Comparison of Decentralized and Centralized Update Paradigms for Distributed Remote Estimation

Sunjung Kang, Atilla Eryilmaz and Changhee Joo

***Abstract*—In this work, we perform a comparative study of *centralized* and *decentralized* update strategies for the basic remote tracking problem of many distributed users/devices with randomly evolving states. Our goal is to reveal the impact of the fundamentally different tradeoffs that exist between information accuracy and communication cost under these two update paradigms. In one extreme, decentralized updates are triggered by distributed users/transmitters based on *exact local state-information*, but also at a *higher cost* due to the need for uncoordinated multi-user communication. In the other extreme, centralized updates are triggered by the common tracker/receiver based on *estimated global state-information*, but also at a *lower cost* due to the capability of coordinated multi-user communication. We use a generic superlinear function to model the communication cost with respect to the number of simultaneous updates for multiple sources. We characterize the conditions under which transmitter-driven decentralized update policies outperform their receiver-driven centralized counterparts for symmetric sources, and vice versa. Further, we extend the results to a scenario where system parameters are unknown and develop learning-based update policies that asymptotically achieve the minimum cost levels attained by the optimal policies.**

I. Introduction

In recent years, there has been a growing number of applications requiring real-time updates of system status, especially in cyber-physical systems such as smart homes and buildings or health-care monitoring systems [2]. These systems rely on sensors that gather time-varying information and transmit it to a central controller or monitor, which then makes decisions based on the aggregated data from multiple sources. Although it is ideal to maintain the controller up-to-date all the way, this is often impractical due to limited resources of communication networks.

To ensure timely updates, there have been various studies conducted in the fields of Age of Information (AoI) [3]–[7] and Remote Estimation (RE) [8]–[25], where the value of information is measured with freshness and accuracy, respectively. The age is a quantitative measure used as a performance metric to assess the freshness of information. It is defined as the time elapsed since the most recent packet available at the destination was generated.

S. Kang and A. Eryilmaz are with ECE, The Ohio State University, USA {kang.853, eryilmaz.2}@osu.edu

C. Joo (corresponding author) is with CSE, Korea University, Korea changhee@korea.ac.kr

The work of S. Kang and A. Eryilmaz is funded in part by NSF grants: NSF AI Institute (AI-EDGE) 2112471, CNS-NeTS-2106679, CNS-NeTS-2007231, CNS-SpecEES-1824337; and the ONR Grant N00014-19-1-2621. The work of C. Joo is funded, in part by the MSIT, Korea, under the ICT Creative Consilience program (IITP-2022-2020-0-01819) supervised by the IITP, and in part by the NRF grant (MSIT, Korea) (No. NRF-2021R1A2C2013065, 2022R1A5A1027646).

The early version of this paper has been presented in IEEE INFOCOM 2021 [1].

In the context of the single-source single-destination scenario, prior work [4] has explored the determination of the optimal update rate for minimizing Age of Information (AoI) when random transmission times are considered. Another perspective from [7] delves into the examination of scheduling policies that factor in transmission costs. This introduces a compelling trade-off between managing the age of information and mitigating communication expenses. In the multiple-source single-destination scenario, a scheduling problem under communication constraints has been studied in [5], [6]. In [5], at most one source can transmit a packet via a channel, where a packet is dropped with some probability. In [6], a channel is modeled as a FIFO queue with random service time.

Similarly, in the context of RE, the estimation error is employed to measure the accuracy of information held by a remote monitor, in comparison to the actual information. It has been observed that a sampling strategy that minimizes the AoI does not necessarily minimize the estimation error [8], [9]. This observation is particularly evident in scenarios involving Wiener processes and Ornstein-Uhlenbeck processes, where the channel is modeled as a First-In-First-Out (FIFO) queue, as studied in [8], [9]. In [10], remote estimation problems with a packet-drop channel for both finite state Markov source and first-order autoregressive source are investigated, where a channel state changes over time horizon following finite-state Markov chain, and the packet-drop probability depends on a channel state and the transmission power level.

In [11], the Automatic Repeat reQuest (ARQ)-based remote estimation framework are studied for the linear time-invariant (LTI) system, where a sensor's observation is noisy and a channel's gain changes over time following finite-state Markov chain. In this domain, several works have tackled the scheduling problem under communication constraints [12]–[16]. In [12], [13], the minimization problem of Mean Squared Error (MSE) of an estimator (or monitor) is considered when the number of transmissions over finite-time horizon is constrained. The scheduling problem with per-transmission communication cost has been studied in [14], [15], [26]. In these contexts, each transmission carries an associated communication cost, and the overarching goal revolves around devising schedules that simultaneously minimize the MSE of an estimator and the communication expenses of a transmitter. This optimization objective extends across finite-time horizons [13], [14], [26] as well as infinite-time horizons [15]. The findings have shown that threshold-based update policies prove to be optimal across distinct types of information sources [15], [26]. Further insights emerge in [16], where the authors delve into investigating (mean-square) stability conditions within scenarios where a transmitter and a receiver communicate over multi-state Markov fading channels.

Exploring scenarios involving multiple sources communicating with a single destination has also been a significant area of research. These investigations have been undertaken in works such as [17]–[25], [27]. In [17]–[21], the individual sources are modeled as linear time-invariant (LTI) systems. Within each time slot, at most m out of n transmitters are allowed to update the remote monitor. Scheduling decisions are made either by a centralized controller or the receiver. In particular, periodic scheduling schemes are proposed in [18], [19]. The works in [27], [28] explore scenarios involving a centralized scheduler making scheduling decisions. These decisions are made considering either random changes in channel conditions [27] or the presence of packet loss probability in the channel [28].

In [21], distributed sensors or transmitters, each sensing an LTI system, contribute to scheduling decisions. Notably, only one transmitter can update the monitor in this setup. It is important to distinguish this from [17]–[20], where the primary goal revolves around minimizing the estimation error covariance. In contrast, in [21], [27], the main focus is on minimizing transmission power consumption while ensuring system stability. The works presented in [23]–[25] delve into scenarios that encompass multiple sources and a receiver, engaged in communication over various channels. Specifically, these works investigate stability conditions for remote estimation systems under both Markov fading channels [23] and semi-Markov fading channels [24]. Furthermore, in [25], researchers focus on establishing a sufficient stability condition for multi-source remote estimation and control problems.

Distinguishing our approach from the previously mentioned studies, we investigate the problem of remote estimation with multiple sources, where communication cost is associated with the number of simultaneous updates. Our primary goal is to study the fundamental dynamics underlying the trade-off between policies driven by transmitters and those driven by the receiver. By focusing on this aspect, we aim to understand essential insights that clarify the trade-off between estimation error and the cost associated with coordination between the transmitters. This research provides a fresh perspective on optimizing remote estimation in the presence of multiple sources, contributing to a better understanding of the interplay between transmitter-initiated and receiver-driven approaches.

We consider simple random-walk sources that transmit information through shared wireless channels, and assume that the channels are perfect, i.e, noiseless and no packet drop, as in [15], [22]. It is worth noting that the aim of our paper is not to provide specific efficient policies for any given system that captures certain complexities such as heterogeneous source dynamics or packet drop channels, etc., but rather to understand the fundamental insights into the trade-off between transmitter-driven and receiver-driven policies. Due to the channel sharing, communication cost changes according to the number of simultaneous transmissions of the sources, which will be explained in detail later.

Our contributions can be summarized as follows.

- We formulate the remote estimation problem in shared communication channels, where the estimator remotely tracks the time-varying state of multiple sources. We demonstrate, with an example, that the (communication) cost associated with coordination between distributed transmitters increases super-linearly with respect to the number of transmitters.
- We study information update policies that make decisions of *when* and *which* source information should be transmitted to the estimator when dynamics of sources are symmetric. The update decisions can be triggered either by the distributed transmitters based on exact local state-information or by the receiver based on estimated global state-information assuming that system parameters are known a priori.
- We then study information update policies when dynamics of sources are asymmetric.
- We extend the results to a scenario where system parameters are unknown, and develop learning-based update policies employing the Upper Confidence Bound (UCB) technique from the (stochastic) Multi-Armed Bandit (MAB) literature [29]. Through numerical simulations, we show that our learning-based update policies asymptotically achieve the minimum cost levels.

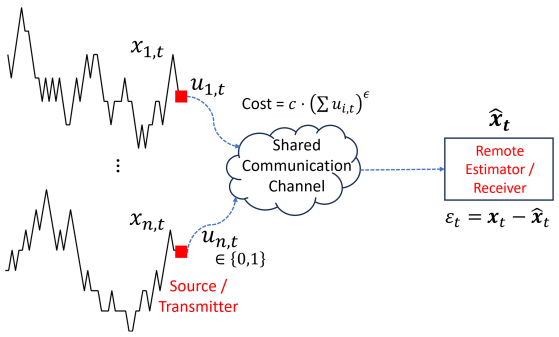


Fig. 1: System model.

The rest of paper is organized as follows. In Section II, we describe the system model and formulate the problem. In Sections III and IV, we study information update policies triggered by the distributed transmitters and by the receiver, respectively, when system parameters are known. In Section V, we compare the performance of the proposed update policies and extend them to the scenarios where the system parameters are unknown. In Section VII, we verify our analysis results through numerical simulations. In Section VIII, we conclude our work.

II. System Model

We present our system model with n information sources (e.g., sensors) and one remote estimator (e.g., sink or collector), where the estimator remotely tracks the time-varying state of the sources through shared wireless channels, as shown in Fig. 1. We describe the cost models of information mismatch and update communication, and then formulate our problem. We use the terms of sensor and transmitter, interchangeably, and similarly for the terms of estimator and receiver. The notations used in this paper are listed in Table I for ease of reference.

A. Value of Information

We consider a time-slotted system. At each time t, the state of each source changes following a random walk process. Specifically, let $w_{i,t}$ be an independent and identically distributed (i.i.d) random process with distribution as

$$w_{i,t} = \begin{cases} 1, & \text{with probability } p_i, \\ 0, & \text{with probability } 1-2p_i, \\ -1, & \text{with probability } p_i, \end{cases} \tag{1}$$

for some $p_i \in [0, 0.5]$. The parameter p_i is known to the receiver[1]. This simple non-biased, scalar-valued model not only captures the essential aspect of the problem, but also can be converted to a biased case by adding a constant drift. Let $x_{i,t}$ denote the state of source i at the beginning of time t, which is a random walk process associated with $w_{i,t}$ as

$$x_{i,t+1} = x_{i,t} + w_{i,t}, \text{ for } t \geq 0, \tag{2}$$

with initial state $x_{i,0} \in \mathbb{R}$.[2]

Let $u_{i,t} \in \{0, 1\}$ denote an update decision of transmitter i in time slot t, where $u_{i,t} = 1$ implies that transmitter i updates the receiver at time slot t. At the end of time slot $t-1$, the update decision $u_{i,t}$ can be made either in a decentralized manner by each transmitter or in a centralized manner by the receiver, based on their own observations up to time $t-1$. The detailed explanation will be made in Section II-D. Then the estimated state of source i at the receiver at time t, denoted by $\hat{x}_i(t)$, evolves as

$$\hat{x}_{i,t} = \begin{cases} x_{i,t}, & \text{when } u_{i,t} = 1, \\ \hat{x}_{i,t-1}, & \text{when } u_{i,t} = 0. \end{cases} \tag{3}$$

Let $\varepsilon_{i,t}$ denote the estimation error between $x_{i,t}$ and $\hat{x}_{i,t}$, i.e., $\varepsilon_{i,t} = x_{i,t} - \hat{x}_{i,t}$. Let $f_p(\varepsilon)$ be a penalty function, which increases with respect to the error ε. In this paper, we consider the squared error:

$$f_p(\varepsilon_{i,t}) = (x_{i,t} - \hat{x}_{i,t})^2. \tag{4}$$

B. Cost of Communication

When transmitter i makes a transmission at time slot t, i.e., $u_{i,t} = 1$, a packet containing the state value $x_{i,t}$ is successfully transmitted, incurring a communication cost. The communication cost may represent energy consumption or protocol overhead, which typically relies on diverse factors such as transmission power and interference intensity. In this paper, we pay attention to the cost associated with coordination between the transmitters, since multiple distributed transmitters should communicate over shared channels. For the sake of tractability, we assume that the communication cost depends on the number of simultaneous transmissions and is not affected by the specific transmitters engaged in transmission.

Let N_t denote the number of transmitters that take the update action simultaneously during time slot t, i.e., $N_t = \sum_{i=1}^{n} u_i(t)$. Then, we define the communication cost of transmitter i during time slot t under a given update policy π as $f^{\pi}_{c,i}(N_t)$. The specific formulation of $f^{\pi}_{c,i}(N_t)$ will be provided as we introduce each individual update paradigm in Section III and IV.

[1]Later in section V-B, we will address the case when the parameter is unknown and has to be learned.

[2]With scalar states and no sensing (or measurement) noise, the state evolution in (2) is a special case of a discrete-time linear time-invariant (LTI) system considered in [12]–[16]. If sources' states are multi-dimensional and each dimension is independent of other dimensions, the similar results obtained in this paper can be applied to the multi-dimensional case. The comparison of centralized and decentralized update paradigms under more general LTI systems is an interesting open problem.

TABLE I: Notations.

Symbol	Description
n	Number of sources / transmitters
$w_{i,t}$	Noise of source i in time slot t
p_i	State dynamic parameter of source i
$x_{i,t}$	State of source i in time slot t
$u_{i,t}$	Update decision of transmitter (Tx) i in time slot t
$\hat{x}_{i,t}$	Estimated state of source i at the receiver in time slot t
$\varepsilon_{i,t}$	Estimate error of source i in time slot t
$f_p(\varepsilon)$	Penalty function for estimate error ε
N_t	Number of transmissions in time slot t
$f^{\pi}_{c,i}(N_t)$	Comm. cost of Tx i in time slot t under policy π
$C^{\pi}_{i,t}$	Per-source cost of Tx i in time slot t under policy π
$g_{\pi}(\cdot)$	Expected average cost under policy π
c_s	Comm. cost constant under decent. update paradigms
ϵ_s	Comm. cost coef. under decent. update paradigms
$\tilde{g}_{\text{TD}}(\gamma)$	Expected avg. cost of TD policy with threshold γ
γ^*_L	Threshold used under the TD-L policy
c_r	Comm. cost constant under cent. update paradigms
ϵ_r	Comm. cost coef. under cent. update paradigms
$\tilde{g}_{\text{RD}}(\tau)$	Expected avg. cost of RD policy with period τ
$\bar{\gamma}$	Upper bound of possible thresholds
$\bar{\tau}$	Upper bound of possible periods
$\hat{r}_j$	Average cost during update interval j
$\hat{r}(\gamma)$ / $\hat{r}(\tau)$	Empirical avg. cost for threshold γ (or period τ)
$\eta(\gamma)$ / $\eta(\tau)$	Number of selection for threshold γ (or period τ)
α	Drift of source dynamics
γ_{max}	Max. threshold for source with asymmetric dynamics
γ_{min}	Min. threshold for source with asymmetric dynamics

C. Problem formulation

Considering the aforementioned costs, the *per-source* cost associated with source i at each time t, under policy π, can be written as

$$C^{\pi}_{i,t} = u_{i,t} f^{\pi}_{c,i}(N_t) + (1 - u_{i,t}) f_p(\varepsilon_{i,t}). \tag{5}$$

Suppose that $\mathbf{x}_0 = \hat{\mathbf{x}}_0$. Our objective is to find an update policy π that minimizes the expected average cost over an infinite time horizon:

$$\text{minimize } g_{\pi}(n), \tag{6}$$

where

$$g_{\pi}(n) = \mathbb{E}\left[\lim_{s\to\infty} \frac{1}{sn} \sum_{t=1}^{s} \sum_{i=1}^{n} C^{\pi}_{i,t}\right]. \tag{7}$$

In this work, we focus on the case of homogeneous transmitters with $p_i = p$ for all i.

D. Decentralized and Centralized Update Paradigms

We organize our investigation under two fundamentally different paradigms: that of decentralized and centralized

TABLE II: Information and control available to the policies.

Policy	TD-L	TD-G	RD
local parms.	p, c_s	p, c_s	p, c_r
global parms.	–	n, ϵ_s	n, ϵ_r
error	$\varepsilon_{i,t}$	$\varepsilon_{i,t}$	–
controller	transmitter i	transmitter i	receiver
control var.	$u_{i,t}$	$u_{i,t}$	$u_{1,t}, \ldots, u_{n,t}$

update policies. These paradigms can also be referred to as transmitter-driven (TD) and receiver-driven (RD) paradigms, respectively, since the update decisions are triggered by each transmitter under the former one, while the update decisions are triggered by the receiver under the latter one.

Under a TD policy, each transmitter independently makes individual decisions based on its own error $\varepsilon_{i,t}$, but without the knowledge of the other's actions, e.g., the number N_t of transmitters in time slot t. On the other hand, under a RD policy, the receiver can collectively decide on the update actions (thus the set of transmitters at time slot t is under control), but it lacks knowledge of the current errors $\varepsilon_{i,t}$. Intuitively, when the communication cost $C^{\pi}_{i,t}$ is relatively small (meaning relatively small N_t), the error cost $f_p(\varepsilon_{i,t})$ dominates the communication cost $f^{\pi}_{c,i}(N_t)$. Consequently, a TD policy may outperform a RD policy. However, when the communication cost becomes sufficiently large, the communication cost starts dominating the error cost, and thus a RD policy will outperform a TD policy. The information and control available for each policy are summarized in Table II.

The objective of this work is to explore and compare TD and RD policies across different scenarios in relation to the number of transmitters and the communication cost structure. Through this study, we aim to understand the fundamental insights into the trade-off between TD and RD policies.

III. Decentralized Update Paradigm

In this section, we begin by establishing the communication cost structure in the context of decentralized update paradigms. Subsequently, we design two different types of TD policies based on their level of coordination: one type is for transmitters with only local information (called as TD-L policy), and the other type is for transmitters with global information as well as the local information (called as TD-G policy).

A. Cost of Communication

In order to motivate and understand the property of the communication cost function under decentralized update paradigms, let us consider a scenario where N transmitters access a shared channel using a Slotted ALOHA scheme. In this setup, each time slot t is divided into mini-slots, in which N transmitters independently attempt to transmit a packet with an identical probability q. During a mini-slot, if a transmission is successful (meaning only one transmitter attempts withing the mini-slot), the corresponding transmitter receives an ACK by the end of the mini-slot and stops transmitting in the subsequent mini-slots (by the end of the time slot).

As the number of transmitters increases, the level of contention increases, leading to a higher probability of collisions.

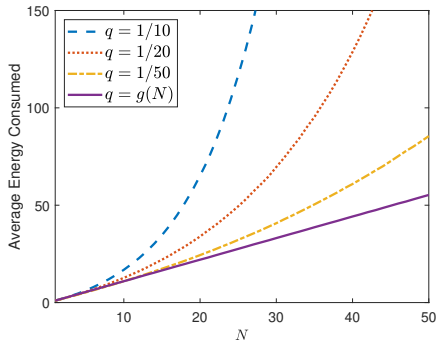


Fig. 2: Average energy consumed for successful update with respect to the number N of simultaneous transmissions, when each transmitter attempts with probability q.

When a collision occurs, no transmission is successfully completed in that mini-slot. Consequently, with an increase in the number of transmitters, the average number of attempts a transmitter makes in a time slot for a successful transmission also increases. Suppose that each transmission consumes a unit of energy (or power), and that each time slot consists of sufficiently many mini-slots to enable all transmitters to succeed in transmitting within the time slot. In this scenario, the average energy cost for a successful transmission, that can be considered as the communication cost per an update of a source, will increase with the number N of simultaneous transmissions.

As an example, we run simulations involving 50 transmitters utilizing the Slotted ALOHA protocol and measure the average update cost of N sources, assuming each transmission consumes a unit of energy. The results, obtained for different combinations of N and q, are shown in Fig. 2, where $q = g(N)$ denotes the utilization of an empirically determined optimal q based on the given N. The results reveal that employing a fixed value of q leads to an exponential increase of the average cost as N increases. The minimum cost is achievable by appropriately adjusting the value of q in accordance with N.

Based on the observation, we model the communication cost under decentralized update paradigms as a function of N_t, the number of simultaneous transmissions at time slot t, in the following form:

$$f^{\pi_{\text{dec}}}_{c,i}(N_t) = c_s N_t^{\epsilon_s}, \tag{8}$$

where π_{dec} denotes an update policy within decentralized update paradigms, and constant $c_s > 0$ and exponent coefficient $\epsilon_s \geq 1$ are involved.[3] We remark that the Slotted ALOHA is used as an example to motivate the super-linearity of a communication cost with respect to the number of sources within the context of decentralized update paradigms. In this paper, we consider a network where simultaneous transmissions at a given time are allowed with high communication cost. We also remark that update policies for distributed remote estimation

[3]In this paper, we consider a single channel. Extension to a practical multi-channel environment where each channel has a different coefficient ϵ_s remains as an interesting open problem.

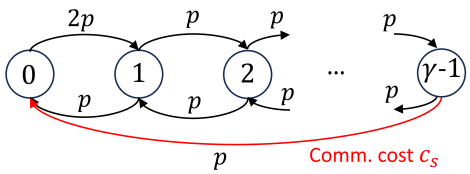


Fig. 3: Evolution of error $|\varepsilon_t|$ using the threshold-based policy as defined in (9).

over a random access channel have been studied in [30]. We denote the expected cost function as $g_\pi(n, c_s, \epsilon_s)$ for a TD policy π to highlight its dependence on n, c_s and ϵ_s.

B. Decentralized Update Policy

In a single source system with constant per-transmission communication cost, prior research [15] has shown that an optimal update policy is *of threshold type* in the forms of

$$u_t^* = \begin{cases} 0, & \text{if } |\varepsilon_t + w_t| < \gamma, \\ 1, & \text{if } |\varepsilon_t + w_t| \geq \gamma, \end{cases} \tag{9}$$

where $\gamma > 0$ denotes a threshold value. Subsequently, the error evolution ε_t can be depicted as a Markov chain, as illustrated in Fig. 3. For notational convenience, we omit subscript i and denote this policy as *th-*γ. Further, given constant per-transmission communication cost $\bar{c}$, it was also shown in [15] that the expected average cost $h(\bar{c})$ over infinite time horizon can be obtained as

$$h(\bar{c}) := g_{th\text{-}\gamma}(1, \bar{c}, \epsilon_s) = \frac{2}{\gamma^2}\left(\bar{c}p + \sum_{i=1}^{\gamma-1} f_p(i)(\gamma - i)\right). \tag{10}$$

Additionally, when considering the MSE $f_p(\varepsilon) = \varepsilon^2$, the optimal threshold that minimizes (10) is $\gamma^* = \lfloor \sqrt[4]{12p\bar{c}} \rfloor$ or $\lceil \sqrt[4]{12p\bar{c}} \rceil$.

Inspired by these results, we consider a TD policy where each transmitter updates the receiver using a threshold γ_i. Consequently, the number N_t of simultaneous transmissions at time slot t becomes a random variable, resulting in the per-transmission communication cost also becoming a random variable as per (8). To characterize the performance of the TD policy, we study the asymptotic behavior of N_t, which will lead to our understanding of the expected average cost $g_{TD}(n, c, \epsilon_s)$.

Note that due to each transmitter independently updating the receiver, the error $\varepsilon_{i,t}$ for transmitter i can be viewed as an independent renewal process, as it resets to 0 with every update. An inter-renewal distribution is called *arithmetic* if the intervals between renewals are integer multiples of some real number. The span of an arithmetic distribution is defined as the largest number ρ for which this property holds. Subsequently, the following theorem provides an asymptotic behavior of renewal probabilities.

Theorem 3.1: [Theorem 4.6.2 in [31]] If an inter-renewal distribution is arithmetic with span ρ, then

$$\lim_{t\to\infty} \mathbb{P}(\textit{Renewal at } t\rho) = \frac{\rho}{\mathbb{E}[T]}, \tag{11}$$

where T denotes the inter-renewal interval.

The renewal process $\varepsilon_{i,t}$ exhibits an arithmetic nature with a span of $\rho = 1$, as the inter-renewal intervals can be $\gamma, \gamma+1, \gamma+2, \ldots$. Moreover, the expectation of the inter-renewal interval under the threshold-type update policy with a threshold γ is known as $\mathbb{E}[T] = \frac{\gamma^2}{2p}$ [15]. Hence, we can obtain that $\lim_{t\to\infty} \mathbb{P}(\text{Renewal at } t) = \lim_{t\to\infty} \mathbb{P}(u_{i,t} = 1) = \frac{1}{\mathbb{E}[T]} = \frac{2p}{\gamma^2}$.

Combined with the independence of the renewal processes, Theorem 3.1 can be used to characterize the asymptotic behavior of N_t.

Lemma 3.1: When n independent (symmetric) transmitters update the receiver with the same threshold γ, the number N_t of simultaneous transmissions at time slot t converges in distribution to a Binomial distribution with parameters n and $\frac{2p}{\gamma^2}$, i.e.,

$$\lim_{t\to\infty} N_t \sim \textit{Binom}\left(n, \tfrac{2p}{\gamma^2}\right), \tag{12}$$

where $Binom(\cdot,\cdot)$ denotes the Binomial distribution.

Lemma 3.1 can be shown using the independence of the transmitters' decision $u_{i,t}$ and Theorem 3.1. We refer to Appendix A for the proof.

Let $\tilde{g}_{\text{TD}}(\gamma, i)$ denote the expected average cost of a TD policy for source i using threshold γ, and let N follow the distribution of $\lim_{t\to\infty} N_t$. Replacing $\bar{c}$ in (10) with the per-transmission communication cost $c_s k^{\epsilon_s}$, and from Lemma 3.1, we can obtain:

$$\begin{aligned} \tilde{g}_{\text{TD}}(\gamma, i) &= \textstyle\sum_{k=1}^{n} \mathbb{P}\left(N = k \mid u_{i,t} = 1\right) h(c_s k^{\epsilon_s}) \\ &= \textstyle\sum_{k=0}^{n-1} \mathbb{P}\left(k; n-1, \frac{2p}{\gamma^2}\right) h(c_s (k+1)^{\epsilon_s}), \\ &= \mathbb{E}[h(c_s (K+1)^{\epsilon_s})], \end{aligned} \tag{13}$$

where $\mathbb{P}(k; n, q)$ is the probability that $N = k$ when $N \sim Binom(n, q)$, and K is a random variable that follows $Binom\left(n-1, \frac{2p}{\gamma^2}\right)$. Due to the symmetry, this holds for all i, and we can write $\tilde{g}_{\text{TD}}(\gamma, i) = \tilde{g}_{\text{TD}}(\gamma)$ for all i.

For the TD policies with local information (i.e., p and c_s) (TD-L), each transmitter i optimizes its threshold γ agnostic about other transmitters, which results in $\gamma_L^* = \lfloor \sqrt[4]{12pc_s} \rfloor$ or $\lceil \sqrt[4]{12pc_s} \rceil$ that leads to the expected average cost

$$g_{\text{TD-L}}(n, c_s, \epsilon_s) = \tilde{g}_{\text{TD}}(\gamma_L^*). \tag{14}$$

On the other hand, for the TD policies with global information (i.e., p, c_s, n and ϵ_s) (TD-G), the transmitters can minimize (13) by further optimizing their threshold γ with respect to n and ϵ_s, which results in the expected average cost

$$g_{\text{TD-G}}(n, c_s, \epsilon_s) = \min_{\gamma \geq 0} \tilde{g}_{\text{TD}}(\gamma). \tag{15}$$

We note that obtaining a closed-form expression for $\tilde{g}_{\text{TD}}(\gamma)$ in (13) is challenging due to the necessity of evaluating the expectation of a non-linear function involving the random variable K, i.e., $\mathbb{E}[(K+1)^{\epsilon_s}]$. This complexity makes the analytical determination of an optimal γ in (15) intractable. Consequently, considering the closed-form threshold γ_L^* obtained from local information might be a viable approach to reduce computational complexity. Alternatively, one could explore the use of a learning policy, as elaborated in Section V-B.

IV. Centralized Update Paradigm

Differing from decentralized update paradigms, in the context of centralized update paradigms, the receiver undertakes the task of managing transmissions among transmitters. Consequently, the communication cost related to transmission coordination is relatively lower compared to that within decentralized update paradigms. Further, unlike the TD policy, wherein each transmitter can monitor errors, the RD policy lacks direct access to error information. As a result, the receiver's decision-making process relies on estimating the current error for each source. This reliance on error estimation, influenced by the error's renewal property, gives rise to periodic updates over time.

A. Cost of Communication

In order to motivate and understand the property of the communication cost structure under centralized update paradigms, we can consider a scenario where the centralized receiver is equipped with advanced multi-user detection techniques, such as successive interference cancellation or interference alignment [32]. These techniques can effectively mitigate the interference caused by simultaneous transmissions, allowing the receiver to reliably decode and recover the information from multiple transmitters in the presence of interference. In this well-designed communication framework, the additional cost incurred by each transmitter during simultaneous updates may not scale linearly with the number of transmitters. Instead, due to the receiver's ability to efficiently separate and decode received signals, the cost escalation could occur at a slower rate. This suggests the potential for achieving sub-linear overhead concerning the number of simultaneous transmissions under certain conditions.

Based on this motivation, we model the communication cost under centralized update paradigms as a function of N_t in the following form:

$$f_{c,i}^{\pi_{\text{cent}}}(N_t) = c_r N_t^{\epsilon_r}, \tag{16}$$

where π_{cent} denotes an update policy within centralized update paradigms, and $c_r > 0$ and $\epsilon_r > 0$ are involved. Note that $f_{c,i}^{\pi_{\text{cent}}}(N_t)$ can be sub-linear, linear or super-linear. This is in contrast to the communication cost within decentralized update paradigms, which we discussed as being super-linear in Section III-A.

B. Expected Error

Before delving into an RD policy, we begin our exploration by studying how the expected error between a source and the estimator evolves over time. For notational convenience, we omit the subscript i in our notations.

Let s denote the time elapsed since a transmitter's update to the receiver. In this context, there exist $2s+1$ potential error states for the source (i.e., $x - \hat{x} \in [-s, s]$). Denote $\mathbf{e}_s = [e_s(-s), ..., e_s(0), ..., e_s(s)]$ as the expected error vector when the receiver is not updated by the transmitter for s consecutive time slots, where $e_s(k)$ corresponds to the probability that the error between the receiver and the source is k (i.e., $x - \hat{x} = k$). Employing (1), the evolution of expected error follows Bayes' rule, given by:

$$e_s(k) = e_{s-1}(k)(1-2p) + (e_{s-1}(k-1) + e_{s-1}(k+1))p, \tag{17}$$

where $k \in -s, ..., s$, while $e_{s-1}(-s-1) = e_{s-1}(-s) = e_{s-1}(s) = e_{s-1}(s+1) = 0$.

Let $\xi(s)$ denote the expected error cost when the receiver has not been updated from the source for s consecutive time slots, i.e.,

$$\xi(s) = \textstyle\sum_{k=-s}^{s} e_s(k) f_p(|k|). \tag{18}$$

In the special case of the mean squared error penalty function $f_p(\varepsilon) = \varepsilon^2$, the expected error cost $\xi(s)$ can be obtained as the following lemma.

Lemma 4.1: If $f_p(\varepsilon) = \varepsilon^2$, the expected error cost after s consecutive time slots since the last update is

$$\xi(s) = 2ps, \text{ for } s \geq 1. \tag{19}$$

We refer to Appendix B for the proof.

C. Single-transmitter Scenario

We begin our investigation by deriving an optimal solution for a single-transmitter problem with the MSE penalty function. In this straightforward scenario, we determine an optimal RD policy. By subsequently comparing it to a TD policy, we can gain insight into the advantages of employing a TD policy over an RD policy when the simultaneous transmission results in relatively small communication costs.

We initiate our analysis by formulating a discrete-time Markov Decision Process (MDP). In this setup, the state at time slot t is denoted by s. Within each state, the receiver has two possible actions: either to update ($u = 1$) or to refrain from updating ($u = 0$). If $u = 0$, the state progresses to $s+1$. On the other hand, if $u = 1$, the state transitions to 1. Assuming a per-transmission communication cost of $\bar{c}$, the expected cost associated with state s and action u can be expressed as $u\bar{c} + (1-u)\xi(s)$. Consequently, we have the following Bellman equation:

$$\phi(s) = \min\{\xi(s) + \phi(s+1) - \lambda, \bar{c} + \phi(1) - \lambda\}, \tag{20}$$

where $\xi(s) = 2ps$, $\phi(\cdot)$ denotes the cost-to-go function, and λ denotes the minimum expected average cost over infinite time horizon [33].

We show in Lemma 4.2 that an optimal policy that solves the Bellman equation (20) is also of threshold type.

Lemma 4.2: There exists a threshold policy that optimally solves the Bellman equation (20). Specifically, given constant update cost $\bar{c}$, the policy has a real-valued time threshold $\tau^*(\bar{c})$ such that

$$u_s^* = \begin{cases} 0, & \text{if } s < \tau^*(\bar{c}), \\ 1, & \text{if } s \geq \tau^*(\bar{c}). \end{cases} \tag{21}$$

Lemma 4.2 can be shown as in [15] using the fact that $\xi(s)$ is increasing in s when $f_p(\varepsilon) = \varepsilon^2$. We refer to Appendix C for the proof. Note that unlike the optimal TD policy (9), the optimal RD policy has a *time threshold* with periodic updates.

Given time threshold τ, the expected average cost $g_{\text{RD}}(\tau)$ under the RD update policy is given by

$$g_{\text{RD}}(\tau) = \frac{1}{\tau}\left(\bar{c} + \sum_{s=1}^{\tau-1}\xi(s)\right). \tag{22}$$

For $f_p(\varepsilon) = \varepsilon^2$, we have $\xi(s) = 2ps$ and thus $g_{\text{RD}}(\tau) = \bar{c}/\tau + p(\tau-1)$, which is convex in $\tau > 0$. Thus, by solving $\frac{dg_{\text{RD}}}{d\tau} = 0$, we can obtain a closed-form expression of an optimal time threshold that solves (20): $\tau^*(\bar{c}) = \lfloor\sqrt{\bar{c}/p}\rfloor$ or $\lceil\sqrt{\bar{c}/p}\rceil$.

Note that, under the TD policy with a single transmitter, we have the expected update interval $\mathbb{E}[T] = \sqrt{3\bar{c}/p}$. That is, the RD policy updates the receiver more frequently than the TD policy on average. This is because the controller does not use the error ε_t and thus it compensates for the lack of information by updating more frequently.

D. Multi-transmitter Scenario

Now, consider a scenario where n (homogeneous) transmitters update the receiver. Unlike the TD policy where each transmitter independently updates the receiver, resulting in a random number of simultaneous transmissions at any given time, an RD policy enables control over the number of simultaneous transmissions to ensure that the communication cost remains within reasonable bounds. Given that we are dealing with homogeneous transmitters, with $p_i = p$ for all i, the update periods (referred to as time thresholds) are consistent across all transmitters.

Note that for a fixed time period τ, the expected estimation error remains constant $p\tau(\tau-1)$. Thus, to minimize the expected average cost, it is necessary to minimize the communication cost. Consider a scenario where m transmitters are allocated among τ time slots, and the objective is to allocate transmitter i to one of these time slots. Let k denote the number of transmitters already assigned to a particular time slot, resulting in a cumulative communication cost of k^{ϵ_r+1} for that slot. If transmitter i be allocated within this slot, the total communication cost increases to $(k+1)^{\epsilon_r+1}$. Given that $\epsilon_r > 0$, the total communication cost for each time slot increases super-linearly with respect to the number of simultaneous transmissions. Thus, the optimal strategy[4] for placing transmitter i is to select the time slot with the smallest number of existing transmitters. Consequently, the optimal strategy for deploying all transmitters across the τ time slots is to uniformly distribute them among the time slots.

Let τ_{n,ϵ_r} denote the update period. The receiver can optimize τ_{n,ϵ_r} by taking into account n and ϵ_r, and control the transmissions by assigning a time slot to each transmitter.

- When $n \le \tau_{n,\epsilon_r}$, an optimal policy involves each transmitter i updating during time slot t such that $(t \mod \tau_{n,\epsilon_r}) = i$. This ensures there is at most one transmission during each time slot. In this case, there are $\tau_0 = \tau_{n,\epsilon_r} - n$ idle (no-update) time slots within the update period τ_{n,ϵ_r}.
- When $n > \tau_{n,\epsilon_r}$, an optimal policy lets each transmitter i update at time slot t such that $(t \mod \tau_{n,\epsilon_r}) = (i \mod \tau_{n,\epsilon_r})$. Then, at each time slot, there are $\lceil\frac{n}{\tau_{n,\epsilon_r}}\rceil$ transmissions or $\lceil\frac{n}{\tau_{n,\epsilon_r}}\rceil - 1$ transmissions. Let $k_{n,\epsilon_r} = \lceil\frac{n}{\tau_{n,\epsilon_r}}\rceil$, and let τ_0 denote the number of time slots where $\lceil\frac{n}{\tau_{n,\epsilon_r}}\rceil - 1$ transmitters update the receiver within an update period τ_{n,ϵ_r}. The structure of an optimal RD policy given n and τ_{n,ϵ_r} is shown by Fig.4. Each slot on the x-axis represents one time slot, and each bin on the y-axis represents one transmission opportunity. The number in each bin is the index of transmitters of $\{1,2,\ldots,n\}$. Note that each of the first $(\tau_{n,\epsilon_r} - \tau_0)$ time slots on a period has k_{n,ϵ_r} simultaneous transmissions, and yields the total cost of $c_r(k_{n,\epsilon_r})^{\epsilon_r}$. Each of the rest τ_0 time slots has $k_{n,\epsilon_r} - 1$ simultaneous transmissions, and the cost of $c_r(k_{n,\epsilon_r} - 1)^{\epsilon_r}$.

Let $\tilde{g}_{\text{RD}}(\tau_{n,\epsilon_r})$ denote the expected average cost given τ_{n,ϵ_r}, which is given by

$$\begin{aligned}\tilde{g}_{\text{RD}}(\tau) &= \frac{k(\tau-\tau_0)}{n\tau}(ck^{\epsilon_r} + p(\tau-1)\tau) \\ &\quad + \frac{(k-1)\tau_0}{n\tau+}(c(k-1)^{\epsilon_r} + p(\tau-1)\tau) \\ &= \frac{c}{n\tau}((\tau-\tau_0)k^{1+\epsilon_r} + \tau_0(k-1)^{1+\epsilon_r}) + p(\tau-1),\end{aligned} \tag{23}$$

where we omit subscripts n and ϵ_r for notational convenience (i.e., $k = k_{n,\epsilon_r}$. Note that $\tau = \tau_{n,\epsilon_r}$) and $k = \lceil n/\tau\rceil$ and $\tau_0 = k\tau - n$. The expected average cost, $g_{\text{RD}}(n, c_r, \epsilon_r)$, under the RD policy is given by

$$g_{\text{RD}}(n, c_r, \epsilon_r) = \min_{\tau\ge 1}\tilde{g}_{\text{RD}}(\tau). \tag{24}$$

Similar to the TD-G policy discussed in Section III-B, it is worth noting the complexity in optimizing $\tilde{g}_{\text{RD}}(\tau)$ concerning τ in (24), primarily due to its non-convex nature. Consequently, rather than directly tackling these intricate optimization challenges, we provide a comprehensive comparison of the asymptotic behaviors exhibited by our proposed policies in Section V-A.

V. Performance Comparison of Decentralized and Centralized Update Paradigms

In this section, we conduct a comparative analysis of the two TD policies and the RD policy. We highlight that it is intractable to solve the optimization problems 15 and 24 due to the complexity of the objective functions. These functions are not only intricate but can also lack convexity, primarily due to their dependence on exponent coefficients ϵ_s and ϵ_r. Instead of solving these intricate optimization problems directly, we offer a comparison of the asymptotic behaviors exhibited by our proposed policies. This approach allows us to understand the fundamental insights into the trade-off between transmitter-driven and receiver-driven policies. Furthermore, we will extend our design to a scenario where the system parameters are unknown, enhancing the practical relevance of our study. .

A. TD-L Policy vs. TD-G Policy vs. RD Policy

We first consider the single-transmitter case of $n = 1$, in which TD-L is equivalent to TD-G. Suppose that $c_s = c_r = c$.

[4]This policy is optimal in the sense that there is no other policy that can make communication cost smaller.

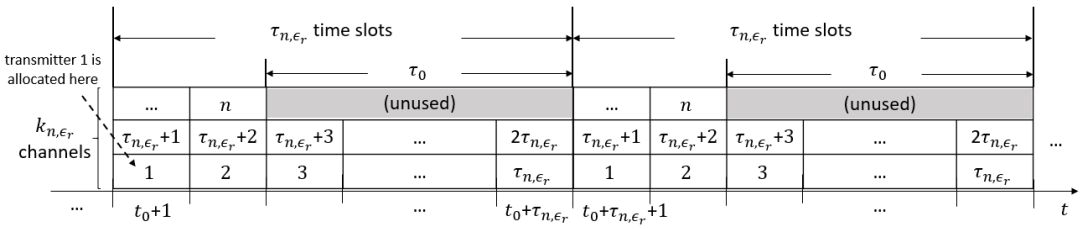


Fig. 4: Time slot and channel allocation of the RD policy.

If $c < 2p$, then the optimal policy is to update at every time slot and we have $g_{\text{TD}}(1, c, \epsilon_s) = g_{\text{RD}}(1, c, \epsilon_r) = c$. Suppose that $c \geq 2p$. Then, from (22) with $\tau = \sqrt{c/p}$ and (10) with $\gamma = \sqrt[4]{12pc}$, we have

$$g_{\text{RD}}(1, c, \epsilon_r) = 2\sqrt{pc} - p \geq \tfrac{2}{\sqrt{3}}\sqrt{pc} - \tfrac{1}{6} = g_{\text{TD}}(1, c, \epsilon_s), \quad (25)$$

where $g_{\text{TD}} = g_{\text{TD-L}} = g_{\text{TD-G}}$ and the inequality comes from that $c \geq 2p$.

Not only this confirms the expected superiority of TD updates to RD updates for the single-transmitter case, but also reveals that the performance improvement is a function of system parameters p and c. Note that when $\epsilon_s = \epsilon_r = 0$, each transmitter pays the same per-transmitter cost c regardless of the number of simultaneous transmissions. Hence, we have $g_{\text{RD}}(n, c, 0) = g_{\text{RD}}(1, c, \epsilon_r) \geq g_{\text{TD}}(1, c, \epsilon_s) = g_{\text{TD}}(n, c, 0)$, i.e., TD policies always outperforms RD policy.

We will begin by examining the scenario with a single transmitter, denoted as $n = 1$, in which TD-L is equivalent to TD-G. Let's assume that $c_s = c_r = c$. If the communication cost c is less than $2p$, then the optimal policy is to update in every time slot, resulting in $g_{\text{TD}}(1, c, \epsilon_s) = g_{\text{RD}}(1, c, \epsilon_r) = c$. On the other hand, if $c \geq 2p$, we can utilize (22) with $\tau = \sqrt{c/p}$ and (10) with $\gamma = \sqrt[4]{12pc}$ to deduce that:

$$g_{\text{RD}}(1, c, \epsilon_r) = 2\sqrt{pc} - p \geq \tfrac{2}{\sqrt{3}}\sqrt{pc} - \tfrac{1}{6} = g_{\text{TD}}(1, c, \epsilon_s), \quad (26)$$

where we have taken into account the inequality $c \geq 2p$.

This comparison not only confirms the expected advantage of TD updates over RD updates for the single-transmitter case but also reveals that the degree of performance enhancement depends on the system parameters p and c. Importantly, when both ϵ_s and ϵ_r are set to 0, indicating that each transmitter incurs the same per-transmitter cost c regardless of the number of simultaneous transmissions, we find that $g_{\text{RD}}(n, c, 0) = g_{\text{RD}}(1, c, \epsilon_r) \geq g_{\text{TD}}(1, c, \epsilon_s) = g_{\text{TD}}(n, c, 0)$. This implies that TD policies consistently outperform RD policies.

Now, we consider when $\epsilon_s \geq 1$, $\epsilon_r > 0$ and $n \gg 1$. Theorem 5.1 shows the asymptotic behavior of $g_{\text{TD-L}}$, $g_{\text{TD-G}}$ and g_{RD}, under the assumption that $c \geq 2p$.

Theorem 5.1: Under the TD-L and TD-G policies, we have the asymptotic lower bounds such that

$$g_{\text{TD-L}}(n, c_s, \epsilon_s) = \Omega(n^{\epsilon_s}), \quad (27)$$

and

$$g_{\text{TD-G}}(n, c_s, \epsilon_s) = \Omega\left(n^{\frac{\epsilon_s}{\epsilon_s+2}}\right), \quad (28)$$

respectively, for $\epsilon_s > 1$. Under the RD policy, we have an asymptotic upper bound such that

$$g_{\text{RD}}(n, c_r, \epsilon_r) = O\left(n^{\frac{\epsilon_r}{\epsilon_r+2}}\right) \quad (29)$$

for $\epsilon_r > 0$.

We refer to Appendix E for the detailed proof.

Given that $g_{\text{RD}}(1, c, \epsilon_r) \geq g_{\text{TD-L}}(1, c, \epsilon_s) = g_{\text{TD-G}}(1, c, \epsilon_s)$ holds for $c = c_s = c_r$ and any ϵ_s and ϵ_r, the implications of Theorem 5.1 become apparent. The theorem indicates the existence of a crossing point where the RD policy begins to surpass the TD-L policy for $\epsilon_s \geq \epsilon_r$, and the RD policy outperforms the TD-G policy for $\epsilon_s > \epsilon_r$. This trend becomes more evident when considering our discussions in Sections III-A and IV-A, where we established that communication costs under decentralized update paradigms tend to be super-linear ($\epsilon_s \geq 1$), while those under centralized update paradigms can exhibit sub-linear behavior ($\epsilon_r \in (0, 1)$).

In essence, changing the strategy depending on parameters n, ϵ_s and ϵ_r for some given p and c improves the system performance. More specifically, when the value of information holds greater significance compared to communication costs (i.e., for relatively small values of n and $\epsilon_s - \epsilon_r$), TD policies are the preferred choice. Conversely, when communication costs outweigh the value of information, an RD policy tends to be more effective.

Remark 5.1: our paper primarily focuses on understanding the trade-off between estimation error and communication cost through an analysis of the joint optimization problem. On the other hand, depending on practical scenarios, the error cost minimization under the communication error constraints, or the communication cost minimization under the error cost constraints can be more relevant. From our results, it can be infer that when the limitation on communication cost is relatively tight, leading to a small number of simultaneous transmissions, the RD policy is expected to perform better than the TD policy. Conversely, if the communication cost constraint becomes less strict, the TD policy might be more favorable in terms of performance compared to the RD policy. Further, when considering communication cost minimization under the error cost constraints, we can similarly expect potential outcomes and behaviors.

B. Learning-based update policy

In this subsection, we consider scenarios where source's dynamic parameter p is *unknown*. We assume that the upper bounds of thresholds $\overline{\gamma}$ and $\overline{\tau}$ are given for TD and RD policies, respectively. In the following, we develop learning-based TD and RD policies employing the Multi-Armed Bandit

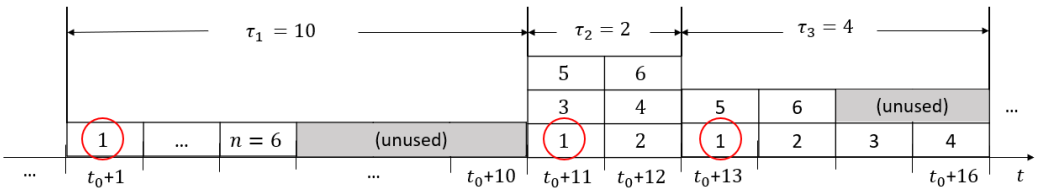


Fig. 5: Example of varying thresholds: Transmitter 1 marked with the circle has the same update intervals with the given thresholds.

(MAB) technique by considering each possible threshold as an arm[5].

Learning-based TD policy[6]*:* Let $s_{i,j}$ denote the time when transmitter i performs its j^{th} update, with $s_{i,0} = 0$ for all transmitters. Let $\Delta_{i,j} := s_{i,j+1} - s_{i,j}$ denote the duration between the $(j+1)^{th}$ and j^{th} updates for transmitter i. In the learning-based algorithm, an update interval corresponds to a round for learning, where we apply the Upper Confidence Bound (UCB) technique [29]. At the beginning of the j^{th} interval, transmitter i chooses threshold $\gamma_{i,j}$. When the error exceeds this threshold, as determined by (9), transmitter i sends its update packet to the receiver, incurring a communication cost of $c_s \left(\sum_{i=1}^{n} u_{i,s_{j+1}}\right)^{\epsilon_s}$. Consequently, at the end of the j^{th} interval, the average cost $\hat{r}_{i,j}$ of transmitter i during interval j can be written as

$$\hat{r}_{i,j} = \frac{1}{\Delta_{i,j}} \left(\sum_{t=s_{i,j}+1}^{s_{i,j+1}} \varepsilon_{i,t}^2 + c_s \left(\sum_{i=1}^{n} u_{i,s_{j+1}}\right)^{\epsilon_s} \right). \quad (30)$$

Transmitter i uses this average cost as feedback for the choice of threshold $\gamma_{i,j}$. For each threshold $\gamma \in [0, \overline{\gamma}]$, transmitter i stores the empirical average cost and the number of selections up to now as $\hat{r}_i(\gamma)$ and $\eta_i(\gamma)$, respectively.

We run the following procedure independently for each transmitter i. For the first $\overline{\gamma}+1$ update intervals, the transmitter selects $\gamma \in [0, \overline{\gamma}]$ exactly once. For each interval $j > \hat{\gamma} + 1$, it decides an action according to the following procedure.

- At the beginning of the j^{th} update interval:
  1) For each γ, $I_i(\gamma) \leftarrow \frac{\hat{r}_i(\gamma)}{\max_{\gamma'} \hat{r}_i(\gamma')} - \sqrt{\frac{2\log(j)}{\eta_i(\gamma)}}$.
  2) $\gamma_{i,j} \leftarrow \arg\min_{\gamma} I_i(\gamma)$.
- When an update occurs and the interval ends:
  1) $\eta_i(\gamma_{i,j}) \leftarrow \eta_i(\gamma_{i,j}) + 1$.
  2) $\hat{r}_i(\gamma_{i,j}) \leftarrow \hat{r}_i(\gamma_{i,j}) \left(1 - \frac{1}{\eta_i(\gamma_{i,j})}\right) + \frac{\hat{r}_{i,j}}{\eta_i(\gamma_{i,j})}$.

Note that, for each possible threshold γ, the empirical average cost $\hat{r}_i(\gamma)$ can be greater than 1. Thus, when the UCB index $I(\gamma)$ is calculated, we normalize the empirical costs with the maximum value among them so that the values lie between 0 and 1.

Learning-based RD policy:

Now, we develop the learning-based RD policy by employing the UCB technique. The receiver learns an optimal period τ^* among the possible periods $\tau \in [1, \overline{\tau}]$. Let τ_j denote the period of the j^{th} interval. At the beginning of the j^{th} interval, the receiver collectively decides the update schedule for transmitter i within that interval as shown in Fig. 4. However, a challenge arises. Due to the changing threshold values between consecutive intervals, from the perspective of an individual source i, the update interval may appear somewhat arbitrary. For example, in Fig. 5, the update interval of source 4 is $(t_0+12)-(t_0+4) = 8$ and $(t_0+16)-(t_0+12) = 5$ when τ_j changes from 10 to 2 and 4. Thus, for the purpose of learning an optimal threshold, the receiver traces the empirical average cost of transmitter 1 only, since transmitter 1 has consistent update interval with τ_j.

As in the learning-based TD policy, the average cost during interval j is written as (30) replacing ϵ_s with ϵ_r. In the RD policy, the update interval Δ_j equals τ_j. Let $\hat{r}(\tau)$ and $\eta(\tau)$ denote the empirical average cost (of transmitter 1) for τ and the number of selections for τ, respectively. The learning-based RD policy is operated as in the learning-based TD policy by replacing γ with τ.

We verify the performance of learning-based TD and RD policies through simulations in Section VII.

VI. Extensions to Asymmetric Source Dynamics

In this section, we consider an information source of which state has an asymmetric noise. More specifically, the state x_t of the source evolves as

$$x_{t+1} = x_t + w_t, \quad (31)$$

where

$$w_t = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p-q, \\ -1, & \text{with probability } q, \end{cases} \quad (32)$$

where $p, q \in [0,1]$ such that $p+q \leq 1$ and $\alpha = p - q$. Then, we have that $\mathbb{E}[x_{t+1} - x_t \mid x_t] = \alpha$, i.e., the state x_t is drifted by α. We assume that $\alpha = \frac{m}{\kappa}$, where $\kappa \in \mathbb{N}$, $m \in \mathbb{Z}$ and the greatest common denominator of m and κ is 1, and that α is known to the receiver so that the receiver can update the estimate $\hat{x}(t)$ as

$$\hat{x}_t = \hat{x}_{t-1} + \alpha, \quad (33)$$

when $u_t = 0$, i.e., no update occurs at time t. Then, the error $\varepsilon_t = x_t - \hat{x}_t$ evolves as

$$\varepsilon_{t+1} = \varepsilon_t + z_t, \quad (34)$$

where

$$z_t = \begin{cases} \frac{\kappa - m}{\kappa}, & \text{with probability } p, \\ -\frac{m}{\kappa}, & \text{with probability } 1-p-q, \\ \frac{-\kappa - m}{\kappa}, & \text{with probability } q, \end{cases} \quad (35)$$

[5]Since time and state space are discrete, we can employ the finite-armed Multi-Armed Bandits.

[6]Each transmitter follows the proposed procedure independently, thus we omit subscripts indicating the indices of transmitters.

when $u_0 = 0$.

A. Decentralized Update Paradigm

we first consider a transmitter-driven update policy. Without loss of generality, we assume that $\alpha = \frac{m}{\kappa} > 0$. Let $\gamma_{\min} = -\frac{l_{\min}}{\kappa}$ and $\gamma_{\max} = \frac{l_{\max}}{\kappa}$, where $l_{\min}, l_{\max} \in \mathbb{N}$, be thresholds so that, if $\varepsilon_t + z_t \geq \gamma_{\max}$ (or $\varepsilon_t + z_t \leq \gamma_{\min}$), then the source sends an update to the receiver with value $\gamma_{\max}$ (or $\gamma_{\min}$) and the error evolves toward 0 by the amount of $\gamma_{\max}$ (or $\gamma_{\min}$). Note that this update policy requires one-bit of information for each update, and that the state space of the error ε_t is $\{l/\kappa : l \in \{-l_{\min}, -l_{\min}+1, ..., 0, ..., l_{\max}-1, l_{\max}\}\}$ given $\alpha = \frac{m}{\kappa}$.

In a scenario of symmetric source dynamics, we remark that the error returns to 0 after every transmission and hitting positive and negative thresholds are equally probable. On the other hand, under the proposed TD policy for asymmetric source dynamics, the error may not return to 0 after a transmission and hitting positive and negative thresholds can be different depending on the returning value. In Section VI-C, we show that hitting positive and negative thresholds under the TD policy becomes equally probable as a threshold γ increases. Next, we describe the transition probability under the proposed TD policy.

Let $R = [r_{jk}]_{j,k=\gamma_{\min}+\frac{1}{\kappa},\ldots,\gamma_{\max}-\frac{1}{\kappa}}$,where $r_{jk} = \mathbb{P}(\varepsilon_{t+1} = k \mid \varepsilon_t = j)$ be the transition probability from states j to k. Then, by the one-bit transmitter-driven update policy, we have that

$$r_{jk} = \begin{cases} p, & \text{if } (j,k) = \left(\frac{l}{\kappa}, \frac{l}{\kappa} + \frac{\kappa - m}{\kappa}\right), \\ & \quad l = -l_{\min}+1, ..., l_{\max} - \kappa + m - 1, \\ & \text{or } (j,k) = \left(\gamma_{\max} - \frac{\kappa-m}{\kappa} + \frac{l}{\kappa}, \frac{l}{\kappa}\right), \\ & \qquad l = 0, ..., \kappa - m - 1, \\ 1-p-q, & \text{if } (j,k) = \left(\frac{l}{\kappa}, \frac{l}{\kappa} - \frac{m}{k}\right), \\ & \qquad l = l_{\min} + m + 1, ..., l_{\max} - 1, \\ & \text{or } (j,k) = \left(\gamma_{\min} + \frac{m}{\kappa} - \frac{l}{\kappa}, -\frac{l}{\kappa}\right), \\ & \qquad l = 0, ..., m - 1, \\ q, & \text{if } (j,k) = \left(\frac{l}{\kappa}, \frac{l}{\kappa} - \frac{\kappa+m}{\kappa}\right), \\ & \quad l = l_{\min} + \kappa + m + 1, ..., l_{\max} - 1, \\ & \text{or } (j,k) = \left(\gamma_{\min} + \frac{\kappa+m}{\kappa} - \frac{l}{\kappa}, -\frac{l}{\kappa}\right), \\ & \qquad l = 0, ..., \kappa + m - 1. \end{cases} \tag{36}$$

Since the error evolution ε_t is a finite state Markov chain, there exists a unique steady state distribution π, which can be obtained by solving

$$\pi = \pi R. \tag{37}$$

Further, from the steady state distribution π of the error evolution ε_t, we can obtain the long-term mean squared error $E_{\gamma_{\min},\gamma_{\max}}$ when thresholds are $\gamma_{\min}$ and $\gamma_{\max}$ as

$$E_{\gamma_{\min},\gamma_{\max}} := \lim_{s\to\infty} \frac{1}{s} \sum_{t=1}^{s} \mathbb{E}[\varepsilon_t^2] = \sum_{l=-l_{\min}+1}^{l_{\max}-1} \pi_{l/\kappa} \left(\frac{l}{\kappa}\right)^2. \tag{38}$$

Let $P_u(l) = \mathbb{P}(u_t = 1 \mid \varepsilon_t = l)$ be the probability that an update occurs at time t given that the error at time t is l when thresholds are $\gamma_{\min}$ and $\gamma_{\max}$, which can be written as

$$\begin{aligned} &P^u_{\gamma_{\min},\gamma_{\max}}(l) \\ &= \mathbb{P}(\varepsilon_t + z_t \geq \gamma_{\max} \text{ or } \varepsilon_t + z_t \leq \gamma_{\min} | \varepsilon_t = l) \\ &= \mathbb{P}(z_t \geq \gamma_{\max} - l \text{ or } z_t \leq \gamma_{\min} - l), \end{aligned} \tag{39}$$

which can be obtained by (35). Then, we have

$$P^u_{\gamma_{\min},\gamma_{\max}} := \lim_{t\to\infty} \mathbb{P}(u_t = 1) = \sum_{l=-l_{\min}+1}^{l_{\max}-1} \pi_{l/\kappa} P_u(l/\kappa). \tag{40}$$

Hence, from Lemma 3.1, (38) and (40), the expected average cost $\tilde{g}_{\text{TD}}(\gamma_{\min}, \gamma_{\max})$ of the transmitter-driven policy for a homogeneous n-source scenario is given by

$$\tilde{g}_{\text{TD}}(\gamma_{\min}, \gamma_{\max}) = \mathbb{E}[\tilde{h}(c(K+1)^{\epsilon_s})], \tag{41}$$

where the expectation is taken over a random variable $K \sim Binom\,(n-1, P_u)$, and

$$\tilde{h}(\bar{c}) = \bar{c} P^u_{\gamma_{\min},\gamma_{\max}} + E_{\gamma_{\min},\gamma_{\max}}. \tag{42}$$

B. Centralized Update Paradigm

We now consider a receiver-driven update policy. Since the receiver adjusts its estimate with the expected drift $\alpha = \mathbb{E}[x_{t+1} - x_t \mid x_t] = p - q$, the expected error $\xi(\tau)$ after τ consecutive time slots since the last update is give by

$$\xi(\tau) = (p + q - (p-q)^2)\tau, \tag{43}$$

which can be shown as Lemma 4.1. For completeness, we refer Appendix D. Then, we can use the results of Lemma 4.2 so that an optimal RD policy for a single source scenario has a time-based threshold. Further, as in (23), we have the expected average cost $\tilde{g}_{\text{RD}}(\tau)$ of the receiver-driven update policy for a homogeneous n-source scenario with threshold τ as

$$\begin{aligned} \tilde{g}_{\text{RD}}(\tau) = \frac{c}{n\tau}&((\tau - \tau_0)k^{1+\epsilon_r} + \tau_0(k-1)^{1+\epsilon_r}) \\ &+ \frac{p + q - (p-q)^2}{2}(\tau - 1), \end{aligned} \tag{44}$$

where $k = \lceil n/\tau \rceil$ and $\tau_0 = k\tau - n$.

C. Learning-based update policy

Now, we consider learning-based TD and RD update policies when system parameters p, q, c and ϵ_s (or ϵ_r) are unknown, and we assume that $\alpha = p - q = \frac{m}{\kappa}$, $\kappa \in \mathbb{N}$, $m \in \mathbb{Z}$, is known to both sources and the receiver.

For a learning-based RD policy, we can employ the UCB technique to find an optimal (time-based) threshold τ as in Section V-B. On the other hand, a TD policy for asymmetric dynamics can have asymmetric thresholds $\gamma_{\min} < 0$ and $\gamma_{\max} > 0$, and finding two optimal thresholds requires more time than finding one symmetric optimal threshold for symmetric dynamics. Hence, instead learning asymmetric thresholds $\gamma_{\min}$ and $\gamma_{\max}$, we let sources learn one symmetric optimal threshold $\gamma = -\gamma_{\min} = \gamma_{\max}$ so that we can use the UCB technique as in Section V-B.

Since z_t is an asymmetric random variable with mean 0, we may not have $\mathbb{P}(\varepsilon_t > 0 \mid |\varepsilon_t| \geq \gamma) = \mathbb{P}(\varepsilon_t < 0 \mid |\varepsilon_t| \geq \gamma)$. However, we show that, for a large threshold γ, the error ε_t is equally likely to be positive or negative when it exceeds threshold γ in the following theorem.

Theorem 6.1: For the error evolution ε_t defined in (34), we have

$$\lim_{\gamma\to\infty} \mathbb{P}(\varepsilon_t > 0 \mid |\varepsilon_t| \geq \gamma) = \lim_{\gamma\to\infty} \mathbb{P}(\varepsilon_t < 0 \mid |\varepsilon_t| \geq \gamma). \tag{45}$$

Note that an optimal threshold for TD-G policy increases as the number n of sources increases as shown in the proof of Theorem 5.1 (Appendix E). That is, the error ε_t is equally likely to be positive or negative when it exceeds threshold γ for a sufficiently large n. The theorem can be shown using analysis of Martingales [34]. The detailed proof is in Appendix F.

VII. Simulation Results

In this section, we compare the performance of TD-L, TD-G and RD policies through simulations. Throughout the simulations, we use $p = 0.3$ and $c = 50$ for sources with symmetric dynamics.

It is obvious that TD-G policy outperforms TD-L policy for all $\epsilon_s \geq 0$ since TD-G policy uses more information than TD-L policy, and thus we do not compare between TD-L and TD-G policies. For numerical simulations, we use thresholds $\gamma_L^* = \lfloor \sqrt[4]{12pc} \rfloor$ or $\lceil \sqrt[4]{12pc} \rceil$ for TD-L policy, γ_G^* that minimizes (13) for TD-G policy, and τ^* that minimizes (24) for RD policy[7]. Based on the given threshold, each transmitter either updates the receiver ($u_{i,t} = 1$) or not ($u_{i,t} = 0$) at every time slot t. Then, the average cost $C(t)$ at time slot t is

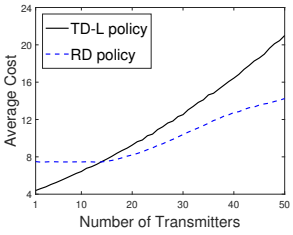
$$C(t) := \frac{1}{tn} \sum_{s=1}^{t} \sum_{i=1}^{n} \left(\varepsilon_{i,t}^2 + u_{i,t} \cdot c \left(\sum_{i=1}^{n} u_{i,t} \right)^{\epsilon} \right), \tag{46}$$

where $\epsilon = \epsilon_s$ for TD-L and TD-G policies and $\epsilon = \epsilon_r$ for RD policy.

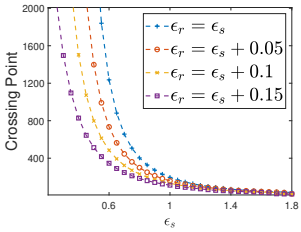
We first compare TD-L and RD policies. We run simulations for $T = 10^4$ time slots, and the results are averaged over 50 repetitions. Fig. 6(a) shows the average cost $C(T)$ at time T when $\epsilon_s = \epsilon_r = 2$. We observe that for a relatively small n, TD-L policy outperforms RD policy. However, as n increases, the gap becomes close to zero and eventually RD policy outperforms TD-L policy. We call the point (the number of transmitters) where RD policy starts to outperform TD-L policy as a *crossing point*. In Fig. 6(a), the crossing point is at 14. Fig. 6(b) shows the crossing point with respect to ϵ_s and ϵ_r. As expected, for relatively large n and ϵ_s, the communication cost of distributed updates dominates the value of (state) information, and the value of information dominates the update cost for small n and ϵ_s.

We now compare TD-G and RD policies. Note that, according to Theorem 5.1, the existence of a crossing point between TD-G and RD policies can be guaranteed only for $\epsilon_s > \epsilon_r > 0$ Fig. 7 shows the ratio of the average cost of RD policy to that of TD-G policy when $\epsilon_s = \epsilon_r = \epsilon$.

[7]Thresholds γ_G^* and τ^* can be found using numerical search methods.



(a) Average cost when $\epsilon_s = \epsilon_r = 2$.



(b) Crossing point with respect to ϵ_s and ϵ_r.

Fig. 6: Performance comparison between TD-L and RD policies.

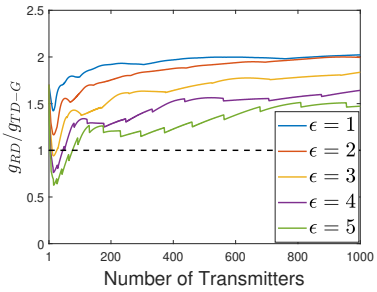
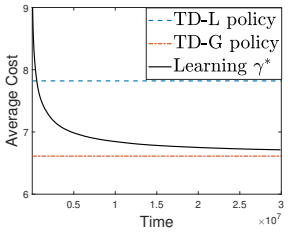


Fig. 7: Performance comparison between TD-G and RD policies when $\epsilon = \epsilon_s = \epsilon_r$.

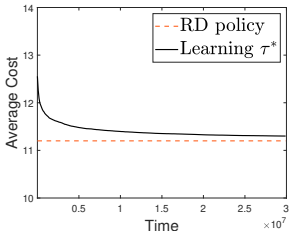
The ratio greater than 1 implies TD-G policy outperforms RD policy. As a specific example, when $\epsilon = 1$, from (55) and (62), we can analytically show that $\lim_{n\to\infty} \frac{g_{\text{RD}}}{g_{\text{TD-G}}} \geq \lim_{n\to\infty} \frac{(4/48^{2/3}+48^{1/3}/6)(p^2cn)^{1/3}-1/6}{(p(2c/p)^{1/3}+c(p/2c)^{2/3})n^{1/3}} \approx 2.08$, which agrees with the simulation result. This implies that when transmitters have global information, i.e., n and ϵ_s, they can adjust their threshold reflecting the distribution of N_t and this leads to significant improvement of the performance of TD-L policy.

We evaluate the learning-based TD and RD policies, where system parameters p, c, ϵ_s and ϵ_r are unknown to both transmitters and the receiver, and n is known to the receiver but not to the transmitters. Only the range of possible values of each parameter is known, and thus each transmitter and the receiver have the set of possible thresholds $\gamma \in [0, \overline{\gamma}]$ and $\tau \in [1, \overline{\tau}]$, respectively. We set $p = 0.3$, $c = 50$, $\epsilon_s = 2$, $\epsilon_r = 1$ and $n = 50$, and assume that $\overline{\gamma} = 10$ and $\overline{\tau} = 30$, respectively. We run simulations for $T = 3 \times 10^7$ time slots.

Fig. 8(a) shows the performance of the learning-based TD policy, which is compared to TD-L and TD-G policies that operate with known system parameters. As shown in Fig. 8(a), the average cost of the learning-based TD policy rapidly approaches that of TD-G policy, which implies that the learning-based TD policies find the global optimal threshold γ_G^*. Fig. 8(b) shows the performance of the learning-based RD policy, which is also compared to RD policy with known parameters. It verifies that the learning-based RD policy finds the optimal threshold τ^* of RD policy. These findings confirm that the findings of our work can be effectively translated into the learning environment where system parameters as well as value and cost functions are unknown.



(a) The learning-based TD policy.



(b) The learning-based RD policy.

Fig. 8: Performance of the learning-based policies when $p = q = 0.4$.

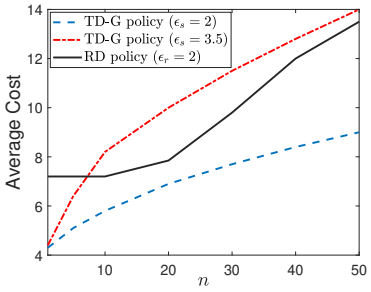
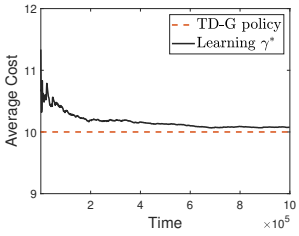


Fig. 9: Performance comparison between TD-G and RD policies when $\epsilon_s = 3.5$ and $\epsilon_r = 2$.

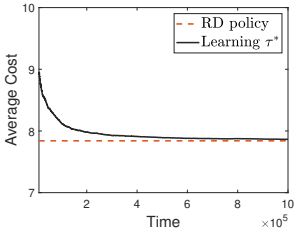
We now consider sources with asymmetric dynamics with $p = 0.4$ and $q = 0.2$ and use $c = 50$. $\epsilon_s = 3.5$ and $\epsilon_r = 2$ throughout the simulations. Fig. 9 shows the average cost $C(T)$ of TD-G (when $\epsilon_s = 2$ and 3.5) and RD (when $\epsilon_r = 2$) policies with respect to the number n of sources at time $T = 10^7$. As can be seen in Fig. 9, there is a crossing point when $\epsilon_s = 3.5$ and $\epsilon_r = 2$. On the other hand, when $\epsilon_s = \epsilon_r = 2$, the existence of a crossing point cannot be guaranteed as discussed in Theorem 5.1. We now evaluate the learning based TD and RD policies described in Section VI-C. Fig. 10 shows the average cost of the learning-based update policies when $n = 20$, $\epsilon_s = 3.5$ (Fig. 10(a)) and $\epsilon_r = 2$ (Fig. 10(b)). The optimal (offline) average costs denoted as TD-G policy in Fig. 10(a) and RD policy in Fig. 10(b) are found by numerical search to minimize the expected average cost in (41) and (44), respectively. It shows that the learning-based update policies find the optimal thresholds γ^* (TD-G policy) and τ^* (RD policy).

VIII. Conclusion

We investigated decentralized (transmitter-driven) and centralized (receiver-driven) update paradigms, where a receiver is updated from multiple sources of which states evolve according to a simple random walk process. In particular, we considered a scenario where each update is accompanied by communication cost, and we modeled communication cost as a superlinear function of the number of simultaneous transmissions at a given time since the transmitters communicate over shared channels. When the cost associated with the information mismatch (error) is the mean squared error, we



(a) The learning-based TD policy.



(b) The learning-based RD policy.

Fig. 10: Performance of the learning-based policies with asymmetric sources of $p = 0.4$ and $q = 0.2$.

obtained the expected average cost for the transmitter-driven and receiver-driven policies, and compared them for different number of transmitters. From the comparison, we provided insights into the tradeoff between the value of fresh information and the cost of distributed communication in the remote tracking of large-scale distributed systems. When simultaneous transmission incurs relatively small communication costs, e.g., small coefficient ϵ or small number n of sources, a decentralized scheme performs better than the centralized scheme, and vice versa. We also developed learning-based policies that asymptotically achieve the minimum costs attained by the optimal policies when the system parameters are unknown. Finally, through numerical simulations, we verified the performance of the proposed policies. Theoretical analysis of the performance of learning-based update policies is an interesting future work in consideration that each transmitter has different update periods. Other interesting future works include studies of heterogeneous sources, vector state estimation and multi-channel systems.

Appendix A
Proof of Lemma 3.1

By the independence of the transmitters' decision $u_{i,t}$ and Theorem 3.1, we have

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}(N_t = k) &= \lim_{t\to\infty}\mathbb{P}\left(\sum_{i=1}^{n} u_{i,t} = k\right)\\ &= \lim_{t\to\infty}\binom{n}{k}\mathbb{P}(u_{i,t}=1)^k\mathbb{P}(u_{i,t}=0)^{n-k} \qquad (47)\\ &= \binom{n}{k}\left(\frac{1}{\mathbb{E}[T]}\right)^k\left(1-\frac{1}{\mathbb{E}[T]}\right)^{n-k},\end{aligned}$$

where $\mathbb{E}[T]$ is the expectation of the inter-renewal interval under the threshold-type update policy with a threshold γ, which is $\frac{2p}{\gamma^2}$ [15].

Appendix B
Proof of Lemma 4.1

With $\varepsilon_0 = 0$, we can write $\varepsilon_\tau = \varepsilon_{\tau-1} + w_{\tau-1}$ for $\tau \geq 1$ as

$$\varepsilon_\tau = \sum_{t=0}^{\tau-1} w_t, \qquad (48)$$

where $\mathbb{E}[w_t]=0$ and $\mathbb{E}[w_t^2]=2p$. Thus, since w_t's are i.i.d., we have

$$\begin{aligned}\xi(\tau)=\mathbb{E}[\varepsilon_\tau^2]&=\mathbb{E}\left[\left(\textstyle\sum_{t=0}^{\tau-1}w_t\right)^2\right]\\&=\mathbb{E}\left[\textstyle\sum_{t=0}^{\tau-1}w_t^2+\sum_{t\neq s}w_tw_s\right]\\&=\textstyle\sum_{t=0}^{\tau-1}\mathbb{E}[w_t^2]=\tau\mathbb{E}[w_1^2]\\&=2p\tau.\end{aligned}\tag{49}$$

Appendix C
Proof of Lemma 4.2

From (20), let $A(s)=\xi(s)+A(s+1)-\lambda$ and $B(s)=c+\phi(1)-\lambda$. Then, an optimal action is to update if $B(s)<A(s)$ and not to update if $B(s)\geq A(s)$. Note that $B(0)=c+\phi(1)-\lambda>\xi(0)+\phi(1)-\lambda=A(0)$ since $\xi(0)=0$ and $c>0$. Thus, $u=0$ is an optimal action for $s=0$. Note that $B(s)=c+\phi(1)-\lambda$ is a constant, and $A(s)$ is an increasing-then-decreasing function or a decreasing function since $A(s+1)-A(s)=\lambda-2ps$.

We show that there exists τ such that $A(s)\leq B(s)$ for $s\leq\tau$ and $A(\tau+1)>B(\tau+1)$ by contradiction. Suppose that such τ does not exist, which implies that $A(s)\leq B(s)$ for all s and thus $u=0$ for all s. Then, the expected cost goes infinity since $\xi(s)=2ps$. If we take an update policy such that $u=1$ for all s, then the expected cost c, which leads to a contradiction. Hence, there exist s such that $A(s)>B(s)$ and, for $\tau=\min\{s:A(s)>B(s)\}$, we have the claim since $A(s)$ is increasing-then-decreasing.

Appendix D
Proof of Equation (43)

With $\varepsilon(0)=0$, we can write $\varepsilon(\tau)=\varepsilon(\tau-1)+w(\tau-1)-\alpha$ for $\tau\geq1$ as

$$\varepsilon(\tau)=\sum_{t=0}^{\tau-1}w_t-\alpha\tau,\tag{50}$$

where $\mathbb{E}[w_t]=p-q$ and $\mathbb{E}[w_t^2]=p+q$. Then, since w_t's are i.i.d., we have $\xi(\tau)=$

$$\begin{aligned}&\mathbb{E}[\varepsilon^2(\tau)]=\mathbb{E}\left[\left(\textstyle\sum_{t=0}^{\tau-1}w_t-\alpha\tau\right)^2\right]\\&=\mathbb{E}\left[\left(\textstyle\sum_{t=0}^{\tau-1}w_t\right)^2-2\alpha\tau\sum_{t=0}^{\tau-1}w_t+\alpha^2\tau^2\right]\\&=\sum_{t=0}^{\tau-1}\mathbb{E}[w_t^2]+\sum_{t\neq s}\mathbb{E}[w_t]\mathbb{E}[w_s]-2\alpha\tau\sum_{t=0}^{\tau-1}\mathbb{E}[w_t]+\alpha^2\tau^2\\&=(p+q)\tau+(p-q)^2(\tau^2-\tau)-2(p-q)^2\tau^2+(p-q)^2\tau^2\\&=(p+q-(p-q)^2)\tau.\end{aligned}\tag{51}$$

Appendix E
Proof of Theorem 5.1

We first show the asymptotic lower bound for TD policies. The expected average cost $\tilde{g}_{\text{TD}}(\gamma)$ of TD policy given threshold γ is, from (13), given by

$$\tilde{g}_{\text{TD}}(\gamma)=\tfrac{2}{\gamma^2}\left(pc\mathbb{E}\left[(K+1)^{\epsilon_s}\right]+\tfrac{\gamma^2(\gamma^2-1)}{12}\right).\tag{52}$$

TD-L policy: Let $f(x)=(x+1)^{\epsilon_s}$ and $\mu=\mathbb{E}[K]=\frac{2p(n-1)}{\gamma^2}$. By expanding the Taylor series of $f(K)$ around μ by the second-order term, we have

$$f(K)=f(\mu)+f'(\mu)(K-\mu)+\tfrac{f''(\alpha)(K-\mu)^2}{2}\tag{53}$$

for some $\alpha\in[0,n-1]$. By taking the expectation on both sides, we have

$$\mathbb{E}[(K+1)^{\epsilon_s}]=(\mu+1)^{\epsilon_s}+\tfrac{\mathbb{E}[f''(\alpha)(K-\mu)^2]}{2}.\tag{54}$$

If $\epsilon_s\geq1$, then $f(x)$ is convex and thus $f''(x)\geq0$ for all $x\in[0,n-1]$. Then, we have

$$\mathbb{E}[(K+1)^{\epsilon_s}]\geq\left(\tfrac{2p(n-1)}{\gamma^2}+1\right)^{\epsilon_s},\tag{55}$$

and thus $\mathbb{E}[(K+1)^{\epsilon_s}]=\Omega(n^{\epsilon_s})$ with $\gamma_L^*=\lfloor\sqrt[4]{12pc}\rfloor$ or $\lceil\sqrt[4]{12pc}\rceil$.

Now, suppose that $0<\epsilon_s<1$. By expanding the Taylor series of $f(K)$ around μ by the third-order term, we have

$$\begin{aligned}f(K)=f(\mu)&+f'(\mu)(K-\mu)\\&+\tfrac{f''(\mu)(K-\mu)^2}{2}+\tfrac{f^{(3)}(\alpha)(K-\mu)^3}{6}\end{aligned}\tag{56}$$

for some $\alpha\in[0,n-1]$. By taking the expectation on both sides, we have $\mathbb{E}[(K+1)^{\epsilon_s}]=$

$$(\mu+1)^{\epsilon_s}+\tfrac{f''(\mu)\text{Var}(K)}{2}+\tfrac{\mathbb{E}[f^{(3)}(\alpha)(K-\mu)^3]}{6}.\tag{57}$$

Note that $f^{(3)}(x)\geq0$ for all $x\in[0,n-1]$ since $f'(x)=\epsilon_s(x+1)^{\epsilon_s-1}$ is convex for $\epsilon_s\in(0,1)$, and $\mathbb{E}[(K-\mu)^3]\geq0$ since, for $X\sim B(m,q)$, $\mathbb{E}[(X-\mathbb{E}[X])^3]=mq(2q-1)(q-1)$ and in our case $q=\frac{2p}{\gamma^2}<0.5$ since $c\geq2p$[8] . Thus, we have

$$\mathbb{E}[(K+1)^{\epsilon_s}]\geq\left(\tfrac{2p(n-1)}{\gamma^2}+1\right)^{\epsilon_s}+\tfrac{f''(\mu)\text{Var}(K)}{2},\tag{58}$$

where $f''(x)=\epsilon_s(\epsilon_s-1)(x+1)^{\epsilon_s-2}$, $\mu=\frac{2p(n-1)}{\gamma^2}$ and $\text{Var}(K)=(n-1)\left(\frac{2p}{\gamma^2}\right)\left(1-\frac{2p}{\gamma^2}\right)$, and thus $\mathbb{E}[(K+1)^{\epsilon_s}]=\Omega(n^{\epsilon_s})$.

TD-G policy: If $\epsilon_s\geq1$, by (52) and (55), we have

$$\begin{aligned}\tilde{g}_{\text{TD}}(\gamma)&\geq\tfrac{2}{\gamma^2}\left(pc\left(\tfrac{2p(n-1)}{\gamma^2}+1\right)^{\epsilon_s}+\tfrac{\gamma^2(\gamma^2-1)}{12}\right)\\&\geq\tfrac{2pc(2p(n-1))^{\epsilon_s}}{\gamma^{2\epsilon_s+2}}+\tfrac{\gamma^2-1}{6}=\overline{g}_{\text{TD}}(\gamma).\end{aligned}\tag{59}$$

Since $\overline{g}_{\text{TD}}(\gamma)$ is convex in γ, by solving $\frac{d\overline{g}_{\text{TD-G}}}{d\gamma}=0$, we have $\gamma^*=\sqrt[2\epsilon_s+4]{6pc(2\epsilon_s+2)(2p(n-1))^{\epsilon_s}}$, with which we have $g_{\text{TD-G}}(n,c,\epsilon_s)=\Omega(n^{\frac{\epsilon_s}{\epsilon_s+2}})$ since $g_{\text{TD}}(n,c,\epsilon_s)=\min_{\gamma>0}\tilde{g}_{\text{TD}}(\gamma)\geq\min_{\gamma>0}\overline{g}_{\text{TD}}(\gamma)$.

If $0<\epsilon_s<1$, by (52) and (58), we have

$$\begin{aligned}\tilde{g}_{\text{TD}}(\gamma)&\geq\tfrac{\gamma^2-1}{6}+\tfrac{2pc(2p(n-1))^{\epsilon_s}}{\gamma^{2\epsilon_s+2}}\left(1-\tfrac{(n-1)}{32(\mu+1)^2}\right)\\&\geq\tfrac{\gamma^2-1}{6}+\tfrac{2pc(2p(n-1))^{\epsilon_s}}{\gamma^{2\epsilon_s+2}}\left(1-o(n)\right),\end{aligned}\tag{60}$$

where $o(n)=\frac{\gamma^4}{128p^2(n-1)}$. Suppose that, for some $\delta\in(0,1)$, there exists an N_δ such that $o(n)\leq\delta$ for all $n\geq N_\delta$. Then, for $n\geq N_\delta$, we have

$$\tilde{g}_{\text{TD}}(\gamma)\geq\tfrac{\gamma^2-1}{6}+\tfrac{2pc(2p(n-1))^{\epsilon_s}}{\gamma^{2\epsilon_s+2}}(1-\delta)=\overline{g}_{\text{TD}}(\gamma).\tag{61}$$

[8]Note that, for $X\sim\mathcal{N}(\mu,\sigma^2)$, the equality holds in (58) since $\mathbb{E}[(X-\mathbb{E}[X])^3]=0$. Since a Binomial distribution, $B(m,q)$, can be approximated by a Gaussian distribution, $\mathcal{N}(mq,mq(1-q))$, for a sufficiently large m, the gap in inequality (58) vanishes as the number n of transmitters increases.

Then, by minimizing $\overline{g}_{\text{TD}}(\gamma)$, we have an optimal threshold $\gamma^* = \sqrt[2\epsilon_s+4]{(1-\delta)6pc(2\epsilon_s+2)(2p(n-1))^{\epsilon_s}}$, with which we have $o(n) = O(n^{\frac{\epsilon_s-2}{\epsilon_s+2}})$. Since $\epsilon_s \in (0,1)$, the result accords with the assumption on $o(n)$. Hence, we have $g_{\text{TD-G}}(n,c,\epsilon_s) = \Omega(n^{\frac{\epsilon_s}{\epsilon_s+2}})$.

RD policy: Under the RD policy, the expected average cost $\tilde{g}_{\text{RD}}(\tau)$ given by τ can be bounded, from (23), as

$$\tilde{g}_{\text{RD}}(\tau) \leq \frac{ck^{1+\epsilon_r}}{n} + p(\tau-1) = \overline{g}_{\text{RD}}(\tau), \tag{62}$$

where $k = \lceil n/\tau \rceil$. Since $g_{\text{RD}}(n,c,\epsilon_r) = \min_{\tau\geq 1} \tilde{g}_{\text{RD}}(\tau) \leq \min_{\tau\geq 1} \overline{g}_{\text{RD}}(\tau) \leq \overline{g}_{\text{RD}}(\tau')$ for any $\tau' \geq 1$, by letting $\tau' = \sqrt[\epsilon_r+2]{(1+\epsilon_r)cn^{\epsilon_r}/p}$, we have $g_{\text{RD}}(n,c,\epsilon_r) = O(n^{\frac{\epsilon_r}{\epsilon_r+2}})$.

Appendix F
Proof of Theorem 6.1

We show that the error $\varepsilon(t)$ is equally likely to be positive or negative when exceeds threshold $\gamma = \frac{l}{\kappa}$, $l, \kappa \in \mathbb{N}$, as the number n of sources increases using analysis of Martingales [34].

First, note that $\mathbb{E}[z_t] = \mathbb{E}[w_t] - \alpha = 0$ in (34) and thus $\mathbb{E}[\varepsilon_{t+1} \mid \varepsilon_t] = \varepsilon_t$, i.e., ε_t is a martingale.

Let $\tau := \min\{t \geq 1 \ : \ |\varepsilon_t| \geq \gamma\}$ with $\varepsilon_0 = k$, and let

$$\begin{aligned} h_\gamma(k) &:= \mathbb{P}(\varepsilon_\tau \geq \gamma \mid \varepsilon_0 = k) \\ h_{-\gamma}(k) &:= \mathbb{P}(\varepsilon_\tau \leq \gamma \mid \varepsilon_0 = k). \end{aligned} \tag{63}$$

When the error ε_t exceeds threshold γ, it returns towards 0 by γ, and thus we have a starting point $k \in ((1-\kappa-m)/\kappa, (-1+\kappa-m)/\kappa)$. By the martingale stopping theorem (Theorem 6.2.2 in [34]), we have, with $\varepsilon_0 = k$, that

$$\begin{aligned} k &= \mathbb{E}[\varepsilon_\tau] \\ &= \mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \geq \gamma]\mathbb{P}(\varepsilon_\tau \geq \gamma) \\ &\quad + \mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \leq -\gamma]\mathbb{P}(\varepsilon_\tau \leq -\gamma). \end{aligned} \tag{64}$$

Using $\mathbb{P}(\varepsilon_\tau \geq \gamma) + \mathbb{P}(\varepsilon_\tau \leq -\gamma) = 1$, we can reorganize (64) as

$$\begin{aligned} h_{-\gamma}(k) &= \mathbb{P}(\varepsilon_\tau \leq -\gamma) \\ &= \tfrac{\mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \geq \gamma] - k}{\mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \geq \gamma] - \mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \leq -\gamma]}. \end{aligned} \tag{65}$$

Further, from (35), we have that

$$\begin{aligned} \gamma \leq \mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \geq \gamma] &\leq \gamma + \tfrac{\kappa-m}{\kappa} \\ -\gamma - \tfrac{\kappa+m}{\kappa} \leq \mathbb{E}[\varepsilon_\tau \mid \varepsilon_\tau \leq -\gamma] &\leq -\gamma. \end{aligned} \tag{66}$$

Hence, we can obtain

$$\tfrac{1}{2} - \tfrac{1+k}{2\gamma+2} \leq h_{-\gamma}(k) \leq \tfrac{1}{2} + \tfrac{\kappa-m}{2\kappa\gamma}, \tag{67}$$

and we have

$$\lim_{\gamma\to\infty} h_{-\gamma}(k) = \lim_{\gamma\to\infty} h_\gamma(k) = \frac{1}{2} \tag{68}$$

for $k \in ((1-\kappa-m)/\kappa, (-1+\kappa-m)/\kappa)$.

References

[1] S. Kang, A. Eryilmaz, and C. Joo, "Comparison of decentralized and centralized update paradigms for remote tracking of distributed dynamic sources," in *IEEE INFOCOM*, 2021, pp. 1–10.

[2] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2015.

[3] A. Kosta, N. Pappas, V. Angelakis, *et al.*, "Age of information: A new concept, metric, and tool," *Foundations and Trends® in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[4] 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 Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[5] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. INFOCOM*, 2018.

[6] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in *MobiHoc*, 2019, pp. 121–130.

[7] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "Aoi-penalty minimization for networked control systems with packet loss," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2019, pp. 189–196.

[8] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the wiener process over a channel with random delay," in *IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 321–325.

[9] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *IEEE WiOpt*, 2019.

[10] J. Chakravorty and A. Mahajan, "Remote estimation over a packet-drop channel with markovian state," *IEEE Transactions on Automatic Control*, 2019.

[11] K. Huang, W. Liu, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Real-time remote estimation with hybrid arq in wireless networked control," *IEEE Transactions on Wireless Communications*, 2020.

[12] X. Gao, E. Akyol, and T. Başar, "Optimal estimation with limited measurements and noisy communication," in *IEEE Conference on Decision and Control (CDC)*, 2015, pp. 1775–1780.

[13] ——, "Optimal sensor scheduling and remote estimation over an additive noise channel," in *American Control Conference (ACC), 2015*, 2015, pp. 2723–2728.

[14] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order lti systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[15] J. Yun, C. Joo, and A. Eryilmaz, "Optimal real-time monitoring of an information source under communication costs," in *IEEE Conference on Decision and Control (CDC)*, 2018, pp. 4767–4772.

[16] W. Liu, D. E. Quevedo, Y. Li, K. H. Johansson, and B. Vucetic, "Remote state estimation with smart sensors over markov fading channels," *IEEE Transactions on Automatic Control*, vol. 67, no. 6, pp. 2743–2757, 2021.

[17] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[18] Y. Mo, E. Garone, and B. Sinopoli, "On infinite-horizon sensor scheduling," *Systems & control letters*, vol. 67, pp. 65–70, 2014.

[19] D. Han, J. Wu, H. Zhang, and L. Shi, "Optimal sensor scheduling for multiple linear dynamical systems," *Automatica*, vol. 75, pp. 260–270, 2017.

[20] S. Wu, X. Ren, S. Dey, and L. Shi, "Optimal scheduling of multiple sensors over shared channels with packet transmission constraint," *Automatica*, vol. 96, pp. 22–31, 2018.

[21] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Random access design for wireless control systems," *Automatica*, vol. 91, pp. 1–9, 2018.

[22] J. Yun and C. Joo, "Information update for multi-source remote estimation in real-time monitoring networks," in *International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 633–635.

[23] W. Liu, D. E. Quevedo, K. H. Johansson, B. Vucetic, and Y. Li, "Remote state estimation of multiple systems over multiple markov fading channels," *arXiv preprint arXiv:2104.04181*, 2021.

[24] W. Liu, D. E. Quevedo, B. Vucetic, and Y. Li, "Remote state estimation of multiple systems over semi-markov wireless fading channels," *arXiv preprint arXiv:2203.16826*, 2022.

[25] W. Liu, K. Huang, D. E. Quevedo, B. Vucetic, and Y. Li, "Deep reinforcement learning for wireless scheduling in distributed networked control," *arXiv preprint arXiv:2109.12562*, 2021.

[26] A. Molin and S. Hirche, "Optimal event-triggered control under costly observations," in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010)*, 2010.
[27] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic control over shared wireless channels," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3140–3155, 2015.
[28] O. Ayan, M. Vilgelm, and W. Kellerer, "Optimal scheduling for discounted age penalty minimization in multi-loop networked control," in *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2020, pp. 1–7.
[29] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.
[30] S. Kang, A. Eryilmaz, and N. B. Shroff, "Remote tracking of distributed dynamic sources over a random access channel with one-bit updates," in *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*. IEEE, 2021, pp. 1–8.
[31] R. G. Gallager, *Discrete stochastic processes*. Springer Science & Business Media, 2012, vol. 321.
[32] S. Gollakota, S. D. Perli, and D. Katabi, "Interference alignment and cancellation," in *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, 2009, pp. 159–170.
[33] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.
[34] S. M. Ross, J. J. Kelly, R. J. Sullivan, W. J. Perry, D. Mercer, R. M. Davis, T. D. Washburn, E. V. Sager, J. B. Boyce, and V. L. Bristow, *Stochastic processes*. Wiley New York, 1996, vol. 2.

Sunjung Kang received her M.S. degree from the school of ECE at Ulsan National Institute of Science and Technology (UNIST) in 2018. She is currently a Ph.D. student in the department of ECE at The Ohio State University. Her research interests include the age of information, remote estimation and optimization.

Atilla Eryilmaz (S'00 / M'06 / SM'17) received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2001 and 2005, respectively. Between 2005 and 2007, he worked as a Postdoctoral Associate at the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. Since 2007, he has been at The Ohio State University, where he is currently a Professor and the Graduate Studies Chair of the Electrical and Computer Engineering Department.

Dr. Eryilmaz's research interests span optimal control of stochastic networks, machine learning, optimization, and information theory. He received the NSF-CAREER Award in 2010 and two Lumley Research Awards for Research Excellence in 2010 and 2015. He is a co-author of the 2012 IEEE WiOpt Conference Best Student Paper, subsequently received the 2016 IEEE Infocom, 2017 IEEE WiOpt, 2018 IEEE WiOpt, and 2019 IEEE Infocom Best Paper Awards. He has served as: a TPC co-chair of IEEE WiOpt in 2014, ACM Mobihoc in 2017, and IEEE Infocom in 2022; an Associate Editor (AE) of IEEE/ACM Transactions on Networking between 2015 and 2019; an AE of IEEE Transactions on Network Science and Engineering between 2017 and 2022; and is currently an AE of the IEEE Transactions on Information Theory since 2022.

Changhee Joo Changhee Joo received the Ph.D. degree from Seoul National University in 2005. He was with Purdue University and The Ohio State University, and then with Korea University of Technology and Education (KoreaTech) and Ulsan National Institute of Science and Technology (UNIST). Since 2019, he has been with Korea University. His research interests are in the broad areas of networking and reinforcement learning. He was a recipient of the IEEE INFOCOM 2008 Best Paper Award, the KICS Haedong Young Scholar Award (2014), the ICTC 2015 Best Paper Award, and the GAMENETS 2018 Best Paper Award. He has served as a general chair of ACM MobiHoc 2022, an Associate Editor of the *IEEE/ACM Transactions on Networking* between 2016 and 2020, an Associate Editor of the *IEEE Transactions Vehicular Technology* since 2020, a Division Editor of the *Journal of Communications and Networks* since 2020.