Sampling of the Wiener Process for Remote Estimation over a Channel with Unknown Delay Statistics

Haoyue Tang
Department of EE and YINS
Yale University
New Haven, CT, USA
haoyue.tang@yale.edu

Yin Sun
Department of ECE
Auburn University
Auburn, AL, USA
yzs0078@auburn.edu

Leandros Tassiulas
Department of EE and YINS
Yale University
New Haven, CT, USA
leandros.tassiulas@yale.edu

ABSTRACT

In this paper, we study an online sampling problem of the Wiener process. The goal is to minimize the mean squared error (MSE) of the remote estimator under a sampling frequency constraint when the transmission delay distribution is unknown. The sampling problem is reformulated into a renewal reward optimization problem, and we propose an online sampling algorithm that can adaptively learn the optimal sampling policy through stochastic approximation. We show that the cumulative MSE regret grows with rate $O(\ln k)$, where k is the number of samples. Through Le Cam's two point method, we show that the worst-case cumulative MSE regret of any online sampling algorithm is lower bounded by $\Omega(\ln k)$. Hence, the proposed online sampling algorithm is minimax order-optimal. Finally, we validate the performance of the proposed algorithm via numerical simulations.

CCS CONCEPTS

• **Networks → Network performance evaluation**; **Network performance evaluation**; Network performance analysis.

KEYWORDS

Age of Information, Online Learning, Stochastic Approximation

ACM Reference Format:
Haoyue Tang, Yin Sun, and Leandros Tassiulas. 2022. Sampling of the Wiener Process for Remote Estimation over a Channel with Unknown Delay Statistics. In *The Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '22), October 17–20, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3492866.3549732

1 INTRODUCTION

The omnipresence of the autonomous driving and the intelligent manufacturing systems involve tasks of sampling and remotely estimating fresh status information. For example, in autonomous driving systems, status information such as the position and the instant speed of cars keep changing, and the controller has to estimate the update-to-date status based on samples collected from the surrounding sensors. To ensure efficient control and system safety, it is important to estimate the fresh status information precisely under limited communication resources and random channel conditions.

To measure the freshness of the status update information, the Age of Information (AoI) metric has been proposed in [11]. By definition, AoI captures the difference between the current time and the time-stamp at which the freshest information available at the destination was generated. It is revealed that the AoI minimum sampling and transmission strategies behave differently from utility maximization and delay minimization [30]. Samples with fresher content should be delivered to the destination timely [22].

When the evolution of the dynamic source can be modeled by a random signal process, the mean square estimation error (MSE) based on the available information at the receiver can be used to capture freshness. Sampling to minimize the MSE of the random process in different communication networks are studied in [9, 15, 17, 21, 28]. When the random process can be observed at the sampler, the optimum sampling policy is shown to have a threshold structure, i.e., a new sample should be taken once the difference between the actual signal value and the estimate based on past samples exceed a certain threshold. The optimum threshold can be computed by iterative thresholding [27] or the bi-section search [21] if the delay distribution and the statistics of the channel are known in advance.

When the statistics of the communication channel is unknown, the problem of sampling and transmissions for data freshness optimization can be formulated into a sequential decision making problem [3–5, 14, 26]. By using the AoI as the freshness metric, [3–5] design online link rate selection algorithms based on stochastic bandits. When the channels are time-varying and the transmitter has an average power constraint, [1, 6, 7, 10, 13] employ reinforcement learning algorithms to minimize the average AoI under unknown channel statistics. Notice that in applications such as the remote estimation, a linear AoI cannot fully capture the data freshness. To solve this problem, Tripathi *et al.* model the information freshness to be a time-varying function of the AoI [26], and a robust online learning algorithm is proposed. The above research tackles with unknown packet loss rate or utility functions, the problem of designing online algorithms under unknown delay statistics are not well studied. The iterative thresholding algorithm proposed in [27] can be applied in the online setting when the delay statistics is unknown, whereas the convergence rate and the optimality of the algorithm are not well understood.

In this paper, we consider an online sampling problem, where a sensor transmits status updates of the Wiener source to a destination through a channel with random delay. Our goal is to design

---

MobiHoc '22, October 17–20, 2022, Seoul, Republic of Korea

ACM ISBN 978-1-4503-9165-8/22/10.
https://doi.org/10.1145/3492866.3549732

a sampling policy that minimizes the estimation error when the delay distribution is unknown a priori. The main contributions of this paper are as follows:

- The design of the MSE minimum sampling policy is reformulated as an optimal stopping problem. By analyzing the sufficient conditions of the optimum threshold, we propose an online sampling policy that learns the optimum threshold adaptively through stochastic approximation. Compared with [23, 24, 27], the operation of the proposed algorithm does not require prior knowledge of an upper bound of the optimum threshold.
- We prove that the time averaged MSE of the proposed algorithm converges almost surely to the minimum MSE if the fourth order moment of the transmission delay is bounded (Theorem 1). In addition, it is shown that the MSE regret, i.e., the sub-optimality gap between the expected cumulative MSE of the proposed algorithm and the optimum offline policy, grows at a speed of $O(\ln k)$, where k is the number of samples (Corollary 1). The perturbed ordinary differential equation (ODE) method is a popular tool for establishing the convergence rate of stochastic approximation algorithms [12]. However, this tool requires either the threshold being learned is in a bounded closed set, or the second moment of the updating directions are bounded. Because our algorithm does not require an upper bound on the optimum threshold, and the essential supremum of the transmission delay could be unbounded, we need to develop a new method for convergence rate analysis, which is based on the Lyapunov drift method for heavy traffic analysis.
- Further by using the classic Le Cam's two point method, we show that for any causal algorithm that makes sampling decision based on historical information, under the worst case delay distribution, the MSE regret is lower bounded by $\Omega(\ln k)$ (Theorem 4). By combining Theorem 1 and Theorem 4, we obtain that the proposed online sampling algorithm achieves the minimax order-optimal regret.
- We validate the performance of the proposed algorithm via numerical simulations. In contrast to [27], the proposed algorithm could meet an average sampling frequency constraint.

2 SYSTEM MODEL AND PROBLEM FORMULATION

2.1 System Model

As is depicted in Fig. 1, we revisit the status update system in [2, 21, 22], where a sensor takes samples from a Wiener process and transmits the samples to a receiver through a network interface queue. The network interface serves the update packets on the First-Come-First-Serve (FCFS) basis. An ACK is sent back to the sensor once an update packet is cleared at the interface. We assume that the transmission duration after passing the network interface is negligible.

Let $X_t \in \mathbb{R}$ denote the value of the Wiener process at time $t \in \mathbb{R}^+$. The sampling time-stamp of the k-th sample, denoted by S_k, is determined by the sensor at will. Based on the FCFS principle, the network interface will start serving the k-th packet after the $(k-1)$-th packet is cleared at the network interface and arrived at the receiver. We assume that the service time D_k are independent and identically distributed (i.i.d) with a probability distribution $\mathbb{P}_D$. The reception time of the k-th packet, denoted by R_k satisfies the following recursive formula: $R_k = \{S_k, R_{k-1}\} + D_k$ and we define $R_0 = 0$ for simplicity. We assume the average transmission delay $\overline{D} := \mathbb{E}_{D\sim\mathbb{P}_D}[D]$ is lower bounded by $\overline{D}_{\text{lb}} > 0$.

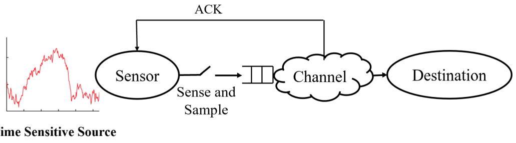

Figure 1: System model.

2.2 MMSE Estimation

Let $i(t) := \max_{k\in\mathbb{N}}\{k|R_k \leq t\}$ be the index of the latest sample received by the destination at time t. The information available at the receiver at time t can be summarized as follows: (i). The sampling time-stamps, transmission delay and the values of previous samples $\mathcal{M}_t := \{(S_j, D_j, X_{S_j})\}_{j=1}^{i(t)}$; (ii). The fact that no packet was received during $(R_{i(t)}, t]$. Similar to [20, 22], we assume that the receiver estimates X_t only based on $\mathcal{M}_t$ and neglects the second part of information. The minimum mean-square error (MMSE) estimator [18] in this case is:

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{M}_t] = X_{S_{i(t)}}. \tag{1}$$

We use a sequence of sampling time instants $\pi \triangleq \{S_k\}_{k=1}^{\infty}$ to represent a sampling policy. The expected time average mean square error (MSE) under π is denoted by $\overline{\mathcal{E}}_\pi$, i.e.,

$$\overline{\mathcal{E}}_\pi \triangleq \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_{t=0}^{T}\left(X_t - X_{S_{i(t)}}\right)^2 \mathrm{d}t\right]. \tag{2}$$

2.3 Problem Formulation

Our goal in this work is to design one sampling policy that can minimize the MSE for the estimator when the delay distribution $\mathbb{P}_D$ is unknown. Specifically, we focus on the set of causal policies denoted by Π, where each policy $\pi \in \Pi$ selects the sampling time S_k of the k-th sample based on the transmission delay $\{D_{k'}\}_{k'<k}$ and Wiener process evolution $\{X_t\}_{t\leq S_k}$ from the past. The transmission delay and the evolution of the Wiener process in the future cannot be used to decide the sampling time. Due to the energy constraint, we require that the sampling frequency should below a certain threshold. The optimal sampling problem is organized as follows:

Problem 1 (MMSE minimization).

$$\mathsf{mse}_{\text{opt}} \triangleq \inf_{\pi\in\Pi} \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_{t=0}^{T}\left(\hat{X}_t - X_t\right)^2 \mathrm{d}t\right], \tag{3a}$$

$$s.t. \quad \limsup_{T\to\infty} \mathbb{E}\left[\frac{i(T)}{T}\right] \leq f_{\max}. \tag{3b}$$

3 PROBLEM SOLUTION

In this section, the MSE minimization problem (i.e., Problem 1) is reformulated into an optimal stopping problem. Let $\pi^\star$ be an optimum policy whose average MSE achieves $\mathsf{mse}_{\text{opt}}$. Sufficient conditions for $\pi^\star$ are provided in Subsection 3.2. The online sampling algorithm π_{online} is provided in Subsection 3.3 and Subsection 3.4 characterizes the behaviors of the online sampling policy.

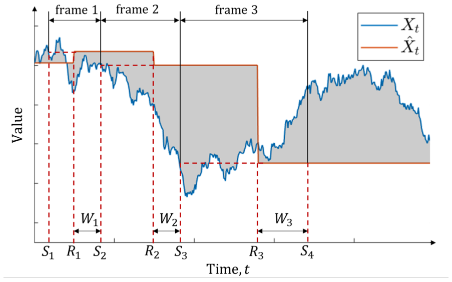

Figure 2: Illustration of the Wiener process and the estimation error. The sampling and reception time-stamp of the k-th sample are denoted by S_k and R_k, respectively. For MMSE estimator, $\hat{X}_t = X_{S_k}, \forall t \in [R_k, R_{k+1})$.

3.1 Markov Decision Reformulation

According to [21, Theorem 1], policy $\pi^\star$ should not take a new sample before the previous sample is delivered to the destination. As is depicted in Fig. 2, the waiting time between the delivery time of the k-th sample and the sampling time of the $(k+1)$-th sample is denoted by W_k. Define frame k as the time interval between the sampling time-stamp of the k-th and the $(k+1)$-th sample. The following corollary enables us to reformulate Problem 1 into a Markov Decision Process.

Lemma 1. *Let $\mathcal{I}_k := (D_k, (X_{S_k+t} - X_{S_k})_{t\geq 0})$ denote the recent information of the sampler in frame k. The set of sampling policies that determine the waiting time W_k only based on the recent information $\mathcal{I}_k$ is denoted by Π_{recent}. Since for each frame k, the difference $X_{S_k+t} - X_{S_k}$ evolves as a Wiener process that is independent of the past $\{X_{S_{k'}+t} - X_{S_{k'}}\}_{k'<k}$, Problem 1 can be reformulated into the following Markov decision process:*

Problem 2 (Markov Decision Process Reformulation).

$$\text{mse}_{\text{opt}} = \inf_{\pi\in\Pi_{\text{recent}}} \limsup_{K\to\infty} \left(\frac{\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right]}{\sum_{k=1}^{K} \mathbb{E}\left[(S_{k+1} - S_k)\right]} + \overline{D} \right), \quad \text{(4a)}$$

$$\text{s.t.} \quad \liminf_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left[(S_{k+1} - S_k)\right] \geq \frac{1}{f_{\max}}. \quad \text{(4b)}$$

The proof is provided in [25, Appendix E].

According to [21, Theorem 1], there exists a stationary policy $\pi^\star$ that selects the waiting time W_k using a conditional probability distribution given the recent $\mathcal{I}_k$. The average MSE $\overline{\mathcal{E}}_{\pi^\star}$ achieves mse_{opt}. Therefore, we can restrict our search of $\pi^\star$ within the set of stationary policies. Next, we will reveal the sufficient conditions for the optimal stationary policy.

3.2 Designing $\pi^\star$ with Known $\mathbb{P}_D$

Let Π_{cons} denote the set of policies that satisfy the sampling frequency constraint. Since $\pi^\star$ achieves the minimum expected time-average MSE among Π_{cons}, we have:

$$\limsup_{K\to\infty} \frac{\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right]}{\sum_{k=1}^{K} \mathbb{E}[D_k + W_k]} \geq \overline{\mathcal{E}}_{\pi^\star} - \overline{D}, \pi \in \Pi_{\text{cons}}. \quad (5)$$

For simplicity, denote $\gamma^\star := \overline{\mathcal{E}}_{\pi^\star} - \overline{D}$, which is the average cost of the MDP when the optimum policy $\pi^\star$ is used, i.e., $\gamma^\star = \limsup_{K\to\infty} \frac{\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right]}{\sum_{k=1}^{K}\mathbb{E}[D_k+W_k]}$. Because $\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[D_k+W_k] > 0$, for any policy $\pi \in \Pi_{\text{cons}}$, inequality (5) can be rewritten as:

$$\begin{aligned}\theta_\pi(\gamma^\star) := \liminf_{K\to\infty} \Bigg(&\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right] \\ &- \gamma^\star \cdot \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[D_k + W_k]\Bigg) \geq 0. \quad (6)\end{aligned}$$

Inequality (6) takes the minimum value 0 if and only if policy π is optimum. Therefore, if the ratio $\gamma^\star$ is known, an optimum policy $\pi^\star$ can be obtained by solving the following functional optimization:

Problem 3 (Functional Optimization Problem).

$$\begin{aligned}\text{mse}_{\text{opt}} = \inf_{\pi\in\Pi} \limsup_{K\to\infty} \Bigg(&\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{6}\left(X_{S_{k+1}} - X_{S_k}\right)^4\right] \\ &- \gamma^\star \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[(D_k + W_k)\right]\Bigg), \quad \text{(7a)}\end{aligned}$$

$$\text{s.t.} \quad \liminf_{K\to\infty} \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}(D_k + W_k)\right] \geq \frac{1}{f_{\max}}. \quad \text{(7b)}$$

To solve Problem 3, we can take the Lagrangian duality of the constraint (7b) with a dual variable ν and obtain the Lagrange function $\mathcal{L}(\pi, \gamma, \nu)$:

$$\begin{aligned}\mathcal{L}(\pi,\gamma,\nu) \triangleq \limsup_{K\to\infty} \Bigg(&\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right] \\ &- (\gamma+\nu)\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[(S_{k+1} - S_k)\right]\Bigg) + \nu\frac{1}{f_{\max}}. \quad (8)\end{aligned}$$

We say that a stationary policy π has a threshold structure, if the waiting time W_k is determined by:

$$W_k = \inf\{w \geq 0 \big| |X_{S_k+D_k+w} - X_{S_k}| \geq \tau\}. \quad (9)$$

Let Z_t be a Wiener process staring from $t = 0$. Let D be the transmission delay following distribution $\mathbb{P}_D$ and the value of the Wiener process at time D is denoted by Z_D. Using the threshold policy (9), the expected frame-length $L_k := D_k + W_k$ and $\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4$ has the following properties:

Lemma 2. *[21, Corollary 1 Restated]*

$$\mathbb{E}[L_k] = \mathbb{E}\left[\max\{\tau^2, Z_D^2\}\right], \quad \text{(10a)}$$

$$\mathbb{E}\left[\frac{1}{6}(X_{S_{k+1}} - X_{S_k})^4\right] = \frac{1}{6}\mathbb{E}\left[\max\{\tau^2, Z_D^2\}^2\right]. \quad \text{(10b)}$$

As is revealed by [21], the optimum policy $\pi^\star$ has a threshold structure as in equation (9). To design an off-line algorithm that can learn the updating threshold $\tau^\star$ of $\pi^\star$, we then reveal the necessary conditions that $\tau^\star$ should satisfy. With slightly abuse of notations, let $\mathcal{L}(\tau, \gamma, \nu)$ denote the expected value of the Lagrange function $\mathcal{L}(\pi, \gamma, \nu)$ when a stationary policy π with threshold τ is used. According to Lemma 2, $\mathcal{L}(\tau,\gamma,\nu)$ can be computed as follows:

$$\begin{aligned}\mathcal{L}(\tau,\gamma,\nu) =&\mathbb{E}\left[\frac{1}{6}\max\{\tau^2, Z_D^2\}^2\right] - (\gamma+\nu)\mathbb{E}[\max\{\tau^2, Z_D^2\}]\\ &+\nu\frac{1}{f_{\max}}.\end{aligned} \tag{11}$$

Condition 1: [21, Theorem 5 Restated] Let $\tau(\gamma,\nu)$ be the optimum sampling threshold that minimizes function $\mathcal{L}(\tau,\gamma,\nu)$, which can be computed as follows:

$$\tau(\gamma,\nu) := \arg\inf_{\tau\geq 0}\mathcal{L}(\tau,\gamma,\nu) = \sqrt{3(\gamma+\nu)}. \tag{12}$$

Condition 2: [21, Eq. (123, 125)]

$$\nu^\star\left(\mathbb{E}\left[\max\{3(\gamma^\star+\nu^\star), Z_D^2\}\right] - \frac{1}{f_{\max}}\right) = 0, \nu^\star \geq 0. \tag{13}$$

Recall that for any policy $\pi\in\Pi_{\text{cons}}$ with threshold τ, inequality (5) implies

$$\theta_\pi(\gamma^\star) = \frac{1}{6}\mathbb{E}\left[\max\{\tau^2, Z_D^2\}^2\right] - \gamma^\star\mathbb{E}\left[\max\{\tau^2, Z_D^2\}\right] \geq 0. \tag{14}$$

According to (12), inequality (14) holds with equality if and only if $\pi^\star$ with threshold $\tau^\star = \sqrt{3(\gamma^\star+\nu^\star)}$ is used. Adding the CS condition (13) on both sides of (14), the necessary condition for $\gamma^\star$ then becomes:

$$\overline{g}_\nu(\gamma^\star) = \theta_{\pi^\star}(\gamma^\star) = 0, \tag{15}$$

where function $\overline{g}_\nu(\gamma) := \mathbb{E}[g_\nu(\gamma; Z_D)]$ is the expectation of function $g_\nu(\gamma; Z_D)$ defined as follows:

$$g_\nu(\gamma; Z_D) := \frac{1}{6}\max\{3(\gamma+\nu), Z_D^2\}^2 - \gamma\max\{3(\gamma+\nu), Z_D^2\}. \tag{16}$$

As is shown by [21, Theorem 7], the duality gap between $\overline{\mathcal{E}}_{\pi^\star}$ and $\sup_{\nu\geq 0}\inf_\pi\mathcal{L}(\pi,\gamma^\star,\nu)$ is zero, and (15) becomes a necessary and sufficient condition.

3.3 An Online Algorithm π_{online}

When $\mathbb{P}_D$ is unknown but $\nu^\star$ is known, we can approximate $\gamma^\star$ by solving equation (15) through stochastic approximation [12, 16, 19]. Notice that the role of $\nu^\star$ is to satisfy the sampling frequency constraint. To achieve this goal, we approximate $\nu^\star$ by maintaining a sequence $\{U_k\}$ that records the sampling constraint violations up to frame k.

The algorithm is initialized by selecting $\gamma_1 = 0$ and $U_1 = 0$. In each frame k, the sampling and updating rules are as follows:

1. Sampling: We treat $\nu_k := \frac{1}{V}U_k^+$ as the dual optimizer ν, where $V > 0$ is fixed as a constant. The waiting time W_{k+1} is selected to minimize the Lagrange function (8), and according to the statement after equation (14), W_k is selected by:

$$W_k = \inf\{w\geq 0 |\left|X_{S_k+D_k+w} - X_{S_k}\right| \geq \sqrt{3\left(\gamma_k+\nu_k\right)}\}. \tag{17}$$

2. Update γ_k: To search for the root $\gamma > 0$ of equation $\overline{g}_{\nu_k}(\gamma) = 0$, we update γ_k through the Robbins-Monro algorithm [19]. In each frame k, we are given an i.i.d sample $\delta X_k = X_{S_k+D_k} - X_{S_k} \sim Z_D$, and the Robbins-Monro algorithm operates by:

$$\gamma_{k+1} = \left(\gamma_k + \eta_k Y_k\right)^+, \tag{18}$$

where $Y_k = g_{\nu_k}(\gamma_k; \delta X_k)$ and function $g_\nu(\cdot)$ is defined in (16). Recall that $\overline{D}_{\text{lb}}$ is a non-zero lower bound of the average delay, the stepsize $\{\eta_k\}$ is selected by:

$$\eta_k = \frac{1}{2\overline{D}_{\text{lb}}}k^{-\alpha}, \alpha\in(0.5, 1]. \tag{19}$$

3. Update U_k: To guarantee that the sampling frequency constraint is not violated, we update the violation U_k up to the end of frame k by:

$$U_{k+1} = U_k + \left(\frac{1}{f_{\max}} - (D_k + W_k)\right). \tag{20}$$

3.4 Theoretical Analysis

We analyze the convergence and optimality of algorithm π_{online}. We assume there is no sampling frequency constraint, i.e., $f_{\max} = \infty$ and make the following assumption on distribution $\mathbb{P}_D$:

ASSUMPTION 1. *The fourth order moment of the transmission delay is upper bounded by B, i.e.,*

$$\mathbb{E}[D^4] \leq B < \infty.$$

The convergence behavior of the optimum threshold $3\gamma^\star$ and the MSE performance are manifested in the following theorems:

THEOREM 1. *The proposed algorithm learns the optimum parameter $\gamma^\star$ almost surely, i.e.,*

$$\lim_{k\to\infty}\gamma_k = \gamma^\star, \quad \textit{w.p.1.} \tag{21}$$

The proof of Theorem 1 is obtained by the ODE method in [12, Chapter 5] and is provided in [25, Appendix A].

THEOREM 2. *The second moment of $(\gamma_k - \gamma^\star)$ satisfies:*

$$\sup_k\mathbb{E}\left[\frac{|\gamma_k - \gamma^\star|^2}{\eta_k}\right] < \infty. \tag{22}$$

Specifically, if $\alpha = 1$ and $\eta_k = \frac{1}{2\overline{D}_{\text{lb}}k}$, then the mean square error decays with rate $\mathbb{E}[(\gamma_k - \gamma^\star)^2] = O(1/k)$.

One challenge in the proof of Theorem 2 is that γ_k is unbounded and the second moment of Y_k is unbounded. We notice that Y_k could become very large when γ_k is much larger than the true value $\gamma^\star$, but the truncation of $(\gamma_k + \eta_k Y_k)^+$ to non-negative part actually prevents the actual update $|(\gamma_k + \eta_k Y_k)^+ - \gamma_k|$ from becoming too large. Based on this observation, we adopt a method from the heavy-traffic analysis by introducing the unused rate $\chi_k := (-(\gamma_k + \eta_k Y_k))^+$, then prove that the variance of the amount of the actual updating $(\eta_k Y_k + \chi_k)$ is finite. Detailed proofs are provided in Section 5.2.

THEOREM 3. *The average MSE under policy π_{online} converges to $\overline{\mathcal{E}}_{\pi^\star}$ almost surely, i.e.,*

$$\limsup_{k\to\infty}\frac{\int_{t=0}^{S_{k+1}}(X_t - \hat{X}_t)^2\mathrm{d}t}{S_{k+1}} = \overline{\mathcal{E}}_{\pi^\star}, \quad \textit{w.p.1.} \tag{23}$$

With the mean-square convergence of γ_k, the proof of Theorem 3 is a direct application of the perturbed ODE method [12] and is provided in [25, Appendix D].

By using Theorem 2 and Theorem 3, we can upper bound the growth rate of the cumulative MSE optimality gap in the following corollary:

Corollary 1. *If $\alpha = 1$, then the growth rate of the cumulative MSE optimality gap up to the k-th sample can be bounded as follows:*

$$\left(\mathbb{E}\left[\int_0^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] - \overline{\mathcal{E}}_{\pi^\star}\mathbb{E}[S_k]\right) = O\left(\ln k\right). \tag{24}$$

The proof of Corollary 1 is provided in [25, Appendix E].

Theorem 4. *For any distribution $\mathbb{P}$, let $\pi^\star(\mathbb{P})$ denote the MSE minimum sampling policy when the delay $D \sim \mathbb{P}$. The threshold obtained by solving equation (15) is denoted by $\gamma^\star(\mathbb{P})$. After k-samples are taken, the minimax estimation error $\gamma^\star(\mathbb{P})$ is lower bounded by:*

$$\inf_{\hat{\gamma}} \sup_{\mathbb{P}} \mathbb{E}\left[(\hat{\gamma} - \gamma^\star(\mathbb{P}))^2\right] = \Omega(1/k). \tag{25}$$

Let $p_w(\mathbb{P}) := \Pr(Z_D^2 \leq 3\gamma^\star(\mathbb{P})|D \sim \mathbb{P})$ denote the probability of waiting by using policy $\pi^\star(\mathbb{P})$ and let $\mathcal{P}_u(\mu) := \{\mathbb{P}|p_w(\mathbb{P}) \geq \mu\}$. Specifically, let $p^\star_{\mathrm{w,uni}} := \Pr(Z_D^2 \leq 3\gamma^\star_{\mathrm{uni}}|D \sim \mathrm{Uni}([0,1]))$. Let Π_h denote the set of policies which the sampling decision S_k is made based on historical information $\mathcal{H}_{k-1}$. We have the following result for $\mu \leq p^\star_{\mathrm{w,uni}}/2$:

$$\inf_{\pi\in\Pi_h} \sup_{\mathbb{P}\in\mathcal{P}_u(\mu)} \left(\mathbb{E}\left[\int_0^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t\right] - \overline{\mathcal{E}}_{\pi^\star(\mathbb{P})}\mathbb{E}[S_{k+1}]\right) \geq \frac{1}{2}\mu\cdot\Omega\left(\ln k\right). \tag{26}$$

As the transmission delay $\mathbb{P}_D$ considered in the paper does not belong to a specific family and could be quite general, obtaining a point-wise converse bound on $\mathbb{E}[(\hat{\gamma}-\gamma^\star(\mathbb{P}))^2]$ for each distribution $\mathbb{P}$ is impossible. As an alternative, a minimax risk bound $\mathbb{E}[(\hat{\gamma} - \gamma^\star(\mathbb{P}))^2]$ over a general distribution set $\mathcal{P}$ can be obtained using Le Cam's two point method for non-parametric estimation [29]. The core idea is to construct two distributions $\mathbb{P}_1, \mathbb{P}_2$, whose ℓ_1 distance $|\mathbb{P}_1^{\otimes k} - \mathbb{P}_2^{\otimes k}|_1$ can be upper bounded by a constant, but $(\gamma^\star(\mathbb{P}_1) - \gamma^\star(\mathbb{P}_2))^2 \geq \Omega(1/k)$ is difficult to distinguish. Such a construction is still challenging because $\gamma^\star(\mathbb{P})$ cannot be obtained in closed form even for the simpliest distribution families such as the delta distribution or exponential distribution. Notice that the estimation error of $\gamma^\star$ is closely related to the estimation error $\overline{g}_\nu(\cdot)$ at a given point. Therefore, the construction of $\mathbb{P}_1$ and $\mathbb{P}_2$ for obtaining the converse bound of Hölder smooth functions [29, Chapter 2] are adopted. The proof of inequality (26) is a direct application of the minimax estimation error (25). Detailed proof of Theorem 4 is provided in Section 5.3.

4 SIMULATION RESULTS

In this section, we provide simulation results to verify the theoretic findings and illustrate the performance of our proposed algorithms. We notice that the MSE minimization problem is closely related to the AoI minimization problem, where the AoI at time t, denoted by $A(t) = t - S_{i(t)}$. For signal-ignorant sampling policies (i.e., the sensor cannot always observe the time-varying process), according to the analysis in [22, Section IV-B], policies that minimize the average AoI achieves the minimum MSE. Therefore, we choose both offline and online AoI minimization policies ($\pi^\star_{\mathrm{AoI}}$ from [22], π_{itr} from [27]) for comparison. To show the convergence of online learning algorithm, we plotted the average MSE performance of the optimum off-line algorithm $\pi^\star$ from [21].

The transmission delay follows the log-normal distribution parameterized by μ and σ such that the density function of the probability measure $\mathbb{P}_D$ is:

$$p(x) := \frac{\mathbb{P}_D(\mathrm{d}x)}{\mathrm{d}x} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right).$$

In simulations, we set $\mu = 0.8$ and $\sigma = 1.2$, the expected time-averaged MSE is computed by taking the average of 20 runs. Fig. 3 depicts the time-average MSE performance up to the k-th frame of different sampling policies.

The asymptotic MSE behaviour is consistent with the convergence results in Theorem 3 and Corollary 1. When there is a sampling frequency constraint, the average MSE and the average sampling interval achieved by policy π_{online} are depicted in Fig. 4 and Fig. 5, respectively. We set $f_{\max} = \frac{1}{10D}$. From these figures, one can observe that the average MSE of π_{online} is close to the optimum MSE $\overline{\mathcal{E}}_{\pi^\star}$ and the sampling frequency can be satisfied. In addition, by choosing a larger V, a smaller MSE performance can be achieved, whereas a larger number of iterations are needed to meet the sampling frequency constraint.

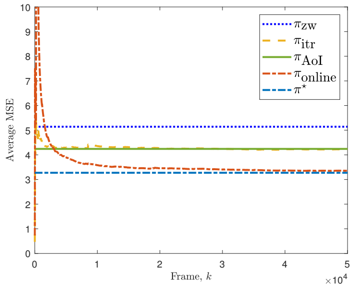

Figure 3: The time average MSE evolution as a function of frame k.

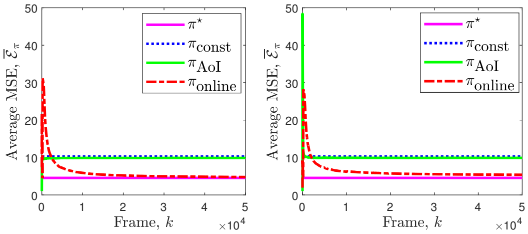

Figure 4: The time average MSE evolution as a function of frame k. (Left: $V = 10$, Right: $V = 1$.)

5 PROOFS OF MAIN RESULTS

5.1 Notations and Preliminary Lemmas

In Table 1, we summarize the notations used in the following proofs. Throughout the proofs, we use $N_1, N_2, \cdots$ to denote absolute constants and $C_1(\cdot), C_2(\cdot)$ to denote polynomials with finite order. For ease of exposition, the specific values and expressions of the constants and functions may vary across different context.

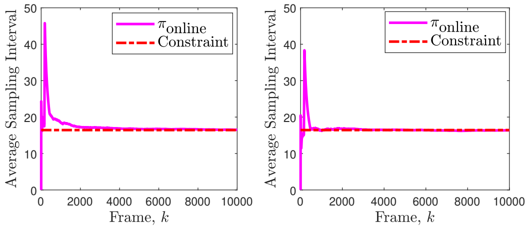

Figure 5: The average sampling interval under different constant V. (Left: $V = 10$, Right: $V = 1$.)

Table 1: Notations

Notation	Meaning
Z_t	a Wiener process staring from time 0
l_γ	length of running time using stopping rule $\tau_\gamma := \inf\{t \geq D \| \|Z_t\| \geq \sqrt{3\gamma}\}$
δX_k	$\delta X_k := X_{S_k+D_k} - X_{S_k}$
Q_k	$Q_k := \frac{1}{6}\left(X_{S_k+D_k} - X_{S_k}\right)^4$
L_k	$L_k := S_{k+1} - S_k = D_k + W_k$, frame length k
E_k	$E_k := \int_{S_k}^{S_{k+1}} (X_t - \hat{X}_t)^2 \mathrm{d}t$, cumulative estimation error in frame k
$q(\gamma)$	$q(\gamma) := \frac{1}{6}\mathbb{E}\left[\max\{3\gamma_k, Z_D^2\}^2\right]$, the expectation of Q_k when $\gamma_k = \gamma$
$l(\gamma)$	$l(\gamma) := \mathbb{E}[\max\{3\gamma, Z_D^2\}]$, expected frame length L_k when $\gamma_k = \gamma$
$\mathcal{I}_k$	$(D_k, (X_t - X_{S_k})_{S_k \leq t < S_{k+1}}$, information in frame k
$\mathcal{H}_k$	$\mathcal{H}_k := \{\mathcal{I}_\kappa\}_{\kappa \leq k}$ historical information up to the end of frame k
$\mathbb{E}_k[\cdot]$	Conditional expectation $\mathbb{E}[\cdot\|\mathcal{H}_{k-1}]$
t_k	$t_k := \sum_{i=1}^k \eta_k$ or $t_k := \sum_{i=1}^k \epsilon_k$, the cumulative step-sizes depending on the context
$m(t)$	$m(t)$ is the unique k so that $t_k \leq t \leq t_{k+1}$

Lemma 3. *Let $M := \mathbb{E}[D^2]$, the optimum ratio $\gamma^\star$ is upper and lower bounded by:*

$$\frac{1}{6}\overline{D} \leq \gamma^\star \leq \frac{1}{2}\frac{M + 2\overline{D}\frac{1}{f_{\max}} + \frac{1}{f_{\max}^2}}{\overline{D} + \frac{1}{f_{\max}}}. \tag{27}$$

The proof is provided in [25, Appendix F].

Lemma 4. *For threshold $\gamma < \infty$, the first, second and fourth order moments of the stopping time τ_γ are bounded, i.e.,*

$$\mathbb{E}[l_\gamma] \leq 3\gamma + \overline{D}, \tag{28a}$$

$$\mathbb{E}[l_\gamma^2] \leq \frac{10}{3}\left((3\gamma)^2 + 3\sqrt{B}\right), \tag{28b}$$

$$\mathbb{E}\left[l_\gamma^4\right] < 4^3\left((3\gamma)^4 + 105B\right) < \infty. \tag{28c}$$

The proof of Lemma 4 is provided in [25, Appendix G].

Lemma 5. *Function $\overline{g}_0(\gamma) = q(\gamma) - \gamma l(\gamma)$ and has the following properties:*

- (i) *$\overline{g}_0(\gamma)$ is concave and monotonically decreasing. The second order derivative $-3 \leq \overline{g}_0''(\gamma) \leq 0$.*
- (ii) *$\overline{g}_0(\gamma^\star) = 0$*
- (iii) *For $\gamma \neq \gamma^\star$, $(\gamma - \gamma^\star)\overline{g}_0(\gamma) \leq -l(\gamma^\star)(\gamma - \gamma^\star)^2 \leq 0$.*

The proof of Lemma 5 is provided in [25, Appendix H].

Corollary 2. *For each $\gamma_k < \infty$, if the fourth order moment of the delay satisfies $\mathbb{E}[D^4] < B < \infty$, given historical transmission $\mathcal{H}_{k-1}$, the conditional second order moment of the cumulative error in frame $E_k = \int_{S_k}^{S_{k+1}} (X_t - \hat{X}_t)^2 dt$ can be bounded as follows:*

$$\begin{aligned}\mathbb{E}_k[E_k^2] = \; & 3(X_{S_k} - X_{S_{k-1}})^2\sqrt{B} \\ & + 12C_1(\gamma_k, B)(X_{S_k} - X_{S_{k-1}})^2 + 3C_2(\gamma_k, B) < \infty,\end{aligned} \tag{29}$$

where C_1 and C_2 are fourth order polynomials of γ.

The proof of Corollary 2 is provided in [25, Appendix I].

With these Lemmas we can proceed to prove main results in Section 3:

5.2 Proof of Theorem 2

The analysis of the convergence rate is obtained through Lyapunov analysis, where the Lyapunov function is denoted by $V(\gamma) := \frac{1}{2}(\gamma - \gamma^\star)^2$. The proof is divided into two steps: first we will upper bound the Lyapunov drift for each γ_k by showing the following equation holds:

$$\mathbb{E}_k[V(\gamma_{k+1})] - V(\gamma_k) \leq -\eta_k \overline{D}_{\text{lb}} V(\gamma_k) + O(\eta_k^2 N_1). \tag{30}$$

Then, based on (30), we can then compute $\mathbb{E}[V(\gamma_k)]$ directly.

Step 1: Bounding the Lyapunov Drift: The analysis is divided into two cases, for $\gamma_k \leq 3\gamma^\star$, inequality (30) can be verified easily; for $\gamma_k \geq 3\gamma^\star$ we will first establish the relationship between $\mathbb{E}_k[V(\gamma_{k+1})] - V(\gamma_k)$ and $\text{Var}[Y_k]$, then upper bound $\text{Var}[Y_k]$ using the fact that Z_D^2 is sub-Gaussian when D is fourth order bounded.

Case 1: If $\gamma_k \leq 3\gamma^\star$, we have:

$$\begin{aligned}& \mathbb{E}_k[V(\gamma_{k+1})] - V(\gamma_k) \\ = & \mathbb{E}_k\left[\frac{1}{2}\left((\gamma_k + \eta_k Y_k)^+ - \gamma^\star\right)^2\right] - \frac{1}{2}(\gamma_k - \gamma^\star)^2 \\ \leq & \mathbb{E}_k\left[\frac{1}{2}(\gamma_k - \gamma^\star + \eta_k Y_k)^2 - \frac{1}{2}(\gamma_k - \gamma^\star)^2\right] \\ \overset{(a)}{=} & (\gamma_k - \gamma^\star)\eta_k \overline{g}_0(\gamma_k) \\ & + \frac{1}{2}\eta_k^2 \mathbb{E}_k\left[\left(\frac{1}{6}\max\{3\gamma_k, \delta X_k^2\}^2 - \gamma_k \max\{3\gamma_k, \delta X_k^2\}\right)^2\right] \\ \overset{(b)}{\leq} & -2\eta_k l(\gamma^\star)V(\gamma_k) \\ & + \frac{1}{2}\eta_k^2\left(\frac{1}{36}((9\gamma^\star)^4 + B) + (3\gamma^\star)^2((9\gamma^\star)^2 + 3\sqrt{B})\right),\end{aligned} \tag{31}$$

where equality (a) is because $\mathbb{E}_k[Y_k] = \mathbb{E}_k[g_0(\gamma_k; \delta X_k)] = \overline{g}_0(\gamma_k)$; inequality (b) is obtained because according to Lemma 5-(iii), $(\gamma_k - \gamma^\star)\overline{g}_0(\gamma_k) \leq -l(\gamma^\star)(\gamma_k - \gamma^\star)^2 = -2l(\gamma^\star)V(\gamma_k)$ and the assumption that $\gamma_k \leq 3\gamma^\star$.

Case 2: If $\gamma_k \geq 3\gamma^\star$, $\gamma_{k+1} = (\gamma_k + \eta_k Y_k)^+$ is truncated into the non-negative real number. We can view the evolution of γ_k as a queueing system, where the queue γ_k is non-negative, and $\eta_k Y_k$ is the arrival rate minus the service rate. Therefore, it is natural to introduce the "unused rate" from [8], which is denoted by $\chi_k := (-(\gamma_k + \eta_k Y_k))^+$. If $\chi_k = 0$, $(\gamma_k + \eta_k Y_k)\chi_k = 0 = -\chi_k^2$ and if $\chi_k \geq 0$, $\gamma_k + \eta_k Y_k = -\chi_k$, therefore

$$(\gamma_k + \eta_k Y_k)\chi_k = -\chi_k^2. \tag{32}$$

Since $\gamma_k + \eta_k Y_k + \chi_k \geq 0$, we have:

$$-\mathbb{E}_k[\gamma_k + \eta_k Y_k] \leq \mathbb{E}_k[\chi_k]. \tag{33}$$

We can then upper bound $\mathbb{E}_k[V(\gamma_{k+1}) - V(\gamma_k)]$ by:

$$\begin{aligned}& \mathbb{E}_k\left[V(\gamma_{k+1}) - V(\gamma_k)\right] \\ = & \mathbb{E}_k\left[\frac{1}{2}\left(\gamma_k - \gamma^\star + \eta_k Y_k + \chi_k\right)^2 - \frac{1}{2}\left(\gamma_k - \gamma^\star\right)^2\right] \\ = & \mathbb{E}_k\Bigg[\frac{1}{2}\left(\gamma_k - \gamma^\star + \eta_k Y_k\right)^2 - \frac{1}{2}(\gamma_k - \gamma^\star)^2 \\ & + \frac{1}{2}\chi_k^2 + (\gamma_k + \eta_k Y_k)\chi_k - \gamma^\star \chi_k\Bigg] \\ \overset{(c)}{=} & \mathbb{E}_k\left[\frac{1}{2}(\gamma_k - \gamma^\star + \eta_k Y_k)^2 - \frac{1}{2}(\gamma_k - \gamma^\star)^2 - \frac{1}{2}\chi_k^2 - \gamma^\star \chi_k\right]\end{aligned}$$

$$
\begin{aligned}
&\overset{(d)}{\leq} \frac{1}{2}\left(\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k]\right)^2-\frac{1}{2}(\gamma_k-\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
&\quad-\frac{1}{2}\mathbb{E}_k[\chi_k]^2-\gamma^\star\mathbb{E}_k[\chi_k]\\
&=\frac{1}{2}\left(\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k]\right)^2-\frac{1}{2}(\gamma_k-\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
&\quad-\frac{1}{2}\left(-\mathbb{E}_k[\chi_k]-\gamma^\star\right)^2+\frac{1}{2}(\gamma^\star)^2,
\end{aligned}\tag{34}
$$

where equality (c) is because equation (32); inequality (d) is obtained because $\mathbb{E}_k[\chi_k^2]\geq\mathbb{E}_k[\chi_k]^2\geq 0$;

We then further divide the analysis into two cases:

Case 2(a): If $\mathbb{E}_k[\gamma_k+\eta_kY_k]\leq\gamma^\star$, we then have $\mathbb{E}_k[\gamma_k-\gamma^\star+\eta_kY_k]\leq 0$. According to (33), $|-\mathbb{E}_k[\chi_k]-\gamma^\star|\geq|\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k]|$. Therefore, inequality (34) can be upper bounded by:

$$
\begin{aligned}
&\mathbb{E}_k\left[V(\gamma_{k+1})-V(\gamma_k)\right]\\
\leq&-\frac{1}{2}(\gamma_k-\gamma^\star)^2+\frac{1}{2}(\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
\overset{(e)}{\leq}&-\frac{1}{4}(\gamma_k-\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
\leq&-2\eta_k\overline{D}_{\mathrm{lb}}V(\gamma_k)+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k],
\end{aligned}\tag{35}
$$

where inequality (e) is obtained because $\frac{1}{4}(\gamma_k-\gamma^\star)^2\geq(\gamma^\star)^2\geq\frac{1}{2}(\gamma^\star)^2$ by assumption that $\gamma_k\geq 3\gamma^\star$ and the last inequality is obtained because $\eta_k\overline{D}_{\mathrm{lb}}\leq\frac{1}{2}$ by the step-size selection rule in equation (19).

Case 2(b): If $\mathbb{E}_k[\gamma_k+\eta_kY_k]\geq\gamma^\star$, considering that $\mathbb{E}_k[Y_k]=\overline{g}_0(\gamma_k)<0$ for $\gamma_k\geq\gamma^\star$, we have $0>\mathbb{E}_k[\eta_kY_k]\geq-(\gamma_k-\gamma^\star)$. Inequality (34) can be bounded by:

$$
\begin{aligned}
&\mathbb{E}_k[V(\gamma_{k+1})-V(\gamma_k)]\\
&\overset{(f)}{\leq}\frac{1}{2}(\gamma_k-\gamma^\star)(\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k])-\frac{1}{2}(\gamma_k-\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
&\leq\frac{1}{2}\eta_k(\gamma_k-\gamma^\star)\overline{g}_0(\gamma_k)+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
&\overset{(g)}{\leq}-\frac{1}{2}\eta_kl(\gamma^\star)(\gamma_k-\gamma^\star)^2+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k]\\
&=-\eta_kl(\gamma^\star)V(\gamma_k)+\frac{1}{2}\eta_k^2\mathrm{Var}[Y_k],
\end{aligned}\tag{36}
$$

where equality (f) is because $\left(-\mathbb{E}_k[\chi_k]-\gamma^\star\right)^2\geq(\gamma^\star)^2$ and $(\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k])^2\leq(\gamma_k-\gamma^\star+\eta_k\mathbb{E}_k[Y_k])(\gamma_k-\gamma^\star)$; inequality (g) is due to Lemma 5-(iii).

To proceed to show inequality (30) for $\gamma_k\geq 3\gamma^\star$, we need to upper bound $\mathrm{Var}[Y_k]$ in inequalities (35) and (36). First, we compute the expectation $\mathbb{E}[Y_k]$ as follows:

$$
\begin{aligned}
\mathbb{E}_k[Y_k]=&\mathbb{E}\left[\frac{1}{6}\max\{3\gamma_k,Z_D^2\}^2-\gamma_k\max\{3\gamma_k,Z_D^2\}\right]\\
=&-\frac{3}{2}\gamma_k^2+\mathbb{E}\left[(\frac{1}{6}Z_D^4-\gamma_kZ_D^2+\frac{3}{2}\gamma_k^2)\mathbb{I}_{(Z_D^2\geq3\gamma_k)}\right]\\
=&-\frac{3}{2}\gamma_k^2+\mathbb{E}\left[\frac{1}{6}(Z_D^2-3\gamma_k)^2\mathbb{I}_{(Z_D^2\geq3\gamma_k)}\right]\\
\leq&-\frac{3}{2}\gamma_k^2+\mathbb{E}\left[\frac{1}{6}(Z_D^2)^2\right]\\
\leq&-\frac{3}{2}\gamma_k^2+\frac{1}{2}\mathbb{E}[D^2]\leq-\frac{3}{2}\gamma_k^2+\frac{1}{2}\sqrt{B}.
\end{aligned}\tag{37}
$$

Given historical information $\mathcal{H}_{k-1}$, the variance of Y_k can be computed by:

$$
\begin{aligned}
&\mathrm{Var}[Y_k|\mathcal{H}_{k-1}]=\mathbb{E}_k\left[(Y_k-\mathbb{E}_k[Y_k])^2\right]\\
=&\mathbb{E}_k\left[\left(-\frac{3}{2}\gamma_k^2-\mathbb{E}_k[Y_k]\right)^2\mathbb{I}_{(Z_D^2\leq3\gamma_k)}\right]\\
&+\mathbb{E}_k\left[\left(\frac{1}{6}Z_D^4-\gamma_kZ_D^2+\frac{3}{2}\gamma_k^2+(-\frac{3}{2}\gamma_k^2-\mathbb{E}_k[Y_k])\right)^2\mathbb{I}_{(Z_D^2\geq3\gamma_k)}\right]\\
\overset{(h)}{\leq}&\frac{1}{4}B+2\mathbb{E}_k\left[\left(\frac{1}{6}Z_D^4-\gamma_kZ_D^2+\frac{3}{2}\gamma_k^2\right)^2\mathbb{I}_{(Z_D^2>3\gamma_k)}\right]\\
&+2\mathbb{E}_k\left[\left(-\frac{3}{2}\gamma_k^2-\mathbb{E}_k[Y_k]\right)^2\mathbb{I}_{(Z_D^2>3\gamma_k)}\right]\\
\leq&\frac{3}{4}B+\frac{1}{3}\mathbb{E}_k\left[(Z_D^2-3\gamma_k)^4\mathbb{I}_{(Z_D^2\geq3\gamma_k)}\right]\\
\leq&\frac{3}{4}B+\frac{1}{3}\mathbb{E}[Z_D^8]\leq(35+\frac{3}{4})B,
\end{aligned}\tag{38}
$$

where (h) is because $\mathbb{E}_k[Y_k]\leq-\frac{3}{2}\gamma_k^2+\frac{1}{2}\sqrt{B}$ implies $(-\frac{3}{2}\gamma_k^2-\mathbb{E}_k[Y_k])^2\leq\frac{1}{4}B$ and $(a+b)^2\leq2(a^2+b^2)$.

Denote $N_1:=\max\{(35+\frac{3}{4})B,\frac{1}{36}((9\gamma^\star)^4+B)+(3\gamma^\star)^2((9\gamma^\star)^2+3\sqrt{B})\}$, inequalities (31), (35) and (36) then lead to:

$$
\mathbb{E}_k[V(\gamma_{k+1})]-V(\gamma_k)\leq-\eta_k\overline{D}_{\mathrm{lb}}V(\gamma_k)+\eta_k^2N_1.\tag{39}
$$

Step 2: Computing $\mathbb{E}[V(\gamma_k)]$ through iteration Taking the expectation on both sides of (39), we have:

$$
\mathbb{E}[V(\gamma_{k+1})]\leq(1-\eta_k\overline{D}_{\mathrm{lb}})\mathbb{E}[V(\gamma_k)]+\eta_k^2N_1.\tag{40}
$$

Multiplying inequality (40) from $i=1$ to k yields:

$$
\mathbb{E}[V(\gamma_{k+1})]\leq\prod_{i=1}^{k}(1-\eta_i\overline{D}_{\mathrm{lb}})V(\gamma_0)+\sum_{i=1}^{k}\eta_i^2N_1\cdot\prod_{j=i+1}^{k}(1-\eta_j\overline{D}_{\mathrm{lb}}).\tag{41}
$$

Since the stepsize selected by (19) satisfies

$$
\eta_k\longrightarrow0,\liminf_k\min_{k\geq i\geq m(t_k-T)}\frac{\eta_k}{\eta_i}=1
$$

according to [12, p. 343, Eq. (4.8)], term $\prod_{i=1}^{k}(1-\eta_i\overline{D}_{\mathrm{lb}})=O(\eta_k)$. Therefore,

$$
\sup_k\mathbb{E}\left[\frac{(\gamma_k-\gamma^\star)^2}{\eta_k}\right]=\sup_k\mathbb{E}\left[2V(\theta_k)/\eta_k\right]=O(1).\tag{42}
$$

This finishes the proof of Theorem 2.

5.3 Proof of Theorem 4

5.3.1 Proof of Inequality (25). Let $\mathbb{P}_1,\mathbb{P}_2$ be two delay distributions and let $\gamma_1^\star,\gamma_2^\star$ be the solution to (15) when $D\sim\mathbb{P}_1$ and $D\sim\mathbb{P}_2$, respectively. Through Le Cam's inequality [31], we have:

$$
\inf_{\hat{\gamma}}\sup_{\mathbb{P}}\mathbb{E}\left[\left(\hat{\gamma}-\gamma^\star(\mathbb{P})\right)^2\right]\geq(\gamma_1^\star-\gamma_2^\star)^2\cdot\left(\mathbb{P}_1^{\otimes k}\wedge\mathbb{P}_2^{\otimes k}\right),\tag{43}
$$

where $\mathbb{P}\wedge\mathbb{Q}:=\int_\Omega\min\{p(x),q(x)\}\mathrm{d}x$ and $\mathbb{P}^{\otimes k}$ is the product of distribution k i.i.d random variables drawn from $\mathbb{P}$.

To use Le Cam's inequality (65), we need to find two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$, whose ℓ_1 distance $|\mathbb{P}_1^{\otimes k}-\mathbb{P}_2^{\otimes k}|_1$ is bounded, and the difference $(\gamma_1^\star-\gamma_2^\star)^2$ is of order $1/k$. We consider $\mathbb{P}_1$ to be a uniform distribution on $[0, 1]$ and let $\gamma_1^\star$ be the optimum ratio of distribution $\mathbb{P}_1$. Through Corollary 3, we can obtain a loose upper bound on $\gamma_1^\star$ as follows:

$$\gamma_1^\star < \frac{1}{2}\frac{\mathbb{E}[D^2]}{\mathbb{E}[D]} = \frac{1}{3}. \tag{44}$$

Let $c \le \frac{1}{2}$ be a constant and we denote

$$\delta = \min\{1 - 3\gamma_1^\star, 1/3, p^\star_{\text{w, uni}}/2\}. \tag{45}$$

Let $\mathbb{P}_2$ be a probability distribution with probability density function $p_2(x)$ defined as follows:

$$p_2(x) = \begin{cases} 1 - c\sqrt{1/k}, & x \le \frac{1}{2}\delta; \\ 1, & \frac{1}{2}\delta < x \le 1 - \frac{1}{2}\delta; \\ 1 + c\sqrt{1/k}, & x > 1 - \frac{1}{2}\delta; \\ 0, & \text{otherwise}. \end{cases} \tag{46}$$

We will first bound $(\gamma_1^\star - \gamma_2^\star)^2$ (in Step 1) and $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$ (in Step 2) as follows:

Step 1: Lower bounding $\gamma_2^\star - \gamma_1^\star$: For notational simplicity, denote function $h_1(\gamma) := \mathbb{E}_{D\sim\mathbb{P}_1}[\frac{1}{6}\max\{3\gamma, Z_D^2\}^2 - \gamma\max\{3\gamma, Z_D^2\}]$ and $h_2(\gamma) := \mathbb{E}_{D\sim\mathbb{P}_2}[\frac{1}{6}\max\{3\gamma, Z_D^2\}^2 - \gamma\max\{3\gamma, Z_D^2\}]$. According to the definition of $\mathbb{P}_2$ in (46), for each γ, the difference between $h_1(\gamma)$ and $h_2(\gamma)$ can be computed by:

$$\begin{aligned} & h_2(\gamma) - h_1(\gamma) \\ =& \int_{1-\delta/2}^{1} \frac{c}{\sqrt{k}}\mathbb{E}\left[\frac{1}{6}\max\{3\gamma, Z_D^2\}^2 - \gamma\max\{3\gamma, Z_D^2\}\middle|D = x\right]\mathrm{d}x \\ & - \int_0^{\delta/2} \frac{c}{\sqrt{k}}\mathbb{E}\left[\frac{1}{6}\max\{3\gamma, Z_D^2\}^2 - \gamma\max\{3\gamma, Z_D^2\}\middle|D = x\right]\mathrm{d}x \\ \overset{(a)}{=}& \int_{1-\delta/2}^{1} \frac{c}{\sqrt{k}}\mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma)}|D = x\right]\mathrm{d}x \\ & - \int_0^{\delta/2} \frac{c}{\sqrt{k}}\mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma)}|D = x\right]\mathrm{d}x, \end{aligned} \tag{47}$$

where inequality (a) is obtained because

$$\frac{1}{6}\max\{3\gamma, Z_D^2\}^2 - \gamma\max\{3\gamma, Z_D^2\} = -\frac{3}{2}\gamma^2 + \frac{1}{6}(Z_D^2 - 3\gamma)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma)}. \tag{48}$$

Since $\gamma_1^\star$ is the optimum ratio for delay distribution $\mathbb{P}_1$, we have $h_1(\gamma_1^\star) = 0$. According to equation (47), function $h_2(\gamma_1^\star)$ can be lower bounded by:

$$\begin{aligned} & h_2(\gamma_1^\star) \\ \overset{(b)}{\ge}& \frac{c}{\sqrt{k}} \cdot \int_{1-\delta/2}^{1} \mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma_1^\star)}\middle|D = x\right]\mathrm{d}x \\ & - \int_0^{\delta/2} \frac{c}{\sqrt{k}}\frac{1}{2}x^2\mathrm{d}x \\ \ge& \frac{c}{\sqrt{k}} \cdot \int_{1-\delta/2}^{1} \mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma_1^\star)}\middle|D = x\right]\mathrm{d}x - \frac{c}{\sqrt{k}}\frac{1}{6}\left(\frac{\delta}{2}\right)^3. \end{aligned} \tag{49}$$

where inequality (b) is because $\mathbb{E}[\frac{1}{6}(Z_D^2 - 3\gamma)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma)}|D = x] \le \mathbb{E}[\frac{1}{6}Z_D^4|D = x] = \frac{1}{2}x^2$.

We then proceed to lower bound $\mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma_1^\star)}|D = x\right]$ for each delay realization $x \in [1 - \delta/2, 1]$ as follows:

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(Z_D^2 \ge 3\gamma_1^\star)}|D = x\right] \\ \overset{(c)}{\ge}& \mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(3\gamma_1^\star \le Z_D^2 \le x)} + \frac{1}{6}(Z_D^2 - x)^2\mathbb{I}_{(Z_D^2 \ge x)}|D = x\right] \\ \ge& \mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(3\gamma_1^\star \le Z_D^2 \le x)}\right] \\ & + \frac{1}{6}\left(\mathrm{Var}[Z_D^2|D = x] - x^2\mathrm{Pr}\left(Z_D^2 \le x|D = x\right)\right) \\ \overset{(d)}{\ge}& \frac{1}{6}x^2 \ge \frac{1}{6}(1 - \delta/2)^2, \end{aligned} \tag{50}$$

where inequality (c) is because $\delta \ge 1 - 3\gamma_1^\star$ by equation (45), and for the conditional mean $\mathbb{E}[Z_D^2|D = x] = x \ge 1 - \delta/2 \ge 1 - \delta \ge 3\gamma_1^\star$; inequality (d) is because $\mathrm{Var}[Z_D^2|D = x] = 2x^2$ and $x^2\mathrm{Pr}(Z_D^2 \le x) \le x^2$ and $\mathbb{E}\left[\frac{1}{6}(Z_D^2 - 3\gamma_1^\star)^2\mathbb{I}_{(3\gamma_1^\star \le Z_D^2 \le x)}\right] \ge 0$. Plugging inequality (50) into (49) and recall that $\delta < 1$ by definition, we have the lower bound of $h_2(\gamma_1^\star)$:

$$h_2(\gamma_1^\star) \ge \frac{c}{\sqrt{k}}\frac{\delta}{2}\frac{1}{6}\left(\left(1 - \frac{\delta}{2}\right)^2 - \left(\frac{\delta}{2}\right)^2\right) \ge \frac{c}{\sqrt{k}}\frac{\delta}{12}(1 - \delta) > 0. \tag{51}$$

By Lemma 5-(i), function $h_2(\cdot)$ is monotonically decreasing. Since $h_2(\gamma_1^\star) > 0$ and $h_2(\gamma_2^\star) = 0$, we can conclude that $\gamma_2^\star \ge \gamma_1^\star$. We then proceed to bound $\gamma_2^\star - \gamma_1^\star$ through Taylor expansion at $\gamma = \gamma_1^\star$.

$$h_2(\gamma_2^\star) = h_2(\gamma_1^\star) + h_2'(\gamma)(\gamma_2^\star - \gamma_1^\star), \tag{52}$$

where $\gamma \in [\gamma_1^\star, \gamma_2^\star]$. Therefore, $\gamma_2^\star$ can be computed by:

$$\gamma_2^\star - \gamma_1^\star = -\frac{h_2(\gamma_1^\star)}{h_2'(\gamma)}. \tag{53}$$

To lower bound $\gamma_2^\star$, we will first find a loose upper bound of $\gamma_2^\star$ using Lemma 3:

$$\gamma_2^\star \le \frac{1}{2}\frac{\mathbb{E}_{D\sim\mathbb{P}_2}[D^2]}{\mathbb{E}_{D\sim\mathbb{P}_2}[D]} \le \frac{1}{2}\left(\frac{1}{3} + \delta \cdot c\sqrt{1/k}\right), \tag{54}$$

Therefore, since $\delta < 1/3$, we have $|h_2'(\gamma)| \le |h_2'(\gamma_2^\star)| = \mathbb{E}[\max\{3\gamma_2^\star, Z_D^2\}] \le \overline{D} + 3\gamma_{2,\text{ ub}} \le 1 + \frac{1}{2} + \frac{3}{2}c\sqrt{\frac{1}{k}}\delta \le 2$. Then by inequality (51), we have

$$\gamma_2^\star - \gamma_1^\star \ge \frac{-h_2(\gamma_1^\star)}{h_2'(\gamma_2^\star)} \ge \frac{1}{24}(1 - \delta)\delta c\sqrt{\frac{1}{k}}. \tag{55}$$

Step 2: Lower bounding $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$: Let $|\mathbb{P} - \mathbb{Q}| = \int_\Omega |\mathrm{d}\mathbb{P} - \mathrm{d}\mathbb{Q}|$ be the ℓ_1 distance between probability distribution $\mathbb{P}$ and $\mathbb{Q}$. Then

$$\begin{aligned} \mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k} =& \int \min\{\mathbb{P}_1^{\otimes k}(\mathrm{d}x), \mathbb{P}_2^{\otimes k}(\mathrm{d}x)\} \\ =& \int \mathbb{P}_1^{\otimes k}(\mathrm{d}x) \cdot \left(1 - \frac{\left(\mathbb{P}_2^{\otimes k}(\mathrm{d}x) - \mathbb{P}_1^{\otimes k}(\mathrm{d}x)\right)^+}{\mathbb{P}_1^{\otimes k}(\mathrm{d}x)}\right) \\ =& 1 - \int \left(\mathbb{P}_2^{\otimes k}(\mathrm{d}x) - \mathbb{P}_1^{\otimes k}(\mathrm{d}x)\right)^+ \\ =& 1 - \frac{1}{2}|\mathbb{P}_1^{\otimes k} - \mathbb{P}_2^{\otimes k}|_1. \end{aligned} \tag{56}$$

Equality (56) enables us to lower bound $\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}$ by upper bounding the ℓ_1 distance $|\mathbb{P}_1^{\otimes k} - \mathbb{P}_2^{\otimes k}|_1$, which is done through the Pinsker's inequality:

$$\begin{aligned}
\frac{1}{2}\left|\mathbb{P}_1^{\otimes k} - \mathbb{P}_2^{\otimes k}\right|_1 &\leq \sqrt{\frac{1}{2}D_{\mathsf{KL}}(\mathbb{P}_2^{\otimes k}||\mathbb{P}_1^{\otimes k})} = \sqrt{\frac{1}{2}kD_{\mathsf{KL}}(\mathbb{P}_2||\mathbb{P}_1)} \\
&\overset{(e)}{\leq} \sqrt{\frac{1}{2}k\int_0^1 p_2(x)\ln p_2(x)\mathrm{d}x} \\
&\overset{(f)}{\leq} \sqrt{\frac{1}{2}k\int_0^1 \left(p_2(x) - 1 + \frac{1}{\min\{p_2(x), 1\}}(p_2(x)-1)^2\right)\mathrm{d}x} \\
&\overset{(g)}{\leq} \sqrt{\frac{1}{2}k\frac{1}{\inf_{0\leq d\leq 1} p_2(d)}\int_0^1 (p_2(x)-1)^2\mathrm{d}x} \\
&\leq \sqrt{\frac{1}{2}k\frac{1}{1-c\sqrt{1/k}}\delta\frac{c^2}{k}} \leq \sqrt{\delta c^2},
\end{aligned} \tag{57}$$

where inequality (e) is because the density function $p_1(x) = 1$ for uniform distribution, therefore $D_{\mathsf{KL}}(\mathbb{P}_2||\mathbb{P}_1) = \int_0^1 p_2(x)\ln p_2(x)$; inequality (f) is because function $g(t) := (t\ln t)$ is convex, its derivative $g(t)'' = 1/t$, therefore, through Taylor expansion we have $g(t) \leq g(1)+(t-1)+\frac{1}{2}\frac{1}{\min\{t,1\}}(t-1)^2 = (t-1)+\frac{1}{2}\frac{1}{\min\{t,1\}}(t-1)^2$; inequality (g) is because $\int_0^1 p_2(x)\mathrm{d}x = 1$.

By choosing $c = 1/2$ and recall that $\delta < 1$, inequality (57) can be upper bounded by:

$$\frac{1}{2}|\mathbb{P}_1^{\otimes k} - \mathbb{P}_2^{\otimes k}|_1 \leq \frac{1}{2}. \tag{58}$$

Plugging (58) into (56) yields:

$$\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k} \geq \frac{1}{2}. \tag{59}$$

Finally, plugging (59) and (55) into the Le Cam's inequality (43) finishes the proof of inequality (25):

$$\inf_{\hat{\gamma}}\sup_{\mathbb{P}}(\hat{\gamma} - \gamma^\star(\mathbb{P}))^2 \geq \frac{1}{2}\left(\frac{1}{24}(1-\delta)\delta p^\star_{\mathrm{w,uni}}\right)^2 \cdot \frac{1}{k}. \tag{60}$$

5.3.2 *Proof of Inequality* (26).

The proof is divided into three steps: consider a delay distribution $\mathbb{P} \in \mathcal{P}_u(\mu)$, first we will show for each sample policy π with a random sampling interval τ, let $l_\pi := \mathbb{E}[\tau] = \mathbb{E}[Z_\tau^2]$ denote the expected running length, the following inequality holds:

$$\mathbb{E}\left[\int_{t=0}^{\tau} Z_t^2\mathrm{d}t\right] - \gamma^\star\mathbb{E}[\tau] \geq \frac{1}{6}p_w(\mathbb{P})\left(l_\pi - l^\star(\mathbb{P})\right)^2, \tag{61}$$

where $l^\star(\mathbb{P}) := \mathbb{E}_{D\sim\mathbb{P}}[\max\{3\gamma^\star(\mathbb{P}), Z_D^2\}]$ is the average frame length when the optimum policy $\pi^\star(\mathbb{P})$ is used. Next, We will show that given k samples $\delta X_1, \cdots, \delta X_k \overset{\text{i.i.d}}{\sim} Z_D$, where $D \sim \mathbb{P}$, the minimax estimation error satisfies:

$$\inf_{\hat{l}}\sup_{\mathbb{P}\in\mathcal{P}_u(\mu)}\mathbb{E}\left[\left(\hat{l} - l^\star(\mathbb{P})\right)^2\right] \geq N \cdot \frac{1}{k}, \tag{62}$$

where N is a constant independent of k and μ with expressions provided in equation (60).

Finally, notice that:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{S_k}(X_t - \hat{X}_t)^2\mathrm{d}t - (\gamma^\star + \overline{D})S_k\right] \\
&= \sum_{k'=1}^{k}\mathbb{E}\left[(X_{S_{k'+1}} - X_{S_{k'}})^2 D_k + \frac{1}{6}(X_{S_{k'+1}} - X_{S_{k'}})^4\right] \\
&\quad - (\gamma^\star + \overline{D})\sum_{k'=1}^{k}\mathbb{E}[S_{k'+1} - S_{k'}] \\
&\geq \frac{1}{6}p_w(\mathbb{P})\sum_{k'=1}^{k}\left(\mathbb{E}[L_k] - l^\star\right)^2.
\end{aligned} \tag{63}$$

Take $\inf_\pi \sup_{\mathbb{P}\in\mathcal{P}_w(\mu)}$ on both sides of inequality (63), we have:

$$\begin{aligned}
&\min_{\pi}\max_{\mathbb{P}\in\mathcal{P}_u(\mu)}\mathbb{E}\left[\int_0^{S_k}(X_t - \hat{X}_t)^2\mathrm{d}t - (\gamma^\star + \overline{D})S_k\right] \\
&\geq \frac{1}{6}\mu\inf_{\hat{l}}\sup_{\mathbb{P}\in\mathbb{P}_u(\mu)}\mathbb{E}[(\hat{l} - l^\star(\mathbb{P}))^2] \geq \frac{1}{6}\mu N\sum_{k'=1}^{k}\frac{1}{k'} \geq \Omega(\ln k).
\end{aligned} \tag{64}$$

Proof of inequality (61) follows similar ideas as [23, Lemma 4]. Details are provided in [25, Appendix K] due to space limitations. The proof of the minimax risk bound (62) is based on Le Cam's two point method as follows:

5.3.3 *Proof of inequality* (62).

Let $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}_w(\mu)$, through Le Cam's inequality [31], we have:

$$\inf_{\hat{l}}\sup_{\mathbb{P}\in\mathcal{P}_u(\mu)}\mathbb{E}\left[\left(\hat{l} - l^\star(\mathbb{P})\right)^2\right] \geq (l_1^\star - l_2^\star)^2 \cdot \left(\mathbb{P}_1^{\otimes k} \wedge \mathbb{P}_2^{\otimes k}\right). \tag{65}$$

Similarly to the proof of bounding $(\hat{\gamma} - \gamma^\star(\mathbb{P}))$, let $\mathbb{P}_1$ be a uniform distribution and $\mathbb{P}_2$ through the density function in equation (46). For $\mu \leq p^\star_{\text{w, uni}}/2$, it is easy to show that $p_w(\mathbb{P}_2) \in \mathcal{P}_u(\mu)$ as follows:

$$\begin{aligned}
&\Pr(Z_D^2 \leq 3\gamma_2^\star|D \sim \mathbb{P}_2) \\
&= \int_0^\infty \Pr(Z_D^2 \leq 3\gamma_2^\star|D = x)p_2(x)\mathrm{d}x \\
&= \int_0^\infty \Pr(Z_D^2 \leq 3\gamma_2^\star|D = x)p_1(x)\mathrm{d}x \\
&\quad - \int_0^{\delta/2}\Pr(Z_D^2 \leq 3\gamma_2^\star|D = x)\frac{c}{\sqrt{k}}\mathrm{d}x \\
&\quad + \int_{1-\delta/2}^{1}\Pr(Z_D^2 \leq 3\gamma_2^\star|D = x)\frac{c}{\sqrt{k}}\mathrm{d}x \\
&\overset{(i)}{\geq} \int_0^\infty \Pr(Z_D^2 \leq 3\gamma_1^\star|D = x)p_1(x)\mathrm{d}x - c/\sqrt{k}\delta \\
&\geq p^\star_{\text{w, uni}} - c/\sqrt{k}\delta \overset{(j)}{\geq} p^\star_{\text{w, uni}}/2.
\end{aligned} \tag{66}$$

where inequality (i) holds because $\gamma_1^\star \leq \gamma_2^\star$ from inequality (55); inequality (j) holds because $\delta < p^\star_{\text{w, uni}}/2$ by definition from equation (45). The difference between $l_2^\star - l_1^\star$ can be computed by:

$$\begin{aligned}
l_2^\star - l_1^\star = &\int_0^1 \mathbb{E}\left[\max\{3\gamma_2^\star, Z_D^2\}|D = x\right]\mathrm{d}x \\
&+ \int_{1-\delta/2}^{1}\frac{c}{\sqrt{k}}\mathbb{E}\left[\max\{3\gamma_2^\star, Z_D^2\}|D = x\right]\mathrm{d}x \\
&- \int_0^{\delta/2}\frac{c}{\sqrt{k}}\mathbb{E}\left[\max\{3\gamma_2^\star, Z_D^2\}|D = x\right]\mathrm{d}x \\
&- \int_0^1 \mathbb{E}\left[\max\{3\gamma_1^\star, Z_D^2\}|D = x\right]\mathrm{d}x.
\end{aligned} \tag{67}$$

Notice that if $x_1 \geq x_2$,

$$\mathbb{E}[\max\{3\gamma, Z_D^2\}|D = x_1] - \mathbb{E}[\max\{3\gamma, Z_D^2\}|D = x_2] \geq 0. \tag{68}$$

Therefore, inequality (67) can be bounded by:

$$\begin{aligned} l_2^\star - l_1^\star \geq & \int_0^1 \mathbb{E}\left[\max\{3\gamma_2^\star, Z_D^2\}|D = x\right] \mathrm{d}x \\ & - \int_0^1 \mathbb{E}\left[\max\{3\gamma_1^\star, Z_D^2\}|D = x\right] \mathrm{d}x \\ \geq & 3(\gamma_2^\star - \gamma_1^\star)\mathbb{E}_{D\sim\mathbb{P}_1}[\Pr(Z_D^2 \leq 3\gamma_1^\star)] \\ \overset{(g)}{\geq} & \frac{1}{24}(1-\delta)\delta c p_{\mathrm{w,\,uni}}^\star \sqrt{\frac{1}{k}}, \end{aligned} \tag{69}$$

where inequality (g) is obtained by equation (55).

Finally, plugging (58) and (69) into Le Cam's inequality (65) finishes the proof of inequality (62):

$$\inf_{\hat{l}} \sup_{\mathbb{P}\in\mathcal{P}_w(\mu)} (\hat{l} - l^\star(\mathbb{P}))^2 \geq \frac{1}{2}\left(\frac{1}{48}(1-\delta)\delta p_{\mathrm{w,uni}}^\star\right)^2 \cdot \frac{1}{k}. \tag{70}$$

6 CONCLUSIONS

In this work, we studied the problem of sampling a Wiener process for remote estimation over a channel with unknown delay statistics. By reformulating the MSE minimization problem as a renewal-reward process, we proposed an online sampling algorithm that can adaptively learn the optimum algorithm as the number of samples grows. We showed that the average MSE obtained by the proposed algorithm converges to the minimum MSE almost surely, and the cumulative MSE has an order of $O(\ln k)$, where k is the number of samples. We then prove that the cumulative MSE regret of any algorithm is at best $\Omega(\ln k)$. Numerical simulation results validate the convergence behaviors of the proposed algorithm.

ACKNOWLEDGMENTS

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. The work of Y. Sun was supported by the NSF grant CCF-1813078 and the ARO grant W911NF-21-1-0244.

REFERENCES

[1] Mohamed A. Abd-Elmagid, Harpreet S. Dhillon, and Nikolaos Pappas. 2020. A Reinforcement Learning Framework for Optimizing Age of Information in RF-Powered Communication Systems. *IEEE Transactions on Communications* 68, 8 (2020), 4747–4760.

[2] Ahmed Arafa, Roy D. Yates, and H. Vincent Poor. 2019. Timely Cloud Computing: Preemption and Waiting. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 528–535.

[3] Eray Unsal Atay, Igor Kadota, and Eytan Modiano. 2020. Aging Bandits: Regret Analysis and Order-Optimal Learning Algorithm for Wireless Networks with Stochastic Arrivals. arXiv:2012.08682 [eess.SY]

[4] Subhankar Banerjee, Rajarshi Bhattacharjee, and Abhishek Sinha. 2020. Fundamental Limits of Age-of-Information in Stationary and Non-stationary Environments. In *2020 IEEE International Symposium on Information Theory (ISIT)*. 1741–1746.

[5] Kavya Bhandari, Santosh Fatale, Urvidh Narula, Sharayu Moharir, and Manjesh Kumar Hanawal. 2020. Age-of-Information Bandits. In *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. 1–8.

[6] Elif Tuğçe Ceran, Deniz Gündüz, and András György. 2019. Reinforcement Learning to Minimize Age of Information with an Energy Harvesting Sensor with HARQ and Sensing Cost. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 656–661.

[7] Elif Tuğçe Ceran, Deniz Gündüz, and András György. 2021. A Reinforcement Learning Approach to Age of Information in Multi-User Networks With HARQ. *IEEE Journal on Selected Areas in Communications* 39, 5 (2021), 1412–1426.

[8] Atilla Eryilmaz and Rayadurgam Srikant. 2012. Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems* 72, 3 (2012), 311–359.

[9] B. Hajek, K. Mitzel, and S. Yang. 2003. Paging and registration in cellular networks: jointly optimal policies and an iterative algorithm. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, Vol. 1. 524–532 vol.1.

[10] Clement Kam, Sastry Kompella, and Anthony Ephremides. 2019. Learning to Sample a Signal through an Unknown System for Minimum AoI. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 177–182.

[11] Sanjit Kaul, Roy Yates, and Marco Gruteser. 2012. Real-time status: How often should one update?. In *2012 Proceedings IEEE INFOCOM*. 2731–2735.

[12] Harold J. Kushner and G. George Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer New York, New York, NY. https://doi.org/10.1007/0-387-21769-X_5

[13] Shiyang Leng and Aylin Yener. 2019. Age of Information Minimization for Wireless Ad Hoc Networks: A Deep Reinforcement Learning Approach. In *2019 IEEE Global Communications Conference (GLOBECOM)*. 1–6.

[14] Bin Li. 2021. Efficient Learning-based Scheduling for Information Freshness in Wireless Networks. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*. arXiv:2101.00257 [cs.PF]

[15] Ashutosh Nayyar, Tamer Başar, Demosthenis Teneketzis, and Venugopal V. Veeravalli. 2013. Optimal Strategies for Communication and Remote Estimation With an Energy Harvesting Sensor. *IEEE Trans. Automat. Control* 58, 9 (2013), 2246–2260.

[16] Michael J. Neely. 2021. Fast Learning for Renewal Optimization in Online Task Scheduling. , 44 pages. http://jmlr.org/papers/v22/20-813.html

[17] Tasmeen Zaman Ornee and Yin Sun. 2021. Sampling and Remote Estimation for the Ornstein-Uhlenbeck Process Through Queues: Age of Information and Beyond. *IEEE/ACM Transactions on Networking* 29, 5 (2021), 1962–1975.

[18] H Vincent Poor. 2013. *An introduction to signal detection and estimation*. Springer Science & Business Media.

[19] Herbert Robbins and Sutton Monro. 1951. A Stochastic Approximation Method. *The Annals of Mathematical Statistics* 22, 3 (1951), 400 – 407.

[20] Touraj Soleymani, Sandra Hirche, and John S. Baras. 2016. Optimal Information Control in Cyber-Physical Systems. *IFAC-PapersOnLine* 49, 22 (2016), 1–6. 6th IFAC Workshop on Distributed Estimation and Control in Networked Systems NECSYS 2016.

[21] Yin Sun, Yury Polyanskiy, and Elif Uysal. 2020. Sampling of the Wiener Process for Remote Estimation Over a Channel With Random Delay. *IEEE Transactions on Information Theory* 66, 2 (2020), 1118–1135.

[22] Yin Sun, Elif Uysal-Biyikoglu, Roy D. Yates, C. Emre Koksal, and Ness B. Shroff. 2017. Update or Wait: How to Keep Your Data Fresh. *IEEE Transactions on Information Theory* 63, 11 (2017), 7492–7508.

[23] Haoyue Tang, Yuchao Chen, Jintao Wang, Jingzhou Sun, and Jian Song. 2022. Sending Timely Status Updates through Channel with Random Delay via Online Learning. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications (INFOCOM 2022)*. London, United Kingdom (Great Britain).

[24] Haoyue Tang, Yuchao Chen, Jintao Wang, Pengkun Yang, and Leandros Tassiulas. 2022. Age Optimal Sampling Under Unknown Delay Statistics. *arXiv preprint arXiv:2202.13367* (2022).

[25] Haoyue Tang, Yin Sun, and Leandros Tassiulas. 2022. Sampling of the Wiener Process for Remote Estimation over a Channel with Unknown Delay Statistics. https://doi.org/10.48550/ARXIV.2207.08020

[26] Vishrant Tripathi and Eytan Modiano. 2021. An Online Learning Approach to Optimizing Time-Varying Costs of AoI. In *Proceedings of the Twenty-Second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing* (Shanghai, China) *(MobiHoc '21)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/3466772.3467053

[27] Cho-Hsin Tsai and Chih-Chun Wang. 2020. Age-of-Information Revisited: Two-way Delay and Distribution-oblivious Online Algorithm. In *2020 IEEE International Symposium on Information Theory (ISIT)*. 1782–1787.

[28] Cho-Hsin Tsai and Chih-Chun Wang. 2022. Unifying AoI Minimization and Remote Estimation—Optimal Sensor/Controller Coordination With Random Two-Way Delay. *IEEE/ACM Transactions on Networking* 30, 1 (2022), 229–242.

[29] Alexandre B. Tsybakov. 2008. *Introduction to Nonparametric Estimation* (1st ed.). Springer Publishing Company, Incorporated.

[30] Roy D. Yates. 2015. Lazy is timely: Status updates by an energy harvesting source. In *2015 IEEE International Symposium on Information Theory (ISIT)*. 3008–3012.

[31] Bin Yu. 1997. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*. Springer, 423–435.