Since δ is selected to be $\delta \leq p_{\text{w, uni}}/2$, it is easy to show that $p_w(\mathbb{P}_2) \geq \delta$ as follows:

$$\begin{aligned}
&p_w(\mathbb{P}_2) = \mathbb{E}_{\mathbb{P}_2}[\mathbb{I}_{(D\leq\gamma_2)}] \\
=&\int_0^1 \mathbb{I}_{(x\leq\gamma_2)}\mathrm{d}x - \int_0^{\delta/2} \frac{c}{\sqrt{k}}\mathbb{I}_{(x\leq\gamma_2)}\mathrm{d}x \\
&+\int_{1-\delta/2}^1 \frac{c}{\sqrt{k}}\mathbb{I}_{(x\leq\gamma_2)}\mathrm{d}x \\
\overset{(a)}{\geq}&\int_0^1 \mathbb{I}_{(x\leq\gamma_1)}\mathrm{d}x - \frac{c}{\sqrt{k}}\delta \overset{(b)}{\geq} p_{\text{w, uni}}/2.
\end{aligned} \tag{123}$$

where inequality (a) holds because $\gamma_1 \leq \gamma_2$ and inequality (b) holds because $\delta < p_{\text{w, uni}}/2$ by definition.

To use the Le Cam's two point method, we then need to lower bound $l_2 - l_1$ and $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$, respectively. The lower bound on $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$ can be obtained in (102) and lower bound on $l_2 - l_1$ can be obtained as follows:

$$\begin{aligned}
&l_2 - l_1 \\
=&\mathbb{E}_{\mathbb{P}_2}\left[(\gamma_2 - D)^{+} + D\right] - \mathbb{E}_{\mathbb{P}_2}\left[(\gamma_1 - D)^{+} + D\right] \\
=&\int_0^1 \max\{\gamma_2, x\}\mathrm{d}x + \int_{1-\delta/2}^1 \frac{c}{\sqrt{k}}\max\{\gamma_2, x\}\mathrm{d}x \\
&-\int_0^{\delta/2} \frac{c}{\sqrt{k}}\max\{\gamma_2, x\}\mathrm{d}x - \int_0^1 \max\{\gamma_1, x\}\mathrm{d}x \\
\overset{(a)}{\geq}&\int_0^1 \max\{\gamma_2, x\}\mathrm{d}x - \int_0^1 \max\{\gamma_1, x\}\mathrm{d}x \\
\geq&\gamma_1(\gamma_2 - \gamma_1) \\
\overset{(b)}{\geq}&\frac{2N_1 c\gamma_1}{3}\frac{1}{\sqrt{k}},
\end{aligned} \tag{124}$$

where inequality (a) is because for $x \in [0, \delta/2]$, we have $\max\{\gamma_2, x + (1 - \delta)\} - \max\{\gamma_2, x\} \geq 0$ and therefore $\int_{1-\delta/2}^1 \frac{c}{\sqrt{k}}\max\{\gamma_2, x\}\mathrm{d}x - \int_0^{\delta/2} \frac{c}{\sqrt{k}}\max\{\gamma_2, x\}\mathrm{d}x \geq 0$; inequality (b) is from (98).

Plugging (124) and (102) into the Le Cam's inequality (122), we have:

$$\inf_{\hat{l}} \sup_{\mathbb{P}_w(\delta)} \mathbb{E}[(\hat{l}(\mathcal{H}_k) - \overline{L}^{\star}(\mathbb{P}))^2] \geq \frac{2N_1^2 c^2 \gamma_1^2}{9} \cdot \frac{1}{k}. \tag{125}$$

∎

Appendix I
Proof of Theorem 5

Proof: Recall from equation (16d), the sampling debt evolves like a queueing system:

$$U_{k+1} = \left(U_k + \left(\frac{1}{f_{\mathsf{max}}} - L_k\right)\right)^{+}.$$

To show that the proposed policy satisfies the sampling constraint, i.e., the sampling debt queue is stable, it is sufficient to prove that [35, Theorem 2.8]

$$\limsup_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[U_k\right] < \infty. \tag{126}$$

This motivates us to adopt the Lyapunov-Drift-Plus-Penalty approach to prove the virtual queue of the unused sampling frequency is stable. Define the Lyapunov function to be:

$$J(U_k) := \frac{1}{2}U_k^2, \tag{127}$$

and the Lyapunov Drift is defined by

$$\Delta(U_k) := \mathbb{E}\left[J(U_{k+1}) - J(U_k)|\mathcal{H}_{k-1}\right]. \tag{128}$$

To upper bound the Lyapunov drift, notice that U_k^2 can be upper bounded by:

$$\begin{aligned}
U_{k+1}^2 &= \left[\max\{U_k - L_k + \frac{1}{f_{\mathsf{max}}}, 0\}\right]^2 \\
&\leq \left[U_k - L_k + \frac{1}{f_{\mathsf{max}}}\right]^2.
\end{aligned} \tag{129}$$

Then, considering the fact that both the waiting time and delay is upper bounded, i.e., $W_k \leq W_{\mathsf{ub}}$ and $D_k \leq B$, the cycle length satisfies $L_k \leq W_{\mathsf{ub}} + B$, $J(U_{k+1}) - J(U_k))$ can be upper bounded as follows:

$$\begin{aligned} J(U_{k+1}) - J(U_k) &= \frac{1}{2}\left(U_{k+1}^2 - U_k^2\right) \\ &\overset{(a)}{\leq} \frac{1}{2}\left(\left[U_k - L_k + \frac{1}{f_{\mathsf{max}}}\right]^2 - U_k^2\right) \\ &\leq -U_k\left(L_k - \frac{1}{f_{\mathsf{max}}}\right) + \frac{1}{2}\left((B + W_{\mathsf{ub}})^2 + \frac{1}{f_{\mathsf{max}}^2}\right). \quad (130) \end{aligned}$$

where inequality (a) is due to (129).

Taking the conditional expectation of (130) with respect to the transmission delay D_k, the Lyapunov drift $\Delta(U_k) = \mathbb{E}\left[J(U_{k+1}) - J(U_k)|\mathcal{H}_{k-1}\right]$ can be upper bounded by:

$$\begin{aligned} \Delta(U_k) \leq& -U_k\mathbb{E}\left[L_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right] \\ &+ \frac{1}{2}\left((B + W_{\mathsf{ub}})^2 + \frac{1}{f_{\mathsf{max}}^2}\right). \quad (131) \end{aligned}$$

The following Lemma establishes an upper bound on $\mathbb{E}\left[L_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right]$, the proof will be given in Appendix J:

Lemma 6: Assumption 2 enables us to upper bound term $-U_k\mathbb{E}\left[L_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right]$ via the following inequality:

$$\begin{aligned} &-U_k\mathbb{E}\left[L_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right] \\ &\leq -\epsilon U_k + V\left(\frac{1}{2}(B + W_{\mathsf{ub}})^2 + \gamma_{\mathsf{ub}}(B + W_{\mathsf{ub}})\right). \quad (132) \end{aligned}$$

Plugging inequality (132) into (131), the Lyapunov drift can be upper bounded by:

$$\begin{aligned} \Delta(U_k) \leq& -\epsilon U_k + \frac{1}{2}\left((B + W_{\mathsf{ub}})^2 + \frac{1}{f_{\mathsf{max}}^2}\right) \\ &+ V\left(\frac{1}{2}(B + W_{\mathsf{ub}})^2 + \gamma_{\mathsf{ub}}(B + W_{\mathsf{ub}})\right). \quad (133) \end{aligned}$$

For simplicity, denote by

$$\begin{aligned} C :=& \frac{1}{2}\left((B + W_{\mathsf{ub}})^2 + \frac{1}{f_{\mathsf{max}}^2}\right) + \\ &V\left(\frac{1}{2}(B + W_{\mathsf{ub}})^2 + \gamma_{\mathsf{ub}}(B + W_{\mathsf{ub}})\right) < \infty. \quad (134) \end{aligned}$$

Summing up inequality (133) from cycle $k = 1$ to K and taking the expectation with respect to historical information $\mathcal{H}_K$, we have:

$$\mathbb{E}\left[\frac{1}{2}U_{K+1}^2 - \frac{1}{2}U_1^2\right] \leq -\epsilon\mathbb{E}\left[\sum_{k=1}^{K} U_k\right] + KC. \quad (135)$$

Finally, recall that $U_1 = 0$ and $U_{K+1} \geq 0$, adding $\sum_{k=1}^{K}\mathbb{E}[U_k]$ on both sides of inequality (135) yields:

$$\epsilon\sum_{k=1}^{K}\mathbb{E}\left[U_k\right] \leq KC. \quad (136)$$

Taking the limit $K \to \infty$ yields:

$$\limsup_{K\to\infty}\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K} U_k\right] < \frac{C}{\epsilon} < \infty, \quad (137)$$

which verifies condition (126) and shows that the proposed method satisfies the sampling constraint. ■

Appendix J
Proof of Lemma 6

Proof: Denote function

$$f(u, w, d) := -u(w + d) + V\left(\frac{1}{2}(d + w)^2 - \gamma(d + w)\right).$$

The partial derivative with respect to w can be computed by:

$$\frac{\partial f(u, w, d)}{\partial w} = V\left(w + d - \left(\gamma + \frac{1}{V}u\right)\right).$$

Therefore, for given u and d, the optimum $w \geq 0$ that minimizes $f(u, w, d)$ is:

$$\arg\min_{w\geq 0} f(u, w, d) = \left(\gamma + \frac{1}{V}u - d\right)^+. \quad (138)$$

Recall from equation (16a), the selection rule of the waiting time is:

$$W_k = \left(\gamma_k + \frac{1}{V}U_k - D_k\right)^+.$$

Therefore, according to (138), the selection rule W_k of the proposed algorithm minimizes function $f(u, w, d)$ when the sampling frequency violation $u = U_k$ and the transmission delay $d = D_k$. As a result, for any other waiting time specified by policy $W = \pi(D)$, we have

$$\begin{aligned} &-U_k(W_k + D_k) \\ &+ V\left(\frac{1}{2}(D_k + W_k)^2 - \gamma_k(D_k + W_k)\right) \\ \leq& -U_k(\pi(D_k) + D_k) \\ &+ V\left(\frac{1}{2}(D_k + \pi(D_k))^2 - \gamma_k(D_k + \pi(D_k))\right). \quad (139) \end{aligned}$$

Adding $U_k\frac{1}{f_{\mathsf{max}}}$ on both sides of inequality (139), then taking the conditional expectation with respect to delay D_k given historical information $\mathcal{H}_{k-1}$, we have:

$$\begin{aligned} &-U_k\mathbb{E}\left[D_k + W_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right] \\ &+ V\mathbb{E}\left[\frac{1}{2}(D_k + W_k)^2 - \gamma_k(D_k + W_k)|\mathcal{H}_{k-1}\right] \\ \leq& -U_k\mathbb{E}\left[\pi(D_k) + D_k - \frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right] \\ &+ V\mathbb{E}\left[\frac{1}{2}(D_k + \pi(D_k))^2 - \gamma_k(D_k + \pi(D_k))|\mathcal{H}_{k-1}\right]. \end{aligned} \quad (140)$$

According to Assumption 2, the sampling frequency constraint (5b) can be strictly satisfied by using policy π_ϵ, i.e.,

$$\mathbb{E}[D + \pi_\epsilon(D)] \geq \frac{1}{f_{\mathsf{max}}} + \epsilon. \quad (141)$$

Considering that the transmission delay D_k is i.i.d., plugging (141) into (140) yields

$$\begin{aligned} & -U_k\mathbb{E}\left[L_k-\frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right]\\ &+V\,\mathbb{E}\left[\frac{1}{2}(D_k+W_k)^2-\gamma_k(D_k+W_k)|\mathcal{H}_{k-1}\right]\\ \le & -U_k\mathbb{E}\left[D_k+\pi_\epsilon(D_k)-\frac{1}{f_{\mathsf{max}}}\right]\\ &+V\,\mathbb{E}\left[\frac{1}{2}(D_k+\pi_\epsilon(D_k))^2-\gamma_k(D_k+\pi_\epsilon(D_k))|\mathcal{H}_{k-1}\right]\\ \le & -\epsilon U_k\\ &+V\,\mathbb{E}\left[\frac{1}{2}(D_k+\pi_\epsilon(D_k))^2-\gamma_k(D_k+\pi_\epsilon(D_k))|\mathcal{H}_{k-1}\right]. \end{aligned} \tag{142}$$

Notice that $\gamma_k \le \gamma_{\mathsf{ub}}$ and $D_k \le B, \pi_\epsilon(d) \le W_{\mathsf{ub}}$, inequality (142) can be simplified to:

$$\begin{aligned} & -U_k\mathbb{E}\left[L_k-\frac{1}{f_{\mathsf{max}}}|\mathcal{H}_{k-1}\right]\\ \le & -\epsilon U_k+V\left(\frac{1}{2}(B+W_{\mathsf{ub}})^2+\gamma_{\mathsf{ub}}(B+W_{\mathsf{ub}})\right). \end{aligned} \tag{143}$$

■

Acknowledgment

The authors would like to thank Prof. Yin Sun for pointing out the ODE approach for the proof of Theorem 2 and for insightful suggestions on Theorem 4, Prof. Yuhao Wang for discussions on the proof of Theorem 5, and Dr. Hengjie Yang for a careful proofreading of the draft.

References

[1] H. Tang, Y. Chen, J. Sun, J. Wang, and J. Song, "Sending timely status updates through channel with random delay via online learning," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2022, pp. 1819–1827.

[2] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[3] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Guest editorial age of information," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1179–1182, May 2021.

[4] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.

[5] Y. Wang and W. Chen, "Adaptive power and rate control for real-time status updating over fading channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3095–3106, May 2021.

[6] B. Wang, S. Feng, and J. Yang, "When to preempt? Age of information minimization under link capacity constraint," *J. Commun. Netw.*, vol. 21, no. 3, pp. 220–232, Jun. 2019.

[7] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the Internet of Things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7468–7482, Nov. 2019.

[8] H. Tang, J. Wang, L. Song, and J. Song, "Minimizing age of information with power constraints: Multi-user opportunistic scheduling in multi-state time-varying channels," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 854–868, Mar. 2020.

[9] E. T. Ceran, D. Gunduz, and A. Gyorgy, "Average age of information with hybrid ARQ under a resource constraint," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[10] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in RF-powered communication systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4747–4760, Aug. 2020.

[11] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Online timely status updates with erasures for energy harvesting sensors," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 966–972.

[12] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Optimal sampling and scheduling for timely status updates in multi-source networks," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 4019–4034, Jun. 2021.

[13] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 2022–2027.

[14] E. Najm, R. Nasser, and E. Telatar, "Content based status updates," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3846–3863, Jun. 2020.

[15] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 3008–3012.

[16] K. Bhandari, S. Fatale, U. Narula, S. Moharir, and M. K. Hanawal, "Age-of-information bandits," in *Proc. 18th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOPT)*, Jun. 2020, pp. 1–8.

[17] E. U. Atay, I. Kadota, and E. Modiano, "Aging wireless bandits: Regret analysis and order-optimal learning algorithm," in *Proc. 19th Int. Symp. Modeling Optim. Mobile, Ad hoc, Wireless Netw. (WiOpt)*, 2021, pp. 1–8, doi: 10.23919/WiOpt52861.2021.9589673.

[18] S. Banerjee, R. Bhattacharjee, and A. Sinha, "Fundamental limits of age-of-information in stationary and non-stationary environments," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1741–1746.

[19] V. Tripathi and E. Modiano, "An online learning approach to optimizing time-varying costs of AoI," in *Proc. Twenty-second Int. Symp. Theory, Algorithmic Found., Protocol Design Mobile Netw. Mobile Comput.*, Jul. 2021, pp. 241–250.

[20] B. Li, "Efficient learning-based scheduling for information freshness in wireless networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2021, pp. 1–10.

[21] E. T. Ceran, D. Gunduz, and A. Gyorgy, "Reinforcement learning to minimize age of information with an energy harvesting sensor with HARQ and sensing cost," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 656–661.

[22] E. T. Ceran, D. Gunduz, and A. Gyorgy, "A reinforcement learning approach to age of information in multi-user networks with HARQ," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1412–1426, May 2021.

[23] C. Kam, S. Kompella, and A. Ephremides, "Learning to sample a signal through an unknown system for minimum AoI," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 177–182.

[24] S. Leng and A. Yener, "Age of information minimization for wireless ad hoc networks: A deep reinforcement learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[25] C.-H. Tsai and C.-C. Wang, "Age-of-information revisited: Two-way delay and distribution-oblivious online algorithm," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1782–1787.

[26] C.-H. Tsai and C.-C. Wang. (2022). *Distribution-Oblivious Online Algorithms for Age-of-Information Penalty Minimization*. [Online]. Available: https://docs.lib.purdue.edu/ecetr/759/

[27] A. Arafa, R. D. Yates, and H. V. Poor, "Timely cloud computing: Preemption and waiting," in *Proc. 57th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2019, pp. 528–535.

[28] S. M. Ross, *Applied Probability Models With Optimization Applications*. Chelmsford, MA, USA: Courier Corporation, 2013.

[29] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.

[30] H. Tang, Y. Sun, and L. Tassiulas, "Sampling of the Wiener process for remote estimation over a channel with unknown delay statistics," in *Proc. 33rd Int. Symp. Theory, Algorithmic Found., Protocol Design Mobile Netw. Mobile Comput.*, Oct. 2022, pp. 51–60.

[31] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396–409, Apr. 2008.

[32] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer, 2003.

[33] B. Yu, *Assouad, Fano, and Le Cam*. New York, NY, USA: Springer, 1997, pp. 423–435.

[34] L. L. Cam, *Asymptotic Methods in Statistical Decision Theory*. Cham, Switzerland: Springer, 2012.

[35] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lect. Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.

Haoyue Tang (Student Member, IEEE) received the B.Eng. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2017 and 2022, respectively. She was a Visiting Student at Technische Universitat München from September 2015 to February 2016 and Télécom Paris from January 2019 to March 2019. She is currently a Post-Doctoral Research Associate with Yale University. Her research interests include age of information, stochastic network optimization, and statistical learning theory.

Yuchao Chen received the B.Eng. degree in electrical engineering from Tsinghua University, Beijing, China, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His research interests include stochastic networking optimization, online learning, and wireless scheduling.

Jintao Wang (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2001 and 2006, respectively. From 2006 to 2009, he was an Assistant Professor at the Department of Electronic Engineering, Tsinghua University. Since 2009, he has been an Associate Professor and Ph.D. Supervisor. He is the Standard Committee Member of the Chinese National Digital Terrestrial Television Broadcasting Standard. He has authored or coauthored more than 100 journal and conference papers and holds more than 40 national invention patents. His research interests include space-time coding, MIMO, and OFDM systems.

Pengkun Yang received the B.E. degree from the Department of Electronic Engineering, Tsinghua University, in 2013, the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign. He is currently an Assistant Professor with the Center for Statistical Science, Tsinghua University. His research interests include statistical inference, learning, and optimization and systems. He was a recipient of the Jack Keil Wolf ISIT Student Paper Award from the 2015 IEEE International Symposium on Information Theory.

Leandros Tassiulas (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1991, and the Diploma degree in electrical engineering from the Aristotele University of Thessaloniki, Greece. He was a Faculty Member at the Polytechnic University, New York, NY, USA, University of Maryland, and University of Thessaly, Greece. He is currently the John C. Malone Professor of electrical engineering with Yale University, New Haven, CT, USA. His most notable contributions include the max-weight scheduling algorithm and the back-pressure network control policy, opportunistic scheduling in wireless, the maximum lifetime approach for wireless network energy management, and the consideration of joint access control and antenna transmission management in multiple antenna wireless systems. He was worked in the field of computer and communication networks with emphasis on fundamental mathematical models and algorithms of complex networks, wireless systems and sensor networks. His current research interests include intelligent services and architectures at the edge of next generation networks including the Internet of Things, sensing and actuation in terrestrial, and non terrestrial environments. His research has been recognized by several awards, including the IEEE Koji Kobayashi Computer and Communications Award in 2016, the ACM SIGMETRICS achievement award 2020, the Inaugural INFOCOM 2007 Achievement Award for fundamental contributions to resource allocation in communication networks, the INFOCOM 1994 and 2017 Best Paper Awards, the National Science Foundation (NSF) Research Initiation Award in 1992, the NSF CAREER Award in 1995, the Office of Naval Research Young Investigator Award in 1997, and the Bodossaki Foundation Award in 1999. He is a several best paper awards including the INFOCOM 1994, 2017 and Mobihoc 2016. He is a Fellow of ACM in 2020.