Sampling for Remote Estimation of an Ornstein-Uhlenbeck Process through Channel with Unknown Delay Statistics

Yuchao Chen*, Haoyue Tang*, Jintao Wang, Pengkun Yang, and Leandros Tassiulas

***Abstract*—In this paper, we consider sampling an Ornstein-Uhlenbeck (OU) process through a channel for remote estimation. The goal is to minimize the mean square error (MSE) at the estimator under a sampling frequency constraint when the channel delay statistics is unknown. Sampling for MSE minimization is reformulated into an optimal stopping problem. By revisiting the threshold structure of the optimal stopping policy when the delay statistics is known, we propose an online sampling algorithm to learn the optimum threshold using stochastic approximation algorithm and the virtual queue method. We prove that with probability 1, the MSE of the proposed online algorithm converges to the minimum MSE that is achieved when the channel delay statistics is known. The cumulative MSE gap of our proposed algorithm compared with the minimum MSE up to the $(k+1)$th sample grows with rate at most $\mathcal{O}(\ln k)$. Our proposed online algorithm can satisfy the sampling frequency constraint theoretically. Finally, simulation results are provided to demonstrate the performance of the proposed algorithm.**

***Index Terms*—Online learning, Ornstein-Uhlenbeck process, stochastic approximation**

I. Introduction

With the rapid development of the autonomous vehicles [1] and intelligent machine communications [2], status update information (e.g., the speed of the vehicles) is becoming a major part in future communication networks [3]. Those status information are delivered to the destination through communication channels, and to guarantee the system safety and efficient control, it is necessary to ensure that the controller has an accurate estimation of the system state.

To measure the information freshness at the destination, the metric, age of information (AoI), has been proposed in [4]. According to the definition, AoI measures the difference between the current time and the generation time of the latest information received at the destination. Previous work [5], [6] have shown that AoI minimization is different from the traditional throughput and delay optimization. Specifically in the data generation procedure, a new data sample should be made only when the data stored at the destination is old. Numerous research have been conducted to minimize the AoI in various networks [4]–[11]. The average AoI optimization in the queueing system is studied in [4], [7]. Age-optimal scheduling policies in a multi-user wireless network are also investigated in [9]–[12]. For minimizing the more general non-linear age function, [6], [8] also design the optimal sampling strategies.

However, when the signal model is known, AoI itself cannot reflect the different signal evolution. As an alternative, a better metric to capture information freshness at the destination is the mean square error (MSE) [13]–[21]. The sampling strategy to minimize the estimation MSE of a Wiener process is studied in [14], [15], [20]. Sampling strategy to minimize an Ornstein-Uhlenbeck (OU) process is investigated in [14], [21]. It is revealed that the optimum sampling threshold depends on signal evolution and channel delay statistics. When the channel delay statistics is known, the aforementioned optimum sampling thresholds can be computed numerically by fixed-point iteration [19] or bi-section search [20], [21].

When the channel statistics of the communication link is unknown, finding the optimum policy (i.e., the optimum AoI [6] or signal difference threshold [20], [21]) is challenging. Designing an adaptive sampling and transmission strategy under unknown channel statistics for data freshness optimization can be formulated into a sequential decision-making process [22]–[29]. Based on the stochastic multi-armed bandit, [22]–[24] design online channel selection algorithms to minimize average AoI performance for the ON-OFF channel with unknown transition probability. For channels with more efficient communication protocols, [30]–[32] use reinforcement learning to minimize the AoI performance under unknown channel statistics. For communication channels with random delay, [28], [29], [33] apply the stochastic approximation method to design adaptive sampling algorithms to optimize AoI performance. The stochastic approximation

Manuscript received May 29, 2023; revised August 1, 2023; approved for publication by Yin Sun, Guest Editor, August 17, 2023.

*Equal Contribution.

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0101700; in part by the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant JSGG20211029095003004. The work of P. Yang is supported by NSFC Grant 12101353, Tsinghua University Initiative Scientific Research Program. The work of H. Tang and L. Tassiulas was supported by the NSF CNS-2112562 AI Institute for Edge Computing Leveraging Next Generation Networks (Athena) and the ONR N00014-19-1-2566.

Y. Chen and J. Wang are with Beijing National Research Center for Information Science and Technology (BNRist) and the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. J. Wang is also with Key Laboratory of Digital TV System of Guangdong Province and Shenzhen City, Research Institute of Tsinghua University in Shenzhen, Shenzhen, China, email: {cyc20@mails.; wangjintao@}tsinghua.edu.cn.

P. Yang is with the Center for Statistical Science, Tsinghua University, Beijing 100084, China, email: yangpengkun@tsinghua.edu.cn.

H. Tang and L. Tassiulas are with the Department of Electrical Engineering and Institute for Network Science, Yale University, New Haven, CT, USA, email: {haoyue.tang;leandros.tassiulas}@yale.edu.

H. Tang is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2023.000037

1229-2370/23/$10.00 © 2023 KICS

method can also be extended to online estimation of the signals with simple evolution model, i.e., the Wiener process [34].

Notice that the Wiener process is the simplest time-varying signal model, and we are interested in extending the results to handle more general and complex signal models. In this paper, we consider a point-to-point link with a sensor sampling an OU process and transmitting the sampled packet to the destination through a channel with random delay for remote estimation. Our goal is to design an online sampling policy to minimize the average MSE under a frequency constraint when the channel statistics is unknown. The main contributions of the work are listed as follows:

- We reformulated the MSE minimum sampling problem under the unknown channel statistics as an optimal stopping problem by providing a novel frame division algorithm that is different from [21]. This novel approach of frame division enables us to propose an online sampling algorithm to learn the optimal threshold adaptively through stochastic approximation and virtual queue method.
- When there is no sampling frequency constraint, we proved that the expected average MSE of the proposed algorithm can converge to the minimum MSE almost surely. Specifically, we first utilized the property of the OU process to bound the threshold parameter (Lemma 2 and Lemma 6), and then we proved the cumulative MSE regret grows at the speed of $\mathcal{O}(\ln K)$, where K is the number of samples (Theorem 2) we have taken.
- When there exists a sampling frequency constraint, by viewing the sampling frequency debt as a virtual queue, we proved that the sampling frequency constraint can be satisfied in the sense that the virtual queue is stable (Theorem 3).

The rest of the paper is organized as follows. In Section II, we introduce the system model and formulate the MSE minimization problem. In Section III, we reformulate the problem into an optimal stopping optimization and then propose an online sampling algorithm. The theoretical analysis of the proposed algorithm is provided in Section IV. In Section V, we present the simulation results. Finally, conclusions are drawn in Section VI.

II. Problem Formulation

A. System Model

As depicted in Fig. 1, we study a status update system similar to [21], where a sensor observes a time-varying process and sends the sampled data to the remote estimator through a channel. Let $X_t \in \mathbb{R}, \forall t \geq 0$ denote the value of the time-varying process at time t. To model these time-varying first-order auto-regressive processes, we assume X_t to be an OU process in this work. This general process is the only nontrivial continuous-time process that is stationary, Gaussian, and Markovian [35]. The OU process evolution parameterized by $\mu, \theta, \sigma \in \mathbb{R}^+$ can be modeled by the following stochastic differential equation (SDE) [35]:

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t,$$

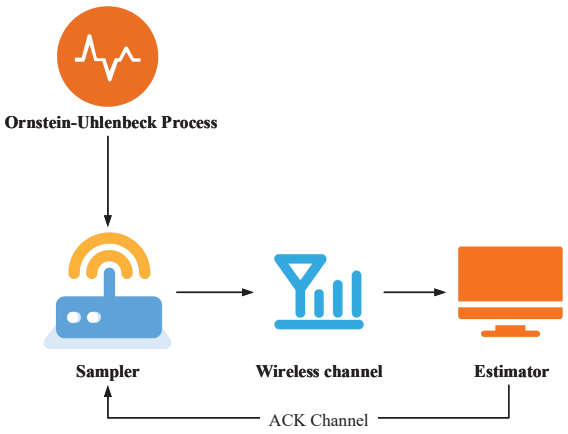


Fig. 1. A point-to-point status update system.

where W_t is a Wiener process.

Suppose the sensor can sample the process at any time $t \in \mathbb{R}^+$ at his own will. Let S_k be the sampling time-stamp of the kth sample. Once sample k is transmitted over the channel, it will experience a random delay $D_k \in [0, \infty)$ to reach the destination. We assume the transmission delay is independent and identically distributed (i.i.d.) following a probability measure $\mathbb{P}_D$.

Due to the interference constraint, only one sample can be transmitted over the channel at one time. Once the transmission of an update finishes, an ACK signal will be sent to the sensor without error immediately. Let R_k be the reception time of the kth sample. Then we can compute R_k iteratively by

$$R_k = \max\{S_k, R_{k-1}\} + D_k. \tag{1}$$

B. Minimum Mean Squared Error (MMSE) Estimation

The receiver attempts to estimate the value of X_t based on the received packets and the transmission results before time t. Let $i(t) = \max_{k\in\mathbb{N}}\{k|R_k \leq t\}$ be the index of the latest received sample at time t. The evolution of X_t can be rewritten using the strong Markov property of the OU process [21, equation (8)] as follows.

$$\begin{aligned} X_t =& X_{S_{i(t)}} e^{-\theta(t-S_{i(t)})} + \mu\left[1 - e^{-\theta(t-S_{i(t)})}\right] \\ &+ \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-S_{i(t)})} W_{e^{2\theta(t-S_{i(t)})}-1} \end{aligned} \tag{2}$$

Let $\mathcal{H}_t := \left(\{S_k, D_k, X_{S_k}\}_{k=1}^{i(t)}, t\right)$ be the historical information up to time t. Then, the MMSE estimator at the destination is the conditional expectation [36]:

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{H}_t] = X_{S_{i(t)}} e^{-\theta(t-S_{i(t)})} + \mu\left[1 - e^{-\theta(t-S_{i(t)})}\right] \tag{3}$$

Combined with (2), the instant estimation error at time t, denoted by Δ_t can be computed as

$$\Delta_t = X_t - \hat{X}_t = \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-S_{i(t)})} W_{e^{2\theta(t-S_{i(t)})}-1}, \tag{4}$$

which can be viewed as an OU process starting at time $t = S_{i(t)}$.

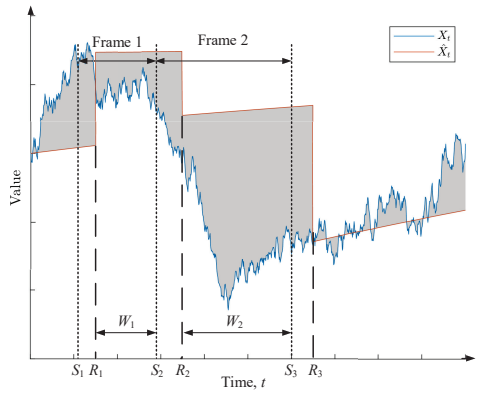


Fig. 2. Illustration of the OU process and the estimation error.

To better demonstrate the MMSE estimation, we draw Fig. 2 as an example. The blue line is a sample path of an OU process, and the orange line is the MMSE estimator computed by (3). Then the difference between these two lines, i.e., the shaded area, is the cumulative estimation error between the two samples.

C. Optimization Problem

The goal of the sampler is to find a sampling policy represented by a series of sampling times, i.e., $\pi := \{S_1, S_2, \cdots\}$ to minimize the estimation MSE of the OU process at the destination. We assume that the sampler knows the statistical information of the OU process, i.e., parameters θ, μ, σ, while the channel delay statistics $\mathbb{P}_D$ is unknown. Here we focus on the set of causal sampling policies denoted by Π. The sampling time S_k selected by each policy $\pi \in \Pi$ is determined only by the historical information. No future information can be used for the sampling decision. Moreover, due to the hardware constraint and energy conservation, the average sampling frequency during the transmission should be below a certain threshold $f_{\max}$. Then, the optimization problem can be formulated as

Problem 1 (MSE minimization):

$$\text{mmse} \triangleq \inf_{\pi\in\Pi} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T (X_t - \hat{X}_t)^2 \mathrm{d}t\right], \tag{5a}$$

$$\text{s.t.} \limsup_{T\to\infty} \mathbb{E}\left[\frac{i(T)}{T}\right] \le f_{\max}. \tag{5b}$$

III. Problem Resolution

In this section, we first reformulate the Problem 1 into an optimal stopping problem. Then, an online sampling algorithm is proposed to approach the optimal mmse.

A. Optimal Stopping Problem Reformulation

Notice that Problem 1 is a constrained continuous-time Markov decision process (MDP) with a continuous state space. It has been proven in [21, Lemma 6] that it is sub-optimal to take a new sample before the last packet is received by the receiver. In other words, to achieve the optimal mmse, the sampling time-stamp S_k should be larger than R_{k-1}. Then (1) can be simplified as $R_k = S_k + D_k$. Let $W_k = S_{k+1} - R_k$ be the waiting time before taking the $(k+1)$th sample. Then, designing a sampling policy $\pi = \{S_1, S_2, \cdots\}$ is equivalent to choosing a sequence of waiting time $\{W_1, W_2, \cdots\}$. To facilitate further analysis, define frame k to be the time interval between S_k and S_{k+1}. Then, we introduce the following lemma to reformulate the Problem 1 into the packet-level MDP.

Lemma 1: Define $\mathcal{I}_k = (D_k, \{X_t\}_{t\ge S_k})$ to be the information in frame k, and Π_r to be the set of stationary sampling policies whose W_k only depends on $\mathcal{I}_k$. Let D be the random delay following distribution $\mathbb{P}_D$. Then Problem 1 can be reformulated into the following MDP:

Problem 2 (Packet-level MDP reformulation):

$$\alpha^\star \triangleq \sup_{\pi\in\Pi_r} \left(\lim_{K\to\infty} \frac{\sum_{k=1}^K \mathbb{E}[O^2_{D_k+W_k}]}{\sum_{k=1}^K \mathbb{E}\left[D_k + W_k\right]}\right), \tag{6a}$$

$$\text{s.t.} \liminf_{K\to\infty} \frac{1}{K}\sum_{k=1}^K \mathbb{E}\left[W_k + D_k\right] \ge \frac{1}{f_{\max}}, \tag{6b}$$

where O_t is an OU process with initial state $O_t = 0$ and parameter $\mu = 0$, which is the solution to the SDE:

$$\mathrm{d}O_t = -\theta O_t \mathrm{d}t + \sigma \mathrm{d}W_t. \tag{7}$$

Moreover, the optimum value $\alpha^\star$ satisfies:

$$\alpha^\star = \left(\frac{\sigma^2}{2\theta} - \text{mmse}\right)\frac{2\theta}{\mathbb{E}[e^{-2\theta D}]} \ge 0. \tag{8}$$

The proof of Lemma 1 is provided in Appendix B.

Assumption 1: The expectation of delay D_k is bounded and known to the transmitter, i.e.,

$$0 < D_{\text{lb}} \le \overline{D} \triangleq \mathbb{E}_{\mathbb{P}_D}[D_k] \le D_{\text{ub}} < \infty. \tag{9}$$

Lemma 2: Define $\hat{W} = 1/f_{\max} + c$, where $c > 0$ is an arbitrary constant. If Assumption 1 is satisfied, then we can bound $\alpha^\star$ as

$$\alpha_{\text{lb}} \le \alpha^\star \le \alpha_{\text{ub}}, \tag{10}$$

where α_{lb} and α_{ub} can be chosen as

$$\alpha_{\text{lb}} = \frac{\sigma^2(1 - e^{-2\theta\hat{W}})}{2\theta(D_{\text{ub}} + \hat{W})} > 0, \tag{11}$$

$$\alpha_{\text{ub}} = \sigma^2. \tag{12}$$

The proof of Lemma 2 is provided in Appendix C. The lower bound is obtained by constructing a feasible and constant sampling policy whose waiting time is always $\hat{W}$ and then using (6a). The constant c is introduced to ensure $\hat{W} > 0$ when there is no frequency constraint. The upper bound is obtained by using (8) and the fact $\text{mmse} \ge \sigma^2/2\theta\mathbb{E}[1-e^{-2\theta D}]$.

B. Optimal Sampling with Known $\mathbb{P}_D$

In the sequel, we will derive the optimum policy $\pi^\star$ that achieves optimal mmse when $\mathbb{P}_D$ is known. The structure of the optimal policy can help us design the algorithm under unknown channel statistics, and the average MSE obtained by $\pi^\star$ will be used to measure the performance of the proposed online learning algorithm in Section III-C

According to (6a), the cost obtained by any policy π that satisfies the sampling constraint (6b) is less or equal to $\alpha^\star$. In other words, we have

$$-\lim_{K\to\infty}\frac{\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[O^2_{D_k+W_k}]}{\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}\geq-\alpha^\star. \tag{13}$$

Multiplying $(1/K)\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]$ on both sides of (13) and then adding $\alpha^\star\lim_{K\to\infty}(1/K)\mathbb{E}[D_k+W_k]$ on both sides, we are able to solve Problem 2 by minimizing the following objective function:

Problem 3:

$$\rho^\star\triangleq\inf_{\pi\in\Pi_r}\limsup_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\big(-\mathbb{E}[O^2_{D_k+W_k}]+\alpha^\star\mathbb{E}[D_k+W_k]\big), \tag{14a}$$

$$\text{s.t. }\liminf_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[W_k+D_k\right]\geq\frac{1}{f_{\max}}, \tag{14b}$$

Similar to Dinkelbach's method [37] for the non-linear fractional programming, we can deduce that the optimal value $\rho^\star$ of Problem 3 equals 0, and the optimum policy that achieves mmse in Problem 1 and $\rho^\star$ in Problem 3 are identical. Therefore, we proceed to solve Problem 3 using the Lagrange multiplier approach. Let $\lambda\geq0$ be the Lagrange multiplier of the sampling frequency constraint (14b), the Lagrange function for Problem 3 is as follows:

$$\mathcal{L}(\pi,\lambda)=\limsup_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\big(-\mathbb{E}[O^2_{D_k+W_k}]+(\alpha^\star-\lambda)\,\mathbb{E}\left[D_k+W_k\right]+\lambda\frac{1}{f_{\max}}\big). \tag{15}$$

Notice that the transmission delay D_k is i.i.d., and O_t is an OU process starting at time $t=0$. Then for fixed λ, selecting the optimum waiting time W_k to minimize (15) becomes a per-sample optimal stopping problem by finding the optimum stop time w to minimize the following expectation:

$$\min_{w}\mathbb{E}\left[-O^2_{D_k+w}+(\alpha^\star-\lambda)w|O_{D_k},D_k\right]. \tag{16}$$

For simplicity, let $V_w=O_{D_k+w}$ be the value of the OU process at time D_k+w and $V_0=O_{D_k}$ by definition. Then problem (16) is one instance of the following optimal stopping problem when $\beta=\alpha^\star-\lambda$:

$$\sup_{\tau}\mathbb{E}_{v_0}\left[V_\tau^2-\beta\tau\right], \tag{17}$$

where $\mathbb{E}_{v_0}$ is the conditional expectation given $V_0=v_0$. The optimum policy to (17) is obtained in the following Lemma:

Lemma 3: If $0<\beta\leq\sigma^2$, then the solution to minimize (17) has a threshold property, i.e.,

$$W_k=w(O_{D_k};\beta):=\inf\{t\geq0:|O_{D_k+t}|\geq v(\beta)\}, \tag{18}$$

where

$$v(\beta)=\frac{\sigma}{\sqrt{\theta}}G^{-1}\left(\frac{\sigma^2}{\beta}\right), \tag{19}$$

and $G^{-1}(\cdot)$ is the inverse function of

$$G(x)=\frac{e^{x^2}}{x}\int_0^x e^{-t^2}\mathrm{d}t,\ x\in[0,\infty). \tag{20}$$

The proof of Lemma 3 is provided in Appendix D.

Since [21, Theorem 6] has proven the strong duality of Problem 3, i.e., $\rho^\star=\max_\lambda\min\mathcal{L}(\pi,\lambda)$. For notational simplicity, let $o(\beta)$ and $l(\beta)$ denote the expected estimation error and frame length by using threshold β, i.e.,

$$o(\beta):=\mathbb{E}[O^2_{D+w(O_D;\beta)}], \tag{21a}$$

$$l(\beta):=\mathbb{E}[D+w(O_D;\beta)]. \tag{21b}$$

by substituting O_{D_k+w} with $(X_{R_k+w}-\hat{X}_{R_k+w})$ in (18), the optimal sampling time $S_{k+1}=R_k+W_k$ to Problem 3 is as follows:

Lemma 4: [21, Theorem 2 Restated] The optimal solution to Problem 1 is:

$$S_{k+1}=\inf\{t\geq R_k:|X_t-\hat{X}_t|\geq v(\alpha^\star-\lambda^\star)\},$$

where $v(\cdot)$ is defined in (19), $\lambda^\star=\arg\sup_\lambda\mathcal{L}(\pi,\lambda)$ is the dual optimizer, and $\alpha^\star$ is the solution to the following equation:

$$0=g_{\lambda^\star}(\alpha):=o(\alpha-\lambda^\star)-\alpha l(\alpha-\lambda^\star), \tag{22}$$

where we recall that $o(\beta)=\mathbb{E}[O^2_{D+w(O_D;\beta)}]=\mathbb{E}[(X_{S_{k+1}}-\hat{X}_{S_{k+1}})^2]$ is the expected squared estimation error by using threshold β, and $l(\beta)=\mathbb{E}[D+w(O_D;\beta)]$ is the expected framelength.

Remark 1: If the frequency constraint is inactive, then according to the complementary slackness, we have $\lambda^\star=0$, and the threshold becomes $v(\alpha^\star)$. Otherwise, the optimal $\alpha^\star-\lambda^\star<\alpha^\star$. Then according to (19), the sampling threshold is larger than $v(\alpha^\star)$ to satisfy the sampling frequency constraint.

Remark 2: In [21, Theorem 2], the optimum sampling threshold to minimize the MSE is

$$v(\beta')=\frac{\sigma}{\sqrt{\theta}}G^{-1}\left(\frac{\mathrm{mse}_\infty-\mathrm{mse}_D}{\mathrm{mse}_\infty-\beta'}\right), \tag{23}$$

where

$$\mathrm{mse}_\infty=\mathbb{E}[O^2_\infty]=\frac{\sigma^2}{2\theta}; \tag{24a}$$

$$\mathrm{mse}_D=\mathbb{E}[O^2_{D_k}]=\frac{\sigma^2}{2\theta}\mathbb{E}[1-e^{-2\theta D}]. \tag{24b}$$

The optimum sampling threshold is taken when $\beta'=\mathrm{mmse}+\lambda'$, i.e.,

$$v(\beta')=\frac{\sigma}{\sqrt{\theta}}G^{-1}\left(\frac{\sigma^2}{\left(\frac{\sigma^2}{2\theta}-\mathrm{mmse}\right)\frac{2\theta}{\mathbb{E}[e^{-2\theta D}]}-\lambda'\frac{2\theta}{\mathbb{E}[e^{-2\theta D}]}}\right)$$

Algorithm 1 Online learning sampling algorithm

1: **Parameters**: V.
2: **Initialization**: $\alpha_1 = 0,\ U_1 = 0$.
3: **for** $k = 1, 2, \cdots, K$ **do**
4: Set $\lambda_k = \frac{1}{V} U_k$.
5: According to the last sampling generation time S_k and delay D_k, choose the waiting time W_k as

$$W_k = \inf\{w \geq 0 : |X_{R_k+w} - \hat{X}_{R_k+w}| \geq v((\alpha_k - \lambda_k)^+)\}.$$

6: Update α_k:

$$\alpha_{k+1} = (\alpha_k + \eta_k(O_{L_k}^2 - \alpha_k L_k))_{\alpha_{\text{lb}}}^{\alpha_{\text{ub}}},$$

where

$$O_{L_k} = X_{S_{k+1}} - \hat{X}_{S_{k+1}}, \tag{26}$$

$$L_k = D_k + W_k. \tag{27}$$

7: Update U_k:

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

8: **end for**

$$\overset{(a)}{=} \frac{\sigma}{\sqrt{\theta}} G^{-1}\left(\frac{\sigma^2}{\alpha^\star - \lambda' \frac{2\theta}{\mathbb{E}[e^{-2\theta D}]}}\right), \tag{25}$$

where (a) holds by (8). Comparing (25) with (19), we find the conclusions coincide.

C. Online Algorithm

Notice that the optimal sampling in Section III-B is determined by $\alpha^\star - \lambda^\star$ through (19). However, when the channel statistics $\mathbb{P}_D$ is unknown, $\alpha^\star$ and $\lambda^\star$ are unknown, making direct computation of $v(\alpha^\star - \lambda^\star)$ impossible. To overcome the challenge, we propose an online learning algorithm to approximate these two parameters $\alpha^\star$ and $\lambda^\star$ respectively.

Notice that $\alpha^\star$ is the solution to (22) when $\lambda = \lambda^\star$. This motivates us to approximate $\alpha^\star$ using the Robbins-Monro algorithm [38] for stochastic approximation. For $\lambda^\star$, we construct a virtual queue U_k to record the cumulative sampling constraint violation up to frame k.

As concluded in Algorithm 1, the proposed algorithm consists of two parts: Sampling (step 5) and updating (step 6 and 7). For the sampling step, the algorithm uses the current estimation α_k and λ_k to compute the threshold, i.e.,

$$W_k = \inf\{w \geq 0 : |X_{R_k+w} - \hat{X}_{R_k+w}| \geq v((\alpha_k - \lambda_k)^+)\}, \tag{28}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. After sample $(k+1)$ is taken at time $R_k + W_k$, we can compute the instant estimation error $O_{L_k} := X_{S_{k+1}} - \hat{X}_{S_{k+1}}$ and the frame length $L_k := D_k + W_k$. According to (4), O_{L_k} is an instance of $O_{D+w(O_D;\alpha-\lambda)}$ when $\lambda = \lambda_k$ and $\alpha = \alpha_k$.

We then update α_{k+1} according to the Robbins-Monro algorithm:

$$\alpha_{k+1} = (\alpha_k + \eta_k(O_{L_k}^2 - \alpha_k L_k))_{\alpha_{\text{lb}}}^{\alpha_{\text{ub}}}, \tag{29}$$

where $(x)_a^b$ is the projection of x onto the interval $[a, b]$; α_{lb} and α_{ub} are the lower and upper bound of $\alpha^\star$ defined in (11) and (12); η_k is the step size, which can be chosen as

$$\eta_k = \begin{cases} \frac{1}{2D_{\text{lb}}}, & k = 1; \\ \frac{1}{(k+2)D_{\text{lb}}}, & k \geq 2. \end{cases}$$

For estimating $\lambda^\star$, we construct a virtual queue U_k which evolves as

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

Then $\lambda_k = U_k/V$, where $V > 0$ is the hyper-parameter. Notice that $1/f_{\max} - L_k$ is the violation of sampling constraint in frame k. Therefore U_k can be interpreted as the cumulative violation up to frame k. The Algorithm 1 attempts to stabilize U_k to satisfy the sampling frequency constraint.

Remark 3: In (28), we choose $(\alpha_k - \lambda_k)^+$ to ensure the positive input for $v(\cdot)$. We should also avoid the estimation $\alpha_k - \lambda_k$ to be zero, which will make the threshold v to be infinite. This requires the algorithm cannot choose V to be too small. Also in practice one can set an arbitrarily small positive value $\eta > 0$ as a lower bound for $\alpha_k - \lambda_k$ to avoid the infinite threshold.

IV. Theoretical Analysis

In this section, we analyze the convergence and optimality of Algorithm 1.

Assumption 2: The second moment of delay D_k is bounded, i.e., [1]

$$0 < M_{\text{lb}} \leq \mathbb{E}_{\mathbb{P}_D}[D_k^2] \leq M_{\text{ub}} < \infty. \tag{30a}$$

First, we assume that there is no sampling frequency constraint, i.e., $f_{\max} = \infty$ and thus $\lambda = 0$. Finally, we will prove that in general case $f_{\max} < \infty$, Algorithm 1 will still satisfy the constraint.

Theorem 1: The time average MSE $\frac{\int_0^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t}{S_{k+1}}$ of the proposed online learning algorithm converges to mmse with probability 1, i.e.,

$$\frac{\int_0^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t}{S_{k+1}} \overset{\text{a.s.}}{=} \text{mmse}. \tag{31}$$

Theorem 2: Let $\mathcal{R}_k := \mathbb{E}\left[\int_0^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] - \text{mmse} \cdot \mathbb{E}[S_{k+1}]$ denote the expected cumulative MSE regret up to the $(k+1)$th sample. We can upper bound $\mathcal{R}_k$ as follows:

$$\mathcal{R}_k \leq \max_{\alpha \in [\alpha_{\text{lb}}, \alpha_{\text{ub}}]} |R_1'(v(\alpha)) v'(\alpha)| \frac{\mathbb{E}[e^{-2\theta D}]}{2\theta} \frac{C}{D_{\text{lb}}^2} \ln k, \tag{32}$$

where C is a constant independent of k and is defined (42).

The proof of Theorem 1 and Theorem 2 are provided in Appendix E and Appendix F, respectively.

Now we consider the sampling frequency constraint. Here we assume that the constraint is feasible, i.e.,

[1]The assumptions is presented here mainly for theoretical analysis. In fact the proposed algorithm discussed in Section III-C does not need the assumption.

Assumption 3: There exists a constant $\epsilon > 0$, and a stationary sampling policy π_ϵ satisfies

$$\mathbb{E}\left[D_k + W_k^\epsilon\right] \geq \frac{1}{f_{\max}} + \epsilon, \tag{33}$$

where the expectation is taken over the channel statistics and the policy π_ϵ.

Theorem 3: Under Algorithm 1, the sampling frequency constraint can be satisfied, i.e.,

$$\lim_{K\to\infty} \inf \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}(D_k + W_k)\right] \geq \frac{1}{f_{\max}}. \tag{34}$$

The proof of Theorem 3 is provided in Appendix H.

V. Simulation Results

In this section, we provide some simulation results to demonstrate the performance of our proposed algorithm. The parameters of the monitored OU process are $\sigma = 1, \theta = 0.2$, and $\mu = 3$. The channel delay follows the log-normal distribution with $\mu_D = \sigma_D = 1$. The expected MSE is computed by taking the average of 100 simulation runs for $K = 10^4$ packet transmission frames.

A. Without a Sampling Frequency Constraint

First, we consider the case with no frequency constraint, i.e., $f_{\max} = \infty$. We compare the MSE performance using the following policies:

- **Zero-Wait Policy** π_{zw}: Take a new sample immediately after the reception of the ACK of the last sample, i.e., $W_k = 0$.
- **Signal-Aware MSE Optimum Policy** $\pi^\star$: Signal aware MSE optimum policy when $\mathbb{P}_D$ is known [21].
- **Signal-Agnostic AoI Minimum Policy** π_{AoI}: Signal agnostic sampling policy for AoI minimization [6].
- **Proposed Online Policy** π_{online}: Described in Algorithm 1.

The estimation performance is depicted in Fig. 3. From Fig. 3, we can verify that the expected MSE performance of the proposed policy π_{online} converges to the optimum policy $\pi^\star$, and achieves a smaller MSE performance compared with the signal-agnostic AoI minimum sampling and zero-wait policy. Previous work [21] has shown that the zero-wait policy is far from optimality when the channel delay is heavy tail. For the AoI optimal policy, while [20] reveals the relationship between average AoI and estimation error for the Wiener process, it is sub-optimal for MSE optimization of the OU process, even worse than the zero-wait policy.

Next, we consider the estimation of the threshold $v(\alpha^\star - \lambda^\star)$. Obviously, the fast and accurate estimation of the threshold is the necessary condition for the convergence of MSE performance. As depicted in Fig. 4, the proposed algorithm can approximate the optimal threshold as the time goes to infinity. Besides, the variance of the threshold estimation will also become small, which guarantees the convergence of MSE.

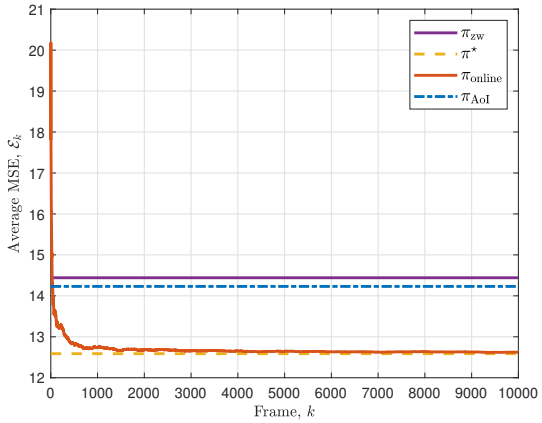


Fig. 3. MSE performance with no frequency constraint.

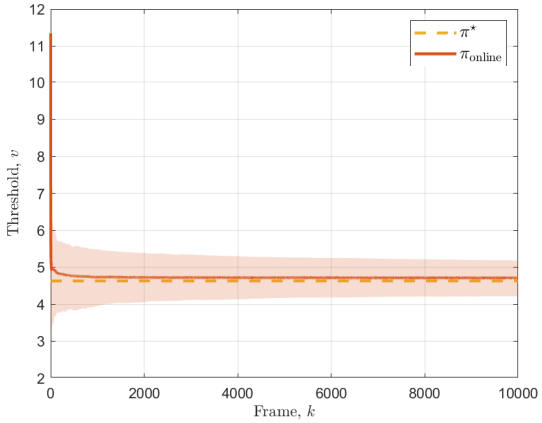


Fig. 4. Threshold evolution without frequency constraint.

B. With A Sampling Frequency Constraint

In this part, we depict the simulation results when a sampling constraint exists. The parameters of the system are the same as in Fig. 3, and we set $f_{\max} = 0.02$. In other words, the minimum average frame length $1/f_{\max} = 50$. Notice that now the zero-wait policy does not satisfy the sampling constraint. Therefore, we consider a frequency conservative policy π_{freq}, which selects W_k as

$$W_k = \max\left\{\frac{k}{f_{\max}} - \sum_{k'=1}^{k-1} L_{k'} - D_k, 0\right\}.$$

We set the parameter $V = 500$ and depict the MSE performance and average frame length in Fig. 5 and Fig. 6. These two figures verify that the proposed algorithm can also approximate the lower bound while satisfying the frequency constraint.

Finally, we investigate the impact of V on the MSE performance and average frame length. We choose three different values of $V = \{300, 500, 800\}$ and compare the MSE performance and average frame length, as depicted in Figs. 7(a) and 7(b) respectively. Generally speaking, the MSE performance of proposed algorithm with different V can all

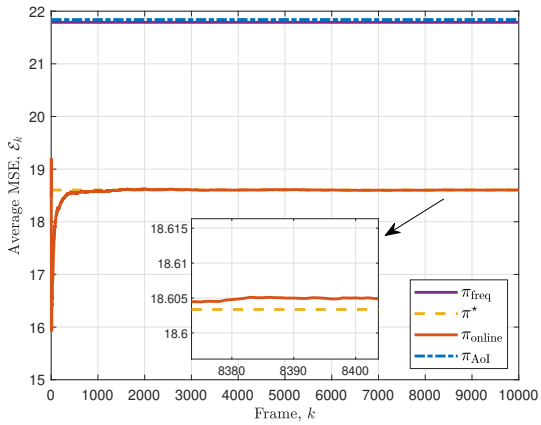


Fig. 5. MSE performance under frequency constraint $f_{\max} = 0.02$.

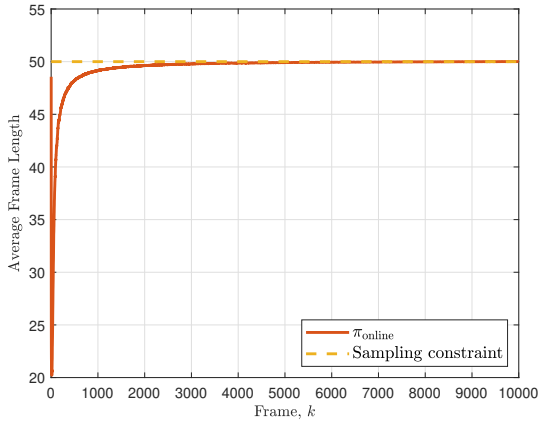
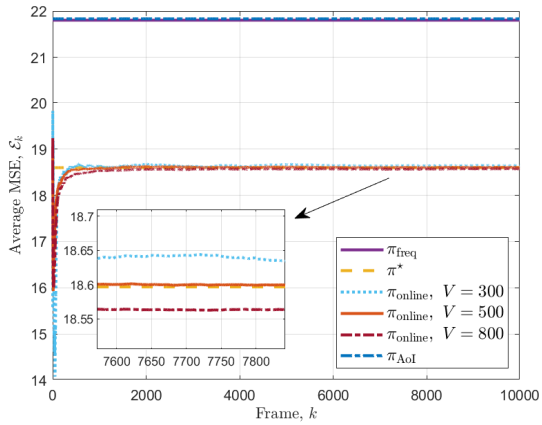
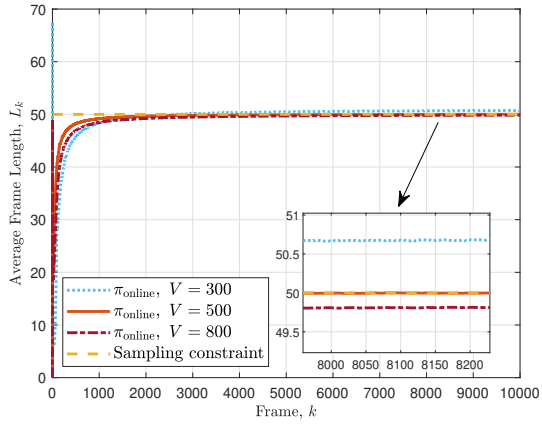


Fig. 6. Average frame length under frequency constraint $f_{\max} = 0.02$.



(a) MSE performance.



(b) Average frame length.

Fig. 7. MSE performance and average frame length with different parameter V.

converge to the optimal MMSE, and the average inter-update interval of the proposed algorithms are near the frequency constraint. Notice that V is a hyper parameter controlling the estimation of the Lagrange multiplier. A larger V indicates less emphasis on the frequency constraint. By using a larger $V = 800$, the algorithm will take a longer time to converge to the sampling frequency constraint. Since for $t < 8000$ the sampling frequency of the algorithm slightly violates the sampling frequency constraint, the MSE is smaller.

VI. Conclusion

In this work, we studied the sampling policy for remote estimation of an OU process through a channel with transmission delay. We aim at designing an online sampling policy that can minimize the mean square error when the delay distribution is unknown. Finding the MSE minimum sampling policy can be reformulated into an optimal stopping problem, we proposed a stochastic approximation algorithm to learn the optimum stopping threshold adaptively. We prove that, after taking k samples, the cumulative MSE regret of our proposed algorithm grows with rate $\mathcal{O}(\ln k)$, and the expected time-averaged MSE of our proposed algorithm converges to the minimum MSE almost surely. Numerical simulation validates the superiority and convergence performance of the proposed algorithm.

References

[1] M. N. Ahangar, Q. Z. Ahmed, F. A. Khan, and M. Hafeez, "A survey of autonomous vehicles: Enabling communication technologies and challenges," *Sensors*, vol. 21, no. 3, p. 706, 2021.

[2] S. Chen, R. Ma, H.-H. Chen, H. Zhang, W. Meng, and J. Liu, "Machine-to-machine communications in ultra-dense networks-A survey," *IEEE Commun. Surveys Tut.*, vol. 19, no. 3, pp. 1478–1503, 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, 2021.

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

[5] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE ISIT*, 2015.

[6] 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, 2017.

[7] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

[8] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[9] I. Kadota, A. Sinha, and E. Modiano, "Scheduling algorithms for optimizing age of information in wireless networks with throughput constraints," *IEEE/ACM Trans. Netw.*, vol. 27, no. 4, pp. 1359–1372, 2019.
[10] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 15–28, 2020.
[11] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Trans. Mobile Comput.*, vol. 20, no. 3, pp. 1173–1185, 2021.
[12] 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, 2020.
[13] V. S. Jog, R. J. La, and N. C. Martins, "Channels, learning, queueing and remote estimation systems with A utilization-dependent component," *CoRR*, vol. abs/1905.04362, 2019. [Online]. Available: http://arxiv.org/abs/1905.04362
[14] M. Rabi, G. V. Moustakides, and J. S. Baras, "Adaptive sampling for linear state estimation," *SIAM J. Control. Optim.*, vol. 50, no. 2, pp. 672–702, 2012.
[15] K. Nar and T. Başar, "Sampling multidimensional wiener processes," in *Proc. IEEE CDC*, 2014.
[16] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order lti systems," *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2013–2025, 2011.
[17] X. Gao, E. Akyol, and T. Başar, "Optimal communication scheduling and remote estimation over an additive noise channel," *Automatica*, vol. 88, pp. 57–69, 2018.
[18] J. Chakravorty and A. Mahajan, "Remote estimation over a packet-drop channel with markovian state," *IEEE Trans. Auto. Control*, vol. 65, no. 5, pp. 2016–2031, 2020.
[19] C.-H. Tsai and C.-C. Wang, "Unifying AoI minimization and remote estimation-optimal sensor/controller coordination with random two-way delay," *IEEE/ACM Trans. Netw.*, vol. 30, no. 1, pp. 229–242, 2022.
[20] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the wiener process for remote estimation over a channel with random delay," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, 2020.
[21] T. Z. Ornee and Y. Sun, "Sampling and remote estimation for the ornstein-uhlenbeck process through queues: Age of information and beyond," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 1962–1975, 2021.
[22] S. Banerjee, R. Bhattacharjee, and A. Sinha, "Fundamental limits of age-of-information in stationary and non-stationary environments," in *Proc. IEEE ISIT*, 2020.
[23] E. U. Atay, I. Kadota, and E. H. Modiano, "Aging wireless bandits: Regret analysis and order-optimal learning algorithm," in *Proc. IEEE WiOpt*, J. Ghaderi, E. Uysal, and G. Xue, Eds., 2021.
[24] S. Fatale, K. Bhandari, U. Narula, S. Moharir, and M. K. Hanawal, "Regret of age-of-information bandits," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 87–100, 2022.
[25] B. Li, "Efficient learning-based scheduling for information freshness in wireless networks," in *Proc. IEEE INFOCOM*, 2021.
[26] V. Tripathi and E. H. Modiano, "An online learning approach to optimizing time-varying costs of AoI," in *Proc. ACM MobiHoc*, 2021.
[27] H. Tang, Y. Chen, J. Wang, P. Yang, and L. Tassiulas, "Age optimal sampling under unknown delay statistics," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 1295–1314, 2023.
[28] C.-H. Tsai and C.-C. Wang, "Age-of-information revisited: Two-way delay and distribution-oblivious online algorithm," *Proc. IEEE ISIT*, 2020.
[29] C.-H. Tsai and C.-C. Wang, "Distribution-oblivious online algorithms for age-of-information penalty minimization," *IEEE/ACM Trans. Netw.*, pp. 1–16, 2023.
[30] S. Leng and A. Yener, "Age of information minimization for wireless ad hoc networks: A deep reinforcement learning approach," in *Proc. IEEE GLOBECOM*, 2019.
[31] 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, 2020.
[32] E. T. Ceran, D. Gündüz, and A. György, "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, 2021.
[33] 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*, 2022.
[34] H. Tang, Y. Sun, and L. Tassiulas, "Sampling of the wiener process for remote estimation over a channel with unknown delay statistics," in *Proc. ACM MobiHoc*, 2022.
[35] J. L. Doob, "The brownian movement and stochastic equations," *Annals of Mathematics*, pp. 351–369, 1942.
[36] H. V. Poor, *An Introduction to Signal Detection and Estimation*, ser. Springer Texts in Electrical Engineering. Springer, 1994. [Online]. Available: https://doi.org/10.1007/978-1-4757-2341-0
[37] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.
[38] H. Robbins and S. Monro, "A stochastic approximation method," *The annals of mathematical statistics*, pp. 400–407, 1951.
[39] M. J. Neely, "Fast learning for renewal optimization in online task scheduling," *J. Machine Learning Research*, vol. 22, no. 279, pp. 1–44, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-813.html
[40] S. M. Ross, *Applied probability models with optimization applications*. Courier Corporation, 2013.
[41] G. Peskir and A. Shiryaev, *Optimal stopping and free-boundary problems*. Springer, 2006.
[42] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY: Springer New York, 2003.
[43] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, ser. Synthesis Lectures on Communication Networks. Morgan & Claypool Publishers, 2010. [Online]. Available: https://doi.org/10.2200/S00271ED1V01Y201006CNT007
[44] D. A. Darling and A. J. F. Siegert, "The First Passage Problem for a Continuous Markov Process," *The Annals of Mathematical Statistics*, vol. 24, no. 4, pp. 624–639, 1953.

APPENDIX A
LEMMAS AND NOTATIONS

First, we state the auxiliary lemmas and corollaries that will be used in the following proofs. Proofs for these lemmas and corollaries are provided in

Lemma 5: [21, Lemma 1 Restated]

$$\begin{aligned}&\mathbb{E}[D_k+W_k]\\&=\mathbb{E}[D_k]+\mathbb{E}[\max\{R_1(v((\alpha_k-\lambda_k)^+))-R_1(|O_{D_k}|),0\}],\end{aligned}\tag{35}$$

where

$$R_1(v)=\frac{v^2}{\sigma^2}{}_2F_2\left(1,1;\frac{3}{2},2;\frac{\theta}{\sigma^2}v^2\right),\tag{36a}$$

$${}_2F_2\left(1,1;\frac{3}{2},2;z\right)=\sum_{n=0}^{\infty}\frac{2^n n!n!}{(n+1)!(2n+1)!!}\frac{z^n}{n!}.\tag{36b}$$

Moreover, since $R_1(\cdot)$ is a monotonically increasing function, $v(\beta)=\sigma/\sqrt{\theta}G^{-1}\left(\sigma^2/\beta\right)$ and $G(x)=e^{x^2}/x\int_0^x e^{-t^2}\mathrm{d}t$ is monotonic increasing, we have $R_1(v(\alpha))$ is monotically decreasing.

Corollary 1: Recall that function $l(\beta)=\mathbb{E}[D+w(O_D;\beta)]$ is the expected framelength when using sampling threshold $v(\beta)$. When there is no sampling frequency constraint and $\lambda=0$, function $l(\alpha)$ has the following property:

$$|l(\alpha)-l(\alpha^\star)|\le N|\alpha-\alpha^\star|,\tag{37}$$

where $N=\max_{\alpha\in[\alpha_{\mathrm{lb}},\alpha_{\mathrm{ub}}]}|R_1'(v(\alpha))v'(\alpha)|$ is a constant independent of α.

The proof is provided in Appendix I-A

Lemma 6: Recall that $\mathbb{E}[D]\le D_{\mathrm{ub}}$ and $\mathbb{E}[D^2]\le M_{\mathrm{ub}}$ and α_k is truncated into interval $[\alpha_{\mathrm{lb}},\alpha_{\mathrm{ub}}]$ using Lemma 2, when

there is no sampling frequency constraint and $\lambda_k \equiv 0$, we have the following bounds for each frame k:

$$0 \le \mathbb{E}[O_{L_k}^2] < \frac{\sigma^2}{2\theta}; \tag{38a}$$

$$0 \le \mathbb{E}[O_{L_k}^4] < \frac{3\sigma^4}{4\theta^2}; \tag{38b}$$

$$0 \le \mathbb{E}[L_k] \le D_{\text{ub}} + \frac{v(\alpha_{\text{lb}})^2}{\sigma^2} e^{\frac{2\theta}{\sigma^2} v(\alpha_{\text{lb}})^2} \triangleq L_{\text{ub}}; \tag{38c}$$

$$0 \le \mathbb{E}[L_k^2] \le M_{\text{ub}} + 2D_{\text{ub}} \frac{v(\alpha_{\text{lb}})^2}{\sigma^2} e^{\frac{2\theta}{\sigma^2} v(\alpha_{\text{lb}})^2} + \frac{2v(\alpha_{\text{lb}})^3}{\sigma^3} \sqrt{\frac{\pi}{\theta}} e^{\frac{3\theta}{\sigma^2} v(\alpha_{\text{lb}})^2} \triangleq L_{\text{ub2}}. \tag{38d}$$

The proof of Lemma 6 is provided in Appendix I-B.

Lemma 7: For fixed λ, function $g_\lambda(\alpha) = o(\alpha - \lambda) - \alpha l(\alpha - \lambda)$ is continuous, monotonically decreasing and convex. Moreover, there exists a constant N so that function $g_0(\alpha)$

$$g_0(\alpha) \ge -l(\alpha^\star)(\alpha - \alpha^\star) + N(\alpha - \alpha^\star)^2, \tag{39a}$$

$$|g_0(\alpha)| \le l(\alpha^\star)|\alpha - \alpha^\star|. \tag{39b}$$

Proof for Lemma 7 is provided in Appendix I-C.

Theorem 4: The estimation α_k computed in Algorithm 1 can converge to $\alpha^\star$ with probability 1, and we have

$$\mathbb{E}[(\alpha_k - \alpha^\star)^2] \le \frac{C}{kD_{\text{lb}}^2} \sim \mathcal{O}\left(\frac{1}{k}\right), \tag{40}$$

where C is a constant independent of k, i.e.,

$$C = \frac{3\sigma^4}{4\theta^2} + \alpha_{\text{ub}}^2 \big(M_{\text{ub}} + 2D_{\text{ub}} \frac{v(\alpha_{\text{lb}})^2}{\sigma^2} e^{\frac{2\theta}{\sigma^2} v(\alpha_{\text{lb}})^2} \tag{41}$$

$$+ \frac{2v(\alpha_{\text{lb}})^3}{\sigma^3} \sqrt{\frac{\pi}{\theta}} e^{\frac{3\theta}{\sigma^2} v(\alpha_{\text{lb}})^2} \big). \tag{42}$$

The proof of Theorem 4 is the same as [39, Lemma 6].

APPENDIX B
PROOF OF LEMMA 1

The ultimate goal is to rewrite the averaged MMSE (5a) obtained by a stationary policy as the time-averaged cost of each frame. The waiting time W_k set by any stationary policy π can be viewed as a stopping time. The information, i.e., tuple $\{(D_k, \Delta_{S_{k+1}})\}$ is a regenerative sequence as the instant estimation error $\Delta_t, t \ge S_k + D_k$ is an OU process starting from time $t = S_k$. Therefore, for stationary policy, the cumulative estimation error in frame k, i.e., $E_k := \int_{S_k}^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t$ and $L_k := S_{k+1} - S_k$ are generative random processes. Then according the renewal-reward theory [40], both the average cumulative MSE in each frame $\{\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^K E_k\right]\}$ and the average frame-length $\{\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^K L_k\right]\}$ have limits. Then according to the renewal reward theory [40], the time averaged MMSE can be computed by:

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_{t=0}^{T} \left(X_t - \hat{X}_t\right)^2 \mathrm{d}t\right] = \limsup_{K\to\infty} \frac{\sum_{k=1}^K \mathbb{E}\left[\int_{S_k}^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t\right]}{\sum_{k=1}^K \mathbb{E}\left[(S_{k+1} - S_k)\right]}. \tag{43}$$

Then to compute the average cost in each frame k, we introduce the following properties of the stopping time of an OU process:

Lemma 8 (Lemma 5, [21] Restated): Let O_t be an OU process with initial state zero and parameter $\mu = 0$, and τ is a stopping time with $\mathbb{E}[\tau] < \infty$, the integral of O_t^2 from 0 to t can be computed by

$$\mathbb{E}\left[\int_0^\tau O_t^2 \mathrm{d}t\right] = \mathbb{E}\left[\frac{\sigma^2}{2\theta}\tau - \frac{1}{2\theta}O_\tau^2\right]. \tag{44}$$

We then proceed to compute the expected cumulative error of stationary policy π using Lemma 8. Notice that the interval $[S_k, S_{k+1})$ can then be divided into two intervals $[S_k, S_k + D_k)$ and $[S_k + D_k, S_k + D_k + W_k)$. The cumulative estimation error during $[S_k, S_k + D_k)$ can be computed as follows:

$$\begin{aligned}
&\mathbb{E}\left[\int_{S_k}^{S_k+D_k} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] \\
&= \mathbb{E}\left[\int_{S_{k-1}}^{S_{k-1}+D_{k-1}+W_{k-1}+D_k} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] \\
&\quad - \mathbb{E}\left[\int_{S_{k-1}}^{S_{k-1}+D_{k-1}+W_{k-1}} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1} + W_{k-1} + D_k) - \frac{1}{2\theta}O^2_{D_{k-1}+W_{k-1}+D_k}\right] \\
&\quad - \mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1} + W_{k-1}) - \frac{1}{2\theta}O^2_{D_{k-1}+W_{k-1}}\right],
\end{aligned} \tag{45}$$

where (a) is because during interval $[S_k, S_k + D_k)$, the instant $X_t - \hat{X}_t$ from (4) is equivalent to an OU process starting from time $t = S_{k-1}$, and the cumulative MSE can be computed by Lemma 8. Notice that the delay distribution D_k is independent of $O_{D_{k-1}+W_{k-1}}$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[O^2_{D_{k-1}+W_{k-1}+D_k}\right] \\
&= \mathbb{E}\left[\left(O_{D_{k-1}+W_{k-1}} e^{-\theta D_k} + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta D_k} W_{e^{2\theta D_k}-1}\right)^2\right] \\
&= \mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}] + \frac{\sigma^2}{2\theta}\mathbb{E}\left[1 - e^{-2\theta D_k}\right].
\end{aligned} \tag{46}$$

Plugging (46) into (45), we have:

$$\begin{aligned}
&\mathbb{E}\left[\int_{S_k}^{S_k+D_k} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] \\
&= \mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1} + W_{k-1} + D_k)\right] \\
&\quad - \frac{1}{2\theta}\mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}] - \frac{\sigma^2}{4\theta^2}\mathbb{E}\left[1 - e^{-2\theta D_k}\right] \\
&\quad - \mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1} + W_{k-1}) - \frac{1}{2\theta}O^2_{D_{k-1}+W_{k-1}}\right],
\end{aligned} \tag{47}$$

Similarly, the second part of the cumulative MSE, i.e., the cumulative MSE during interval $[S_k + D_k, S_k + D_k + W_k)$ can be computed by

$$\mathbb{E}\left[\int_{S_k+D_k}^{S_k+D_k+W_k} (X_t - \hat{X}_t)^2 \mathrm{d}t\right]$$

$$
\begin{aligned}
&=\mathbb{E}\left[\int_{S_k}^{S_k+D_k+W_k}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\
&\quad-\mathbb{E}\left[\int_{S_k}^{S_k+D_k}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\
&\overset{(b)}{=}\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_k+W_k\right)-\frac{1}{2\theta}O^2_{D_k+W_k}\right]\\
&\quad-\mathbb{E}\left[\frac{\sigma^2}{2\theta}D_k-\frac{1}{2\theta}O^2_{D_k}\right],
\end{aligned}
\tag{48}
$$

where (b) is obtained because the instant estimation error $X_t-\hat{X}_t, t\geq S_k+D_k$ is an OU process starting at time S_k according to (4).

By summing up (47) and (48), we are able to compute the expected cumulative error for stationary policy π:

$$
\begin{aligned}
&\mathbb{E}\left[E_k\right]=\mathbb{E}\left[\int_{S_k}^{S_{k+1}}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\
&=\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_{k-1}+W_{k-1}+D_k\right)\right]\\
&\quad-\frac{1}{2\theta}\mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}]-\frac{\sigma^2}{4\theta^2}\mathbb{E}\left[1-e^{-2\theta D_k}\right]\\
&\quad-\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_{k-1}+W_{k-1}\right)-\frac{1}{2\theta}O^2_{D_{k-1}+W_{k-1}}\right]\\
&\quad+\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_k+W_k\right)-\frac{1}{2\theta}O^2_{D_k+W_k}\right]\\
&\quad-\mathbb{E}\left[\frac{\sigma^2}{2\theta}D_k-\frac{1}{2\theta}O^2_{D_k}\right]\\
&\overset{(c)}{=}\mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1}+W_{k-1})\right]+\frac{1}{2\theta}\mathbb{E}\left[O^2_{D_k}\right]\\
&\quad-\frac{1}{2\theta}\mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}]-\frac{\sigma^2}{4\theta^2}\mathbb{E}\left[1-e^{-2\theta D_k}\right]\\
&\overset{(d)}{=}\mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1}+W_{k-1})\right]-\frac{1}{2\theta}\mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}],
\end{aligned}
\tag{49}
$$

where equality (c) is obtained because the transmission delay D_k is i.i.d., and therefore

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_{k-1}+W_{k-1}\right)-\frac{1}{2\theta}O^2_{D_{k-1}+W_{k-1}}\right]\\
=&\mathbb{E}\left[\frac{\sigma^2}{2\theta}\left(D_k+W_k\right)-\frac{1}{2\theta}O^2_{D_k+W_k}\right].
\end{aligned}
\tag{50}
$$

and equality (d) is because:

$$
\mathbb{E}[O^2_{D_k}]=\frac{\sigma^2}{2\theta}\mathbb{E}[1-e^{-2\theta D_k}].
$$

Finally, plugging (49) into (43), we have, with probability 1, the time-averaged MSE can be computed by:

$$
\begin{aligned}
&\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_{t=0}^{T}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\
=&\limsup_{K\to\infty}\frac{\sum_{k=1}^{K}\left(\mathbb{E}\left[\frac{\sigma^2}{2\theta}(D_{k-1}+W_{k-1})\right]\right)}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}\\
&-\frac{\sum_{k=1}^{K}\frac{1}{2\theta}\mathbb{E}[O^2_{D_{k-1}+W_{k-1}}]\mathbb{E}[e^{-2\theta D_k}]}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}\\
=&-\frac{\mathbb{E}[e^{-2\theta D_k}]}{2\theta}\times\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}\left[O^2_{D_k+W_k}\right]}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}+\frac{\sigma^2}{2\theta}.
\end{aligned}
\tag{51}
$$

Notice that optimal value of LHS of (51) is indeed mmse. Therefore, the problem is equivalent to

$$
\begin{aligned}
&\text{mmse}\\
&=\inf_{\pi\in\Pi_r}-\frac{\mathbb{E}[e^{-2\theta D_k}]}{2\theta}\times\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}\left[O^2_{D_k+W_k}\right]}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}+\frac{\sigma^2}{2\theta}
\end{aligned}
$$

Denote $\alpha^\star=\left(\sigma^2/2\theta-\text{mmse}\right)2\theta/\mathbb{E}[e^{-2\theta D_k}]$. Rearranging the terms yields

$$
\alpha^\star=\sup_{\pi\in\Pi_r}\lim_{K\to\infty}\frac{\sum_{k=1}^{K}\mathbb{E}\left[O^2_{D_k+W_k}\right]}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}.
\tag{52}
$$

According to [21], we have mmse $\leq\sigma^2/2\theta$. Therefore, $\alpha^\star\geq 0$.

APPENDIX C
PROOF OF LEMMA 2

Notice that

$$
\mathbb{E}[D_k+\hat{W}]>\frac{1}{f_{\max}}+c>\frac{1}{f_{\max}}.
$$

This means $\hat{W}$ is a fixed and feasible waiting solution to the problem. Then according to (6a), we have

$$
\alpha^\star\geq\frac{\mathbb{E}[O^2_{D+\hat{W}}]}{\mathbb{E}[D+\hat{W}]}.
$$

First we bound $\mathbb{E}[D+\hat{W}]\leq D_{\text{ub}}+\hat{W}$. Next we bound $\mathbb{E}[O^2_{D+\hat{W}}]$ as

$$
\begin{aligned}
\mathbb{E}[O^2_{D+\hat{W}}]=&\frac{\sigma^2}{2\theta}\left(1-\mathbb{E}[e^{-2\theta(D+\hat{W})}]\right)\\
\overset{(a)}{\geq}&\frac{\sigma^2}{2\theta}\left(1-e^{-2\theta\hat{W}}\right),
\end{aligned}
$$

where (a) holds since $D\geq 0$ and e^{-x} is decreasing. Combining the above two terms we have

$$
\alpha^\star\geq\frac{\sigma^2(1-e^{-2\theta\hat{W}})}{2\theta(D_{\text{ub}}+\hat{W})}=\alpha_{\text{lb}}.
$$

For the upper bound, according to [21], we have

$$
\text{mmse}\geq\text{mse}_D=\frac{\sigma^2}{2\theta}\mathbb{E}[1-e^{-2\theta D}].
\tag{53}
$$

Plugging (53) into (8) yields

$$
\alpha^\star\leq\left(\frac{\sigma^2}{2\theta}-\frac{\sigma^2}{2\theta}\mathbb{E}[1-e^{-2\theta D}]\right)\frac{2\theta}{\mathbb{E}[e^{-2\theta D}]}=\sigma^2=\alpha_{\text{ub}}.
$$

APPENDIX D
PROOF OF LEMMA 3

To solve the problem, From general optimal stopping theory [41, Chapter 1], we know that the following stopping time should be optimal:

$$\tau_\star = \inf\{t \geq 0 : |V_t| \geq v_\star\}, \tag{54}$$

where $v_\star$ is the optimal stopping threshold to be found.

We solve (17) by the free-boundary approach [41]. To find the $v_\star$, we solve the following free boundary problem:

$$\frac{\sigma^2}{2}H''(v) - \theta v H'(v) = \beta,\ v \in (-v_\star, v_\star), \tag{55a}$$

$$H(\pm v_\star) = v_\star^2, \tag{55b}$$

$$H'(\pm v_\star) = \pm 2v_\star. \tag{55c}$$

where $H(v)$ is the value function of (17).

Let $S(v) = H'(v)$, (55a) implies:

$$S'(v) - \frac{2\theta v}{\sigma^2}S(v) = \frac{2\beta}{\sigma^2}. \tag{56}$$

Multiplying $e^{-\theta v^2/\sigma^2}$ on both sides of (56), we have:

$$[S(v)e^{-\frac{\theta}{\sigma^2}v^2}]' = \frac{2\beta}{\sigma^2}e^{-\frac{\theta}{\sigma^2}v^2}. \tag{57}$$

Then

$$S(v)e^{-\frac{\theta}{\sigma^2}v^2} = C_1 + \int_0^v \frac{2\beta}{\sigma^2}e^{-\frac{\theta}{\sigma^2}u^2}\mathrm{d}u, \tag{58}$$

where C_1 is a constant so that $S(\pm v_\star)$ satisfy (55c). Denote

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t. \tag{59}$$

Then,

$$\begin{aligned} S(v)e^{-\frac{\theta}{\sigma^2}v^2} &= C_1 + \frac{2\beta}{\sigma^2}\sqrt{\frac{\pi}{4}}\frac{\sigma}{\sqrt{\theta}}\mathrm{erf}\left(\frac{\sqrt{\theta}}{\sigma}v\right) \\ &= C_1 + \frac{\beta}{\sigma}\sqrt{\frac{\pi}{\theta}}\mathrm{erf}\left(\frac{\sqrt{\theta}}{\sigma}v\right) \end{aligned} \tag{60}$$

Therefore, we have:

$$\begin{aligned} H'(v) = S(v) &= C_1 e^{\frac{\theta}{\sigma^2}v^2} + \frac{\beta}{\sigma}\sqrt{\frac{\pi}{\theta}}e^{\frac{\theta}{\sigma^2}v^2}\mathrm{erf}\left(\frac{\sqrt{\theta}}{\sigma}v\right) \\ &= C_1 e^{\frac{\theta}{\sigma^2}v^2} + \frac{2\beta}{\sigma\sqrt{\theta}}F\left(\frac{\sqrt{\theta}}{\sigma}v\right),\ v \in (-v_\star, v_\star), \end{aligned} \tag{61}$$

where $F(x) = e^{x^2}\int_0^x e^{-t^2}\mathrm{d}t$. Consider that $H'(v)$ is odd but $e^{\theta v^2/\sigma^2}$ is even, we have $C_1 = 0$. Therefore:

$$H'(v) = \frac{2\beta}{\sigma\sqrt{\theta}}F\left(\frac{\sqrt{\theta}}{\sigma}v\right). \tag{62}$$

Plugging (62) into the boundary condition (55c), we have:

$$\frac{2\beta}{\sigma\sqrt{\theta}}F\left(\frac{\sqrt{\theta}}{\sigma}v_\star\right) = 2v_\star. \tag{63}$$

Multiplying $\sqrt{\theta}/\sigma$ on both sides of (63), we have:

$$\frac{\beta}{\sigma^2}F\left(\frac{\sqrt{\theta}}{\sigma}v_\star\right) = \frac{\sqrt{\theta}}{\sigma}v_\star. \tag{64}$$

Finally, denote $G(x) = F(x)/x$. the optimum threshold $v_\star$ can be obtained by:

$$\frac{\beta}{\sigma^2}G\left(\frac{\sqrt{\theta}}{\sigma}v_\star\right) = 1, \tag{65}$$

Therefore, we have

$$v_\star = \frac{\sigma}{\sqrt{\theta}}G^{-1}\left(\frac{\sigma^2}{\beta}\right).$$

APPENDIX E
PROOF OF THEOREM 1

According to Lemma 6, since α_k and $\mathbb{E}[L_k]$ is bounded by a function of α, to show that the average MSE $(1/S_{k+1})\int_0^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t$ converges to mmse, it is then suffice to show that sequence

$$\xi_k := \frac{1}{k}\left(\int_0^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times S_{k+1}\right) \tag{66}$$

converges to 0 almost surely.

Our proof is based on the perturbed ODE approach [42, Chapter 7] for analyzing stochastic approximation. To use the ODE approach, first we need to rewrite ξ_k in recursive form as follows:

$$\begin{aligned} \xi_k =& \frac{1}{k}\left(\int_0^{S_k}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times S_k \right. \\ &\left. + \int_{S_k}^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times L_k\right) \\ \overset{(a)}{=}& \frac{1}{k}(k-1)\xi_{k-1} \\ &+ \frac{1}{k}\left(\int_{S_k}^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times L_k\right) \\ =& \xi_{k-1} + \frac{1}{k}\underbrace{\left(-\xi_{k-1} + \int_{S_k}^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times L_k\right)}_{=:G_k}, \end{aligned} \tag{67}$$

where (a) is from the definition of ξ_{k-1} in (66). In (67), $1/k$ can be viewed as a step-size of updating ξ_k and G_k is the updating direction. We can further decompose G_k as follows:

$$\begin{aligned} G_k &= -\xi_{k-1} + \int_{S_k}^{S_k+D_k}(X_t - \hat{X}_t)^2\mathrm{d}t \\ &\quad + \int_{S_k+D_k}^{S_k+D_k+W_k}(X_t - \hat{X}_t)^2\mathrm{d}t - \mathrm{mmse} \times L_k \\ &= -\xi_{k-1} + \int_{S_k}^{S_k+D_k}(X_t - \hat{X}_t)^2\mathrm{d}t \end{aligned}$$

$$+\int_{S_k+D_k}^{S_k+D_k+W_k}(X_t-\hat{X}_t)^2\mathrm{d}t-\left(\frac{\sigma^2}{2\theta}-\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star\right)\times L_k$$
$$=-\xi_{k-1}+\underbrace{\int_{S_k}^{S_k+D_k}O^2_{L_{k-1}+(t-S_k)}\mathrm{d}t}_{=:G_{k,1}}$$
$$+\underbrace{\int_{S_k+D_k}^{S_{k+1}}O^2_{D_k+(t-(S_k+D_k))}\mathrm{d}t}_{=:G_{k,2}}-\underbrace{\left(\frac{\sigma^2}{2\theta}-\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star\right)\times L_k}_{=:G_{k,3}}. \tag{68}$$

Let $\mathbb{E}_k[\cdot]\triangleq\mathbb{E}[\cdot|\mathcal{H}_{k-1}]$ be the conditional probability given historical information $\mathcal{H}_{k-1}$. Then according to (47), since the transmission delay D_k is independent of $O_{L_{k-1}}=X_{S_k}-\hat{X}_{S_k}$, the conditional expectation $\mathbb{E}[G_{k,1}]$ can be computed by:

$$\begin{aligned}\mathbb{E}_k[G_{k,1}]=&\mathbb{E}\left[\int_{S_k}^{S_k+D_k}(X_t-\hat{X}_t)^2\mathrm{d}t|\mathcal{H}_{k-1}\right]\\ =&\mathbb{E}\left[\frac{\sigma^2}{2\theta}D_k\right]+\frac{1}{2\theta}O^2_{L_{k-1}}\left(1-\mathbb{E}[e^{-2\theta D}]\right)\\ &-\frac{\sigma^2}{4\theta^2}\mathbb{E}\left[1-e^{-2\theta D}\right],\end{aligned} \tag{69}$$

Similarly, through (48), the conditional expectation of the $G_{k,2}$ can be computed by:

$$\begin{aligned}\mathbb{E}_k[G_{k,2}]=&\mathbb{E}\left[\int_{S_k+D_k}^{S_k+D_k+W_k}(X_t-\hat{X}_t)^2\mathrm{d}t|\mathcal{H}_{k-1}\right]\\ =&\mathbb{E}_k\left[\frac{\sigma^2}{2\theta}W_k-\frac{1}{2\theta}O^2_{L_k}\right]+\mathbb{E}_k\left[\frac{1}{2\theta}O^2_{D_k}\right].\end{aligned} \tag{70}$$

And the conditional expectation of $G_{k,3}$ can be computed by:

$$\mathbb{E}_k[G_{k,3}]=\left(\frac{\sigma^2}{2\theta}-\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star\right)\mathbb{E}_k[L_k]. \tag{71}$$

From (69)–(71), we can compute the conditional expectation of $\mathbb{E}_k[G_k]$ by:

$$\begin{aligned}&\mathbb{E}_k[G_k]\\ &\overset{(b)}{=}-\xi_{k-1}+\cancel{\mathbb{E}\left[\frac{\sigma^2}{2\theta}D_k\right]}+\frac{1}{2\theta}O^2_{L_{k-1}}\left(1-\mathbb{E}[e^{-2\theta D}]\right)\\ &\quad\cancel{-\frac{\sigma^2}{4\theta^2}\mathbb{E}\left[1-e^{-2\theta D}\right]}\\ &\quad+\mathbb{E}_k\left[\cancel{\frac{\sigma^2}{2\theta}}W_k-\frac{1}{2\theta}O^2_{L_k}\right]+\cancel{\mathbb{E}_k\left[\frac{1}{2\theta}O^2_{D_k}\right]}\\ &\quad-\left(\cancel{\frac{\sigma^2}{2\theta}}-\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star\right)\mathbb{E}_k[L_k]\\ &=-\xi_{k-1}+\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}O^2_{L_{k-1}}-\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star l(\alpha_{k-1})\\ &\quad-\frac{1}{2\theta}(\mathbb{E}_k[O^2_{L_k}]-\alpha_k\mathbb{E}_k[L_k])\\ &\quad+\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star l(\alpha_{k-1})\\ &\quad-\frac{1}{2\theta}\alpha_k\mathbb{E}_k[L_k]+\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star\mathbb{E}_k[L_k]\end{aligned}$$
$$\begin{aligned}&=-\xi_{k-1}+\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\left(O^2_{L_{k-1}}-\alpha^\star l(\alpha_{k-1})\right)\\ &\quad-\frac{1}{2\theta}\left(o(\alpha_k)-\alpha_k l(\alpha_k)\right)+\frac{1}{2\theta}\left(\alpha^\star l(\alpha_{k-1})-\alpha_k l(\alpha_k)\right)\\ &\quad+\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star(l(\alpha_k)-l(\alpha_{k-1}))\\ &=-\xi_{k-1}-\frac{1}{2\theta}\left(o(\alpha_k)-\alpha_k l(\alpha_k)\right)\\ &\quad+\underbrace{\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\left(O^2_{L_{k-1}}-o(\alpha_{k-1})\right)}_{=:\beta_{k,1}}\\ &\quad+\underbrace{\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\left(o(\alpha_{k-1})-\alpha^\star l(\alpha_{k-1})\right)}_{=:\beta_{k,2}}\\ &\quad+\underbrace{\frac{1}{2\theta}\alpha^\star(l(\alpha_{k-1})-l(\alpha_k))}_{=:\beta_{k,3}}+\underbrace{\frac{1}{2\theta}(\alpha^\star-\alpha_k)l(\alpha_k)}_{\beta_{k,4}}\\ &\quad+\underbrace{\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\alpha^\star(l(\alpha_k)-l(\alpha_{k-1}))}_{\beta_{k,5}}.\end{aligned} \tag{72}$$

where (b) is obtained because $\mathbb{E}\left[\frac{1}{2\theta}O^2_{D_k}\right]=\frac{\sigma^2}{4\theta^2}\mathbb{E}[1-e^{-2\theta D}]$ by (24b). Terms $\beta_{k,1},\cdots,\beta_{k,5}$ can be viewed as the bias terms in the ODE. Denote $\delta M_k:=G_k-\mathbb{E}_k[G_k]$ be the difference between the actual update and the conditional expectation, and define function:

$$f(\xi,\alpha)=-\xi-\frac{1}{2\theta}\left(o(\alpha)-\alpha l(\alpha)\right). \tag{73}$$

Plugging (72) into (67), we have:

$$\xi_k=\xi_{k-1}+\frac{1}{k}\left(f(\xi_{k-1},\alpha_k)+\sum_{j=1}^{5}\beta_{k,j}+\delta M_k\right). \tag{74}$$

Denote $t_0=0$ and $t_k:=\sum_{j=0}^{k-1}(1/j)$ to be the cumulative step-size sequences. Select $m(t)\in\mathbb{N}^+$ to be the largest integer so that $t_{m(t)}\leq t$. To show that the ODE (74) converges to 0 with almost surely, we will then verify the following statements, whose proof are provided in Appendix G:

Lemma 9: The updating steps $\{G_k\}$ and the difference sequence $\{\delta M_k\}$ have the following properties:

(a) For each constant N, the expectation $\mathbb{E}[|G_k|\mathbb{I}_{(|\xi_{k-1}|\leq N)}]$ is bounded for each k, i.e.,

$$\sup_k\mathbb{E}[|G_k|\mathbb{I}_{(|\xi_{k-1}|\leq N)}]<\infty. \tag{75}$$

(b) Function $f(\xi,\alpha)$ is continuous in ξ for each α.

(c) For any running time T, the following limit holds for all ξ and $\mu>0$:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=m(jT)}^{m(jT+t)-1}\frac{1}{i}(f(\xi,\alpha_i)-f(\xi,\alpha^\star)\right|\geq\mu\right)=0.$$

(d) The difference sequence $\delta M_k=G_k-\mathbb{E}_k[G_k]$ satisfies:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\delta M_i\right|\geq\mu\right)=0..$$

(e) The sum of the bias terms defined in (72) satisfies:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=m(jT)}^{m(jT+t)-1}\sum_{b=1}^{5}\frac{1}{i}\beta_{i,b}\right|\geq\mu\right)=0. \quad (76)$$

(f) Function $f(\xi,\alpha)$ can be decomposed into the sum of function of ξ and a function of α, i.e.,

$$f(\xi,\alpha)=-\xi-\frac{1}{2\theta}g_0(\alpha). \quad (77)$$

Since $g_0(\alpha^\star)=0$, we have $-\xi=f(\xi,\alpha^\star)$. Moreover,

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq n}\sum_{i=m(j\tau)}^{m(j\tau+\tau)-1}|\frac{1}{i}g_0(\alpha_i)|\geq\mu\right)=0. \quad (78)$$

(g) For each ξ,ξ', function $f(\xi,\alpha)$ satisfies:

$$|f(\xi,\alpha)-f(\xi',\alpha)|=|\xi-\xi'|. \quad (79)$$

Finally, according to [42, p. 166, Theorem 1.1], sequence $\{\xi_k\}$ converges to some limits of the ODE:

$$\dot{\xi}=f(\xi,\alpha^\star)=-\xi. \quad (80)$$

Since function $f(\cdot,\alpha^\star)$ is monotonically decreasing, $\xi=0$ is the unique equilibrium point of the ODE (80). Therefore, ξ_k converges to 0 almost surely, and the time-averaged MSE converges to the mmse with probability 1.

APPENDIX F
PROOF OF THEOREM 2

The cumulative regret, i.e., the difference between the expected cumulative MSE using the online algorithm compared with the MSE optimum sampling up to sample $(K+1)$ can be upper bounded as follows:

$$\begin{aligned}\mathcal{R}_K=&\mathbb{E}\left[\int_0^{S_{K+1}}(X_t-\hat{X}_t)^2\mathrm{d}t\right]-\text{mmse}\times\mathbb{E}[S_{k+1}]\\ \overset{(a)}{=}&-\frac{\mathbb{E}[e^{-2\theta D_k}]}{2\theta}\times\left(\sum_{k=1}^{K}\mathbb{E}[O^2_{D_k+W_k}]\right)\\ &+(\text{mse}_\infty-\text{mmse})\times\left(\sum_{k=1}^{K}\mathbb{E}[L_k]\right)\\ \overset{(b)}{=}&\frac{\mathbb{E}[e^{-2\theta D_k}]}{2\theta}\times\left(-\sum_{k=1}^{K}\left(\mathbb{E}[O^2_{D_k+W_k}]-\alpha^\star\mathbb{E}[L_k]\right)\right),\end{aligned} \quad (81)$$

where (a) is obtained by (51) and $\text{mse}_\infty=\sigma^2/2\theta$, and (b) is obtained by substituting $\text{mse}_\infty-\text{mmse}=\alpha^\star\mathbb{E}[e^{-2\theta D}]/2\theta$ from (8).

Then to further bound the cumulative regret computed by (81), let $W_k^\star$ be the waiting time selected by using parameter $\alpha^\star$ (i.e., the MSE minimum sampling policy). Then it is suffice to upper bound each term $-\mathbb{E}[O^2_{D_k+W_k}]+\alpha^\star\mathbb{E}[L_k]$ for each k as follows:

$$\begin{aligned}&-\mathbb{E}[O^2_{D_k+W_k}-\alpha^\star L_k]\\ &=-\mathbb{E}[O^2_{D_k+W_k}-\alpha_k L_k]-\mathbb{E}[(\alpha_k-\alpha^\star)L_k]\\ &\overset{(c)}{\leq}-\mathbb{E}[O^2_{D_k+W_k^\star}-\alpha_k L_k^\star]-\mathbb{E}[(\alpha_k-\alpha^\star)L_k]\\ &=-\mathbb{E}[O^2_{D_k+W_k^\star}-\alpha^\star L_k^\star]-\mathbb{E}[(\alpha_k-\alpha^\star)(l(\alpha_k)-l(\alpha^\star))]\\ &\overset{(d)}{=}-\mathbb{E}[(\alpha_k-\alpha^\star)(l(\alpha_k)-l(\alpha^\star)]\\ &\overset{(e)}{\leq}\max_{\alpha\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]}|R_1'(v(\alpha))v'(\alpha)|\times|\alpha_k-\alpha^\star|.\end{aligned} \quad (82)$$

where (c) is because W_k is the optimum policy that minimizes $-\mathbb{E}[O^2_{D_k+w}]+\alpha_k\mathbb{E}[D_k+w]$ and therefore we have $-\mathbb{E}[O^2_{D_k+W_k}-\alpha_k L_k]\leq-\mathbb{E}[O^2_{D_k+W_k^\star}-\alpha^\star L_k^\star]$; (d) is because $\mathbb{E}[O^2_{D_k+W_k^\star}-\alpha^\star L_k^\star]=0$ by (22); (e) is from Corollary 1.

Finally, plugging (82) into (81) for each term k, the cumulative regret $\mathcal{R}_K$ can be bounded, i.e.,

$$\begin{aligned}\mathcal{R}_K\overset{(f)}{\leq}&\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\times\left(\sum_{k=1}^{K}\frac{C}{D_{\text{lb}}^2}\max_{\alpha\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]}|R_1'(v(\alpha))v'(\alpha)|\frac{1}{k}\right)\\ \leq&\frac{\mathbb{E}[e^{-2\theta D}]}{2\theta}\frac{C}{D_{\text{lb}}^2}\max_{\alpha\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]}|R_1'(v(\alpha))v'(\alpha)|\ln(K+1),\end{aligned} \quad (83)$$

where (f) is obtained by Theorem 4.

APPENDIX G
PROOF OF LEMMA 9

We will verify each statement in Lemma 9 respectively:

(a) By substituting G_k with (68), we can upper bound $\mathbb{E}[|G_k|\mathbb{I}_{(|\xi_{k-1}|\leq N)}]$ as follows:

$$\begin{aligned}&\mathbb{E}[|G_k|\mathbb{I}_{(|\xi_{k-1}|\leq N)}]\\ &\leq\mathbb{E}[|\xi_{k-1}|\mathbb{I}_{(|\xi_{k-1}|\leq N}]+\mathbb{E}\left[\int_{t=S_k}^{S_{k+1}}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\ &\quad+\text{mmse}\mathbb{E}[L_k].\end{aligned} \quad (84)$$

The first term $\mathbb{E}[|\xi_{k-1}|\mathbb{I}_{(|\xi_{k-1}|\leq N}]\leq N<\infty$ is bounded. Then notice that $\mathbb{E}[L_k]$ is bounded by Lemma 6 and $\text{mmse}\leq\text{mse}_\infty$, the third term $\text{mmse}\mathbb{E}[L_k]$ is also bounded. It then remains to show that the second term $\mathbb{E}\left[\int_{S_k}^{S_{k+1}}(X_t-\hat{X}_t)^2\mathrm{d}t\right]$ is bounded. According to (49), the expectation of the second term can be computed by:

$$\begin{aligned}&\mathbb{E}\left[\int_{S_k}^{S_{k+1}}(X_t-\hat{X}_t)^2\mathrm{d}t\right]\\ &=\mathbb{E}\left[\frac{\sigma^2}{2\theta}L_{k-1}\right]-\frac{1}{2\theta}\mathbb{E}[O^2_{L_{k-1}}]\mathbb{E}[e^{-2\theta D_k}].\end{aligned} \quad (85)$$

Since $\alpha_k\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]$ is bounded and function $l(\alpha_{k-1})=\mathbb{E}[L_{k-1}]$, $o(\alpha_{k-1})=\mathbb{E}[O^2_{L_{k-1}}]$ are both bounded for $\alpha\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]$, the expectation of the second term $\mathbb{E}[\int_{S_k}^{S_{k+1}}(X_t-\hat{X}_t)^2\mathrm{d}t]$ is also bounded. This verifies statement (a).

(b) Function $f(\xi,\alpha)$ can be decoupled into $=-\xi-1/2\theta\times g_0(\alpha)$ and is thus continuous in ξ for each α.

To proceed with the proof of statement $(c)-(f)$, we re-state the following lemma, whose proof is provided in [34, Appendix G]

Lemma 10: Let $\{\psi_k\}$ be a sequence. Then $\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\left|\sum_{i=k}^{j}\frac{1}{i}\psi_i\right|\geq\mu\right)=0$ holds if one of the following condition is satisfied:

(1) ψ_k is a martingale sequence and its second order moment is bounded, i.e., $\mathbb{E}_k[\psi_k]=0, \sup_k\mathbb{E}[\psi_k^2]<\infty$. The correlation between each $(k,k'), k\neq k'$ pair satisfies: $\mathbb{E}[\psi_k\psi_{k'}]=0$.
(2) $\mathbb{E}[|\psi_k|]=\mathcal{O}(k^{-\varepsilon}), \varepsilon>0$.

(c) According to Lemma 7, since $g_0(\alpha)$ is monotonic decreasing and convex, the difference $|g_0(\alpha)-g_0(\alpha')|\leq N_1|\alpha-\alpha'|$. Therefore,

$$\begin{aligned}|f(\xi,\alpha_k)-f(\xi,\alpha^\star)|&=\frac{1}{2\theta}|g_0(\alpha_k)-g_0(\alpha^\star)|\\&\leq\frac{N_1}{2\theta}|\alpha_k-\alpha^\star|.\end{aligned}\tag{86}$$

Therefore, the expectation of $f(\xi,\alpha_k)-f(\xi,\alpha^\star)$ can be upper bounded by:

$$\begin{aligned}&\mathbb{E}\left[f(\xi,\alpha_k)-f(\xi,\alpha^\star)\right]\leq\mathbb{E}[\frac{N_1}{2\theta}|\alpha_k-\alpha^\star|]\\&\overset{(a)}{\leq}\frac{N_1}{2\theta}\sqrt{\mathbb{E}[(\alpha_k-\alpha^\star)^2]}\overset{(b)}{=}\frac{N_1}{2\theta}\sqrt{\frac{C}{D_{\text{lb}}^2}}\sqrt{\frac{1}{k}}\end{aligned}\tag{87}$$

where equality (a) is by Cauchy-Schwartz inequality and equality (b) is from Theorem 4. Since term $f(\xi,\alpha_k)-f(\xi,\alpha^\star)$ satisfies condition 2 in Lemma 10, statement (c) is verified.

(d) Denote $\delta M_{k,j}:=G_{k,j}-\mathbb{E}_k[G_{k,j}]$. Since $G_k=G_{k,1}+G_{k,2}-G_{k,3}$, the difference term $\delta M_k=\delta M_{k,1}+\delta M_{k,2}-\delta M_{k,3}$ also consists of three parts. By the union bound,

$$\begin{aligned}&\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\delta M_i\right|\geq\mu\right)\\&\leq\sum_{p=1}^{3}\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\delta M_{i,p}\right|\geq\mu/3\right).\end{aligned}\tag{88}$$

Therefore, to show that statement (d) is satisfied, it is suffice to show that each term $\delta M_{i,p}, p=1,2,3$ satisfies condition (1) in Lemma 10.

Notice that for fixed $O_{L_{k-1}}$, the first difference term $\delta M_{k,1}=G_{k,1}-\mathbb{E}_k[G_{k,1}]$ depends only on D_k and the OU process evolution during $[S_k,S_k+D_k)$. Therefore, $\mathbb{E}[\delta M_{k,1}]=0$ and $\mathbb{E}[\delta M_{k,1}\delta M_{k',1}]=0, \forall k\neq k'$ due to the independence of D_k and $D_{k'}$. Then, notice that $\text{Var}(\delta M_{k,1})\leq\mathbb{E}[\delta M_{k,1}^2]\leq\mathbb{E}[G_{k,1}^2]$. To show that $\text{Var}(\delta M_{k,1})<\infty$ is bounded, it is suffice to show $\mathbb{E}[G_{k,1}^2]$ is bounded, which is shown as follows:

$$\begin{aligned}&\mathbb{E}[G_{k,1}^2]\\&=\mathbb{E}\left[\left(\int_{t=S_k}^{S_k+D_k}O^2_{L_{k-1}+(t-S_k)}\mathrm{d}t\right)^2\right]\\&\overset{(b)}{\leq}\mathbb{E}\left[D_k\int_{t=S_k}^{S_k+D_k}O^4_{L_{k-1}+(t-S_k)}\mathrm{d}t\right]\\&\overset{(c)}{\leq}\mathbb{E}\left[D_k\int_{t=S_k}^{S_k+D_k}3\left(\frac{\sigma^2}{2\theta}(1-e^{-2\theta(L_{k-1}+(t-S_k))})\right)^2\mathrm{d}t\right]\\&\leq 3D_{\textbf{ub}}\left(\frac{\sigma^2}{2\theta}\right)^2.\end{aligned}\tag{89}$$

where (b) is by Cauchy-Schwartz inequality; inequality (c) is from (103). Since $\delta M_{k,1}$ meets the first condition in Lemma 10, we have:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\delta M_{i,1}\geq\frac{1}{3}\mu\right|\right)=0.\tag{90}$$

The difference sequence $\delta M_{k,2}$ and $\delta M_{k,3}$ only depends on the transmission delay D_k and the OU process evolution in frame k. Using similar methods, it can be shown that sequences $\{\delta M_{k,2}\}$ and $\{\delta M_{k,3}\}$ satisfy condition 1 in Lemma 10. Since $\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\delta M_{i,p}\geq\frac{1}{3}\mu\right|\right)=0$ holds for $p=1,2,3$, plugging into (88) verifies statement (d).

(e) Through the union bound, we have:

$$\begin{aligned}&\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\sum_{b=1}^{5}\beta_{i,b}\right|\geq\mu\right)\\&\leq\sum_{b=1}^{5}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=k}^{j}\frac{1}{i}\beta_{i,p}\right|\geq\mu/5\right).\end{aligned}\tag{91}$$

To show that statement (e) holds, it is suffice to show that each of the bias term satisfy:

$$\lim_{k\to\infty}\Pr\left(\sup_{j\geq k}\max_{0\leq t\leq T}\left|\sum_{i=m(jT)}^{m(jT+t)-1}\frac{1}{i}\delta\beta_{i,p}\right|\geq\mu/5\right)=0,\quad\forall p.\tag{92}$$

the second condition in Lemma 10. We will then upper bound the expectation of each bias term $\mathbb{E}[\beta_{k,p}]$, respectively.

The first bias term satisfies $\mathbb{E}[\beta_{k,1}]=0$ and is hence a martingale sequence. We can bound $\mathbb{E}[\beta_{k,1}^2]$ by:

$$\begin{aligned}&\mathbb{E}[\beta_{k,1}^2]=\text{Var}[\beta_{k,1}]\\&=\text{Var}\left[\left(\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}(O^2_{L_{k-1}}-o(\alpha_{k-1}))\right)\right]\\&=\mathbb{E}\left[\left(\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}(O^2_{L_{k-1}}-o(\alpha_{k-1}))\right)^2\right]\\&\leq 2\left(\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\right)^2\mathbb{E}\left[O^4_{L_{k-1}}+o(\alpha_{k-1})^2\right].\end{aligned}\tag{93}$$

Then according to Lemma 6, $\mathbb{E}[L_{k-1}^4]<\infty$ and $\mathbb{E}[o(\alpha_{k-1})^2]=\mathbb{E}[O^2_{L_{k-1}}]^2<\infty$, term $\beta_{k,1}$ satisfies Condition 1, Lemma 10. Therefore, (92) holds for $p=1$. The expectation of the second bias term $\beta_{k,2}$ can be upper bounded by:

$$\begin{aligned}&\mathbb{E}\left[|\beta_{k,2}|\right]\\&=\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\mathbb{E}[|o(\alpha_{k-1})-\alpha^\star l(\alpha_{k-1})|]\\&=\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\left(\mathbb{E}[|o(\alpha_{k-1})-\alpha_{k-1}l(\alpha_{k-1})|]\right.\\&\quad\left.+\mathbb{E}[(\alpha_{k-1}-\alpha^\star)l(\alpha_{k-1})]\right)\\&\leq\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}\mathbb{E}[N_1\times|\alpha_{k-1}-\alpha^\star|]\end{aligned}$$

$$+\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}l(\alpha_{\text{lb}})\mathbb{E}[|\alpha_{k-1}-\alpha^\star|]. \tag{94}$$

Recall that by Theorem 4, $\mathbb{E}[|\alpha_{k-1}-\alpha^\star|] \le \sqrt{\mathbb{E}[(\alpha_{k-1}-\alpha^\star)^2]} = \mathcal{O}(1/\sqrt{k})$, (94) implies $\mathbb{E}[|\beta_{k,1}|] = \mathcal{O}(1/\sqrt{k})$ and satisfies Lemma 10 condition 2. Equation (92) holds for $p=2$.

We then proceed to upper bound the expectation of of the third bias term by:

$$\begin{aligned}
&\mathbb{E}[|\beta_{k,2}|] = \mathbb{E}\left[\frac{1}{2\theta}\alpha^\star(l(\alpha_{k-1})-l(\alpha_k))\right]\\
&\le \frac{1}{2\theta}\alpha^\star\left(\mathbb{E}\left[|l(\alpha_{k-1})-l(\alpha^\star)|\right]+\mathbb{E}\left[|l(\alpha_k)-l(\alpha^\star)|\right]\right)\\
&\overset{(d)}{\le} \frac{1}{2\theta}\alpha^\star N\left(\mathbb{E}[|\alpha_{k-1}-\alpha^\star|]+\mathbb{E}[|\alpha_k-\alpha^\star|]\right)\\
&\le \frac{1}{2\theta}\alpha^\star N\left(\sqrt{\mathbb{E}[|\alpha_{k-1}-\alpha^\star|^2]}+\sqrt{\mathbb{E}[|\alpha_k-\alpha^\star|^2]}\right)\\
&= \mathcal{O}(k^{-1/2}).
\end{aligned} \tag{95}$$

where inequality (d) is obtained by Corollary 1. Therefore, $\beta_{k,2}$ also satisfies the Condition 2 in Lemma 10 and (92) holds for $p=2$. Since term $\alpha_k \in [\alpha_{\text{lb}}, \alpha_{\text{ub}}]$, we can show that $\beta_{k,3}, \beta_{k,4}, \beta_{k,5}$ satisfy Condition 2 Lemma 10 and thus (92) also holds for $p = 3 \sim 5$. Considering that (92) holds for $p = 1 \sim 5$, through the union bound (91), we show that statement (e) holds.

(f) According to the convexity of function $g_0(\cdot)$ from (39b) Lemma 7, we have $|g_0(\alpha)-g_0(\alpha^\star)| \le N_1|\alpha-\alpha^\star|$. Therefore, we can upper bound the expected value of $\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}g_0(\alpha_k)$ as follows:

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}g_0(\alpha_k)\right|\right]\\
\le&\mathbb{E}\left[\left|\frac{1-\mathbb{E}[e^{-2\theta D}]}{2\theta}N_1(\alpha^\star-\alpha_k)\right|\right]\\
\le&\mathcal{O}(1/\sqrt{k}).
\end{aligned} \tag{96}$$

This verifies Condition 2 in Lemma 10 and therefore verifies statement (f).

Appendix H
Proof of Theorem 3

Recall that the sampling debt queue U_k evolves as

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

According to [43], in order to satisfy the sampling constraint, it is sufficient to prove that

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

Here we adopt the Lyapunov drift-plus-penalty method to prove the stability of U_k. Define the Lyapunov function as

$$L(U_k) = \frac{1}{2}U_k^2, \tag{97}$$

and the Lyapunov drift is defined by

$$\Delta(U_k) = \mathbb{E}[L(U_{k+1}) - L(U_k)|U_k]. \tag{98}$$

First we upper bound U_{k+1}^2:

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

Plugging the above inequality into (97) yields

$$\begin{aligned}
&L(U_{k+1}) - L(U_k)\\
\le&\frac{1}{2}\left[\left(U_k + \frac{1}{f_{\max}} - L_k\right)^2 - U_k^2\right]\\
=&-U_k\left(L_k - \frac{1}{f_{\max}}\right) + \frac{1}{2}\left(\frac{1}{f_{\max}} - L_k\right)^2.
\end{aligned}$$

Plugging the above equation into (98) and then take the expectation on both sides of (98) yields

$$\begin{aligned}
&\Delta(U_k)\\
&\overset{(a)}{\le} -U_k\mathbb{E}\left[L_k - \frac{1}{f_{\max}}\middle| U_k\right]\\
&\quad+\frac{1}{2}\left(\frac{1}{f_{\max}^2} + \mathbb{E}[D_k^2] + \mathbb{E}[W_k^2|U_k] + 2\mathbb{E}[D_kW_k|U_k]\right)\\
&\le -U_k\mathbb{E}\left[L_k - \frac{1}{f_{\max}}\middle| U_k\right]\\
&\quad+\frac{1}{2}\left(\frac{1}{f_{\max}^2} + M_{\text{ub}} + \mathbb{E}[W_k^2|U_k] + 2\mathbb{E}[D_kW_k|U_k]\right).
\end{aligned} \tag{99}$$

where (a) holds since D_k is independent of U_k. Similar to the proof of Lemma 6, we can bound $\mathbb{E}[D_kW_k|U_k]$ and $\mathbb{E}[W_k^2|U_k]$ as

$$\begin{aligned}
\mathbb{E}[D_kW_k|U_k] &\le D_{\text{ub}}\frac{v(\eta)^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\eta)^2}\\
\mathbb{E}[W_k^2|U_k] &\le \frac{2v(\eta)^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\eta)^2}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\Delta(U_k) \le& -U_k\mathbb{E}\left[W_k + D_k - \frac{1}{f_{\max}}\middle| U_k\right]\\
&+\frac{1}{2}\left(\frac{1}{f_{\max}^2} + M_{\text{ub}} + \frac{2v(\eta)^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\eta)^2}\right.\\
&\left.+2D_{\text{ub}}\frac{v(\eta)^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\eta)^2}\right).
\end{aligned}$$

Now we upper bound the first term of the RHS of (99). According to (16), the waiting time W_k is the optimal solution to

$$\sup_w \mathbb{E}\left[O_{D_k+w}^2 - (\alpha_k-\lambda_k)\,w|O_{D_k}, D_k\right]. \tag{100}$$

For simplicity, we denote the historical information O_{D_k}, D_k to be $\mathcal{M}_{k-1}$.

Let W_ϵ be the waiting time under policy π_ϵ. According to (100), we have

$$\begin{aligned}
&\mathbb{E}[O_{D_k+W_k}^2|\mathcal{M}_{k-1}] - \mathbb{E}[(\alpha_k-\lambda_k)W_k|\mathcal{M}_{k-1}]\\
&\ge \mathbb{E}[O_{D_k+W_\epsilon}^2|\mathcal{M}_{k-1}] - \mathbb{E}[(\alpha_k-\lambda_k)W_\epsilon|\mathcal{M}_{k-1}].
\end{aligned}$$

Adding $\frac{1}{V}U_k\left(D_k-\frac{1}{f_{\max}}\right)$ on both sides yields

$$\begin{aligned}&\mathbb{E}[O^2_{D_k+W_k}|\mathcal{M}_{k-1}]-\mathbb{E}[(\alpha_k-\frac{1}{V}U_k)W_k|\mathcal{M}_{k-1}]\\&\quad+\frac{1}{V}U_k\left(D_k-\frac{1}{f_{\max}}\right)\\&\geq\mathbb{E}[O^2_{D_k+W_\epsilon}|\mathcal{M}_{k-1}]-\mathbb{E}[(\alpha_k-\frac{1}{V}U_k)W_\epsilon|\mathcal{M}_{k-1}]\\&\quad+\frac{1}{V}U_k\left(D_k-\frac{1}{f_{\max}}\right).\end{aligned}$$

Rearranging the terms yields

$$\begin{aligned}&-U_k\mathbb{E}\left[D_k+W_k-\frac{1}{f_{\max}}\middle|\mathcal{M}_{k-1}\right]\\\leq&-U_k\mathbb{E}\left[D_k+W_\epsilon-\frac{1}{f_{\max}}\middle|\mathcal{M}_{k-1}\right]\\&-V\mathbb{E}[O^2_{D_k+W_\epsilon}-\alpha_kW_\epsilon|\mathcal{M}_{k-1}]\\&+V\mathbb{E}[O^2_{D_k+W_k}-\alpha_kW_k|\mathcal{M}_{k-1}]\\\overset{(a)}{\leq}&-U_k\epsilon+V\mathbb{E}[O^2_{D_k+W_k}+\alpha_kW_\epsilon|\mathcal{M}_{k-1}]\\\overset{(b)}{\leq}&-U_k\epsilon+V\left(\frac{\sigma^2}{2\theta}+\alpha_{\text{ub}}W_{\text{ub}}\right),\end{aligned}$$

where (a) holds by Assumption 3; (b) holds by Lemma 6 and $W_{\text{ub}}=1/f_{\max}+D_{\text{ub}}$ for sufficiently small ϵ.

Now we have

$$\begin{aligned}\Delta(U_k)\leq&-U_k\epsilon+V\left(\frac{\sigma^2}{2\theta}+\alpha_{\text{ub}}W_{\text{ub}}\right)\\&+\frac{1}{2}\left(\frac{1}{f^2_{\max}}+M_{\text{ub}}+\frac{2v(\eta)^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\eta)^2}\right.\\&\left.+2D_{\text{ub}}\frac{v(\eta)^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\eta)^2}\right)\\\triangleq&-U_k\epsilon+C_1,\end{aligned}$$

where

$$\begin{aligned}C_1=&V\left(\frac{\sigma^2}{2\theta}+\alpha_{\text{ub}}W_{\text{ub}}\right)\\&+\frac{1}{2}\left(\frac{1}{f^2_{\max}}+M_{\text{ub}}+\frac{2v(\eta)^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\eta)^2}\right.\\&\left.+2D_{\text{ub}}\frac{v(\eta)^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\eta)^2}\right)\end{aligned}$$

is a constant. Summing up from $k=1$ to K yields

$$\mathbb{E}\left[\frac{1}{2}U^2_{k+1}-\frac{1}{2}U^2_1\right]\leq-\epsilon\sum_{k=1}^K\mathbb{E}[U_k]+KC_1.$$

Notice that $U_1=0$ and $U_{k+1}\geq0$. Thus we have

$$\epsilon\sum_{k=1}^K\mathbb{E}[U_k]\leq KC_1.$$

Rearranging the terms yields

$$\limsup_{K\to\infty}\frac{1}{K}\sum_{k=1}^K\mathbb{E}[U_k]\leq\frac{C_1}{\epsilon}<\infty.$$

APPENDIX I
PROOF OF AUXILIARY LEMMAS AND COROLLARIES

A. Proof of Corollary 1

Proof:

$$\begin{aligned}&|l(\alpha)-l(\alpha^\star)|\\&=|\mathbb{E}[D_k]+\mathbb{E}[\max\{R_1(v(\alpha))-R_1(|O_{D_k}|),0\}]\\&\quad-(\mathbb{E}[D_k]+\mathbb{E}[\max\{R_1(v(\alpha^\star))-R_1(|O_{D_k}|),0\}])|\\&\leq|R_1(v(\alpha))-R_1(v(\alpha^\star))|\\&\leq\max_{\alpha\in[\alpha_{\text{lb}},\alpha_{\text{ub}}]}|R_1'(v(\alpha))v'(\alpha)|\times|\alpha_k-\alpha^\star|.\end{aligned}\tag{101}$$

■

B. Proof of Lemma 6

Since $O^2_{L_k}$ is an instance of $O^2_{D_k+W_k}$, we just bound $\mathbb{E}[O^2_{D_k+W_k}]$ and $\mathbb{E}[O^4_{D_k+W_k}]$. Therefore we have

$$\mathbb{E}[O^2_{D_k+W_k}]=\frac{\sigma^2}{2\theta}\mathbb{E}[1-e^{-2\theta(D_k+W_k)}]\leq\frac{\sigma^2}{2\theta}.\tag{102}$$

$$\mathbb{E}[O^4_{D_k+W_k}]=3\mathbb{E}\left[\left(\frac{\sigma^2}{2\theta}(1-e^{-2\theta(D_k+W_k)})\right)^2\right]\leq\frac{3\sigma^4}{4\theta^2},\tag{103}$$

which verifies (38a) and (38b).

For L_k, according to Lemma 5 we can bound

$$\begin{aligned}\mathbb{E}[L_k]&=\mathbb{E}[D_k]+\mathbb{E}[\max\{R_1(v(\alpha_k))-R_1(|O_{D_k}|),0\}]\\&\leq D_{\text{ub}}+\mathbb{E}[R_1(v(\alpha_k))].\end{aligned}$$

Since $v(\alpha_k)$ is decreasing function with respect to α_k, $v(\alpha_k)$ can be bounded

$$0<v(\alpha_{\text{ub}})\leq v(\alpha_k)\leq v(\alpha_{\text{lb}})\overset{(a)}{<}\infty,\tag{104}$$

where (a) holds by Lemma 2.

Next, we bound $R_1(v)$ as

$$\begin{aligned}R_1(v)=&\frac{v^2}{\sigma^2}{}_2F_2\left(1,1;\frac{3}{2},2;\frac{\theta}{\sigma^2}v^2\right)\\=&\frac{v^2}{\sigma^2}\sum_{n=0}^\infty\frac{2^n}{(n+1)(2n+1)!!}(\frac{\theta}{\sigma^2}v^2)^n\\\overset{(a)}{\leq}&\frac{v^2}{\sigma^2}\sum_{n=0}^\infty\frac{1}{n!}(\frac{2\theta}{\sigma^2}v^2)^n\\=&\frac{v^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v^2}.\end{aligned}$$

where (a) holds by $n!\leq(2n+1)!!$. Then we have

$$0\leq R_1(v(\alpha_k))\leq\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}.\tag{105}$$

Therefore, we can bound $\mathbb{E}[L_k]$ as

$$0\leq\mathbb{E}[L_k]\leq D_{\text{ub}}+\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2},$$

which verifies (38c).

Finally, we rewrite $\mathbb{E}[L^2_k]$ as

$$\begin{aligned}\mathbb{E}[L^2_k]=&\mathbb{E}[(D_k+W_k)^2]\\=&\mathbb{E}[D^2_k]+2\mathbb{E}[D_kW_k]+\mathbb{E}[W^2_k]\end{aligned}$$

$$\begin{aligned}&=\mathbb{E}[D_k^2]+2\mathbb{E}[D_k\mathbb{E}[W_k|D_k]]+\mathbb{E}[W_k^2]\\&\le M_{\text{ub}}+2\mathbb{E}[D_k\mathbb{E}[W_k|D_k]]+\mathbb{E}[W_k^2].\end{aligned}\tag{106}$$

Next we bound $\mathbb{E}[D_k\mathbb{E}[W_k|D_k]]$ and $\mathbb{E}[W_k^2]$, respectively.

$$\begin{aligned}\mathbb{E}[W_k|D_k]&\overset{(a)}{\le}\mathbb{E}[W_k|D_k,O_{D_k}<v(\alpha_k)]\\&\overset{(b)}{=}\mathbb{E}[R_1(v(\alpha_k))-R_1(|O_{D_k}|)|D_k,|O_{D_k}|<v(\alpha_k)]\\&\le\mathbb{E}[R_1(v(\alpha_k))|D_k,|O_{D_k}|<v(\alpha_k)]\\&\le\mathbb{E}[R_1(v(\alpha_{\text{lb}}))]\\&\overset{(c)}{\le}\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2},\end{aligned}$$

where (a) holds because $W_k=0$ if $|O_{D_k}|>v(\alpha_k)$; (b) holds by Lemma 5; (c) holds by (105). Therefore, we have

$$\begin{aligned}\mathbb{E}[D_k\mathbb{E}[W_k|D_k]]&\overset{(a)}{\le}\mathbb{E}[D_k]\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}\\&\le D_{\text{ub}}\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}.\end{aligned}\tag{107}$$

where (a) holds because $\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}$ is a constant.

Now we bound $\mathbb{E}[W_k^2]$ as

$$\begin{aligned}\mathbb{E}[W_k^2]&=\mathbb{E}[\mathbb{E}[W_k^2|D_k,\alpha_k]]\\&\overset{(a)}{\le}\mathbb{E}[\mathbb{E}[W_k^2|D_k,\alpha_k,|O_{D_k}|<v(\alpha_k)]]\end{aligned}$$

where (a) holds because $W_k=0$ if $|O_{D_k}|>v(\alpha_k)$. Now we just need to bound $\mathbb{E}[W_k^2|D_k,\alpha_k,|O_{D_k}|<v(\alpha_k)]$. According to (28), W_k is the stopping time that an OU process exits a bounded set $[-v(\alpha_k),v(\alpha_k)]$ with the initial state O_{D_k}. Denote $t_v^{(1)}(x)$ and $t_v^{(2)}(x)$ to be the first and second moment of W_k with initial state x and bounded set $[-v,v]$. According to [44, Theorem 6.1], we have

$$\frac{\sigma^2}{2}\frac{\mathrm{d}^2t_v^{(2)}(x)}{\mathrm{d}x^2}-\theta x\frac{\mathrm{d}t_v^{(2)}(x)}{\mathrm{d}x}=-2t_v^{(1)}(x),\ x\in[-v,v]$$

where according to Lemma 5

$$t_v^{(1)}(x)=R_1(v)-R_1(x).\tag{108}$$

Let $s(x)=\frac{\mathrm{d}t_v^{(2)}(x)}{\mathrm{d}x}$, and we have

$$\frac{\sigma^2}{2}s'(x)-\theta xs(x)=-2t_v^{(1)}(x).$$

Multiplying $\frac{2}{\sigma^2}e^{-\frac{\theta}{\sigma^2}x^2}$ on both sides yields

$$s'(x)e^{-\frac{\theta}{\sigma^2}x^2}-\frac{2\theta}{\sigma^2}xs(x)e^{-\frac{\theta}{\sigma^2}x^2}=\frac{-4t_v^{(1)}(x)}{\sigma^2}e^{-\frac{\theta}{\sigma^2}x^2}.$$

This is equivalent to

$$\left(s(x)e^{-\frac{\theta}{\sigma^2}x^2}\right)'=\frac{-4t_v^{(1)}(x)}{\sigma^2}e^{-\frac{\theta}{\sigma^2}x^2}.$$

Therefore, we have

$$s(x)=Ce^{\frac{\theta}{\sigma^2}x^2}-e^{\frac{\theta}{\sigma^2}x^2}\int_{-v}^{x}\frac{4t_v^{(1)}(u)}{\sigma^2}e^{-\frac{\theta}{\sigma^2}u^2}\mathrm{d}u,$$

where C is a constant. Since $t_v^{(2)}(x)$ is even and takes the maximum when $x=0$. Therefore, we have

$$C=\int_{-v}^{0}\frac{4t_v^{(1)}(u)}{\sigma^2}e^{-\frac{\theta}{\sigma^2}u^2}\mathrm{d}u.\tag{109}$$

Since $t_v^{(2)}(x)$ is even, we only need to consider $x\in[-v,0]$. When $x\in[-v,0]$, $t_v^{(2)}(x)$ is increasing and $s(x)\ge 0$. Therefore, we have

$$0\le s(x)\le Ce^{\frac{\theta}{\sigma^2}x^2},\ x\in[-v,0].\tag{110}$$

Then for $x\in[-v,0]$

$$\begin{aligned}t_v^{(2)}(x)&=\int_{-v}^{x}s(u)\mathrm{d}u\\&\overset{(a)}{\le}\int_{-v}^{x}Ce^{\frac{\theta}{\sigma^2}x^2}\mathrm{d}u\\&\le\int_{-v}^{0}Ce^{\frac{\theta}{\sigma^2}x^2}\mathrm{d}u\\&\overset{(b)}{=}ve^{\frac{\theta}{\sigma^2}x^2}\int_{-v}^{0}\frac{4t_v^{(1)}(u)}{\sigma^2}e^{-\frac{\theta}{\sigma^2}u^2}\mathrm{d}u.\\&\overset{(c)}{\le}vR_1(v)e^{\frac{\theta}{\sigma^2}x^2}\int_{-v}^{0}\frac{4}{\sigma^2}e^{-\frac{\theta}{\sigma^2}u^2}\mathrm{d}u\\&=vR_1(v)e^{\frac{\theta}{\sigma^2}x^2}\frac{2}{\sigma}\sqrt{\frac{\pi}{\theta}}\mathrm{erf}\left(\frac{\sqrt{\theta}}{\sigma}v\right)\\&\overset{(d)}{\le}vR_1(v)e^{\frac{\theta}{\sigma^2}v^2}\frac{2}{\sigma}\sqrt{\frac{\pi}{\theta}}.\end{aligned}$$

where (a) holds by (110); (b) holds by (109); (c) holds by (108); (d) holds since $\mathrm{erf}(x)\le 1$. Since $t_v^{(2)}(x)$ is even for $x\in[-v,v]$, we have

$$t_v^{(2)}(x)\le\frac{2}{\sigma}\sqrt{\frac{\pi}{\theta}}vR_1(v)e^{\frac{\theta}{\sigma^2}v^2},\ x\in[-v,v].$$

This means

$$\begin{aligned}&\mathbb{E}[W_k^2|D_k,\alpha_k,|O_{D_k}|<v(\alpha_k)]\\&\le\frac{2}{\sigma}\sqrt{\frac{\pi}{\theta}}v(\alpha_k)R_1(v(\alpha_k))e^{\frac{\theta}{\sigma^2}v(\alpha_k)^2}\\&\overset{(a)}{\le}\frac{2}{\sigma}\sqrt{\frac{\pi}{\theta}}v(\alpha_{\text{lb}})\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}e^{\frac{\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}\\&=\frac{2v(\alpha_{\text{lb}})^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\alpha_{\text{lb}})^2},\end{aligned}$$

where (a) holds by (105). Therefore we have

$$\mathbb{E}[W_k^2]\le\frac{2v(\alpha_{\text{lb}})^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}.\tag{111}$$

Plugging (107) and (111) into (106) yields

$$\begin{aligned}\mathbb{E}[L_k^2]\le&M_{\text{ub}}+2D_{\text{ub}}\frac{v(\alpha_{\text{lb}})^2}{\sigma^2}e^{\frac{2\theta}{\sigma^2}v(\alpha_{\text{lb}})^2}\\&+\frac{2v(\alpha_{\text{lb}})^3}{\sigma^3}\sqrt{\frac{\pi}{\theta}}e^{\frac{3\theta}{\sigma^2}v(\alpha_{\text{lb}})^2},\end{aligned}$$

which verifies (38d).

C. Proof of Lemma 7

Proof: For notational simplicity, for each stopping rule w, denote $\tilde{L}(w,\alpha-\lambda):=\mathbb{E}[-O_{D+w}^2+(\alpha-\lambda)w]$, which equals (16) before taking the infimum. Recall that the selection rule

$$w(O_D;\alpha-\lambda)=\inf\{t\ge D:|X_t-\hat{X}_t|\ge v(\alpha-\lambda)\}.$$

is chosen to minimize function (16). We have

$$-g_\lambda(\alpha) = \inf_w \tilde{L}(w, \alpha - \lambda). \tag{112}$$

For each policy w, function $\tilde{L}(w, \alpha - \lambda)$ is a linear increasing function of α. Then by taking the infimum, function $\inf_w \tilde{L}(w, \alpha - \lambda)$ is continuous, concave and increasing. Therefore, function $g_\lambda(\alpha)$ is convex and monotonic decreasing.

When $\lambda^\star = 0$, according to Lemma 4, (22), we have $g_0(\alpha^\star) = 0$. The derivative at $\alpha^\star$ can be computed by:

$$g_0'(\alpha^\star) = o'(\alpha^\star) - \alpha^\star l'(\alpha^\star) - l(\alpha^\star) \overset{(a)}{=} -l(\alpha^\star), \tag{113}$$

where (a) is obtained because $v(\alpha^\star)$ is the optimum threshold the minimizes $\mathbb{E}[-O^2_{D+w} + \alpha^\star(D + w)]$ so that $o'(\alpha) - \alpha^\star l'(\alpha) = 0$. Then according to the convexity of $g_0(\cdot)$, the Taylor expansion at $\alpha^\star$ implies:

$$\begin{aligned} g_0(\alpha) \geq & g_0(\alpha^\star) - l(\alpha^\star)(\alpha - \alpha^\star) \\ & + \frac{1}{2} \min_{\alpha' \in [\alpha_{\text{lb}}, \alpha_{\text{ub}}]} g_0''(\alpha')(\alpha - \alpha^\star)^2. \end{aligned} \tag{114}$$

Since function $g_0(\alpha)$ is monotically decreasing and convex, by taking $N = \frac{1}{2} \min_{\alpha' \in [\alpha_{\text{lb}}, \alpha_{\text{ub}}]} g_0''(\alpha')$, we have:

$$g_0(\alpha) \geq -l(\alpha^\star)(\alpha - \alpha^\star) + N(\alpha - \alpha^\star)^2. \tag{115}$$

From the convexity of $g_0(\cdot)$, we have:

$$g_0(\alpha) \geq g_0(\alpha^\star) - l(\alpha^\star)(\alpha - \alpha^\star). \tag{116}$$

Then notice that $g_0(\alpha)$ is monotonic decreasing, $g_0'(\alpha^\star) < 0$, for $\alpha > \alpha^\star$, $g_0(\alpha) \leq 0$ and for $\alpha < \alpha^\star$, $g_0(\alpha) \geq 0$. Therefore we have

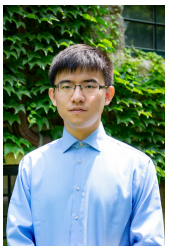
$$|g_0(\alpha)| \leq l(\alpha^\star)|\alpha - \alpha^\star|. \tag{117}$$

■

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

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 Postdoctoral Research Associate at Yale University from 2022-2023. She is currently a Post-Doctoral Research Associate at Meta AI. She was a selected participant at 2022 EECS Rising Stars workshop. Her research interests include age of information, stochastic network optimization, and statistical learning theory.

Jintao Wang (SM'12) received the B.Eng. and Ph.D. degrees in Electrical Engineering both from Tsinghua University, Beijing, China, in 2001 and 2006, respectively. From 2006 to 2009, he was an Assistant Professor in the Department of Electronic Engineering at Tsinghua University. Since 2009, he has been an Associate Professor and Ph.D. Supervisor. He is the Standard Committee Member for the Chinese national digital terrestrial television broadcasting standard. His current research interests include space-time coding, MIMO, and OFDM systems. He has published more than 100 journal and conference papers and holds more than 40 national invention patents.



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.