2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Optimal Scheduling for Remote Estimation with an Auxiliary Transmission Scheme

Zitian Li, Lixin Yang, Yijin Jia, Zenghong Huang, Weijun Lv, Yong Xu

Abstract— **The remote estimation for multiple systems is considered in this paper. Some smart sensors observe the systems and run Kalman filters to compute their state estimates, which are then transmitted to a remote estimator via high frequency wave band. The path loss and signal attenuation make communication at high frequency unreliable. To improve estimation performance, an auxiliary transmission scheme is proposed, where a complementary channel with low frequency wave band is deployed to transmit a duplication of a local state estimate. Since the auxiliary channel consumes extra energy and occupies limited bandwidth, the optimal scheduling needs to be studied, i.e., to determine whether or which sensor to use the auxiliary channel. To tackle this issue, we establish a Markov decision process (MDP) to formulate the optimal scheduling, and prove that the optimal policy is deterministic and stationary. Furthermore, the threshold structure is verified for the optimal policy. The deep reinforcement learning is introduced to approximate an optimal policy. Finally, the threshold structure and the deep reinforcement learning is validated by a numerical example.**

I. Introduction

With the rapid development of communication technology, networked control systems [1], [2] have become popular in both theoretical research and engineering application. Compared with the traditional control system which depends on the point to point communication technology, some useful components of networked control systems are shared in a reliable network. For theoretical research, plenty of scholars pay attention to its stability [3], [4] and performance analysis [5], [6]. For engineering, networked control systems are applied in numbers of fields, such as electric power [7] and medical treatment [8], etc.

Among all the studies of networked control systems, remote estimation is a hot topic [9], [10]. In this scheme, the Kalman filter is run by a sensor or a remote estimator, which depends on the computation capacity of the sensor. However, bandwidth limit [11], [12], power constraint [13], [14] and packet drop [15], [16] inevitably exist. In view of it, many scholars investigate sensor scheduling protocol or transmission strategy to obtain a trade-off between the quality of estimation, communication and power consumption. Some of them get nice research results. To balance estimation performance and communication load, a deterministic event trigger was proposed in [17], where the measurement was allowed to be transmitted to the remote estimator only if the innovation exceeded a given threshold. To preserve the Gaussian property and obtain a closed-form mean square error estimator, a stochastic event trigger was further designed in [18]. Taking into account a composite convex cost function, the optimal scheduling for a single sensor under uncertain channels was studied in [19]. Besides, the optimal policy for multiple sensors over lossy and limited bandwidth channels was further investigated in [20].

The 5th Generation (5G) mobile communication is widely used today [21], [22], and it consists of bands with different frequencies. The trade-off between the capacity of the enhanced Mobile BroadBand (eMBB) and the Ultra-Reliable Low-Latency Communication (URLLC) is a challenging issue. To achieve it, a hybrid transmission strategy was proposed in [23]. In this strategy, the high frequency band and the low frequency band are both considered to transmit data, where the low frequency band is utilized to enhance the reliability of transmission. For example, if the quality of communication (such as the Signal to Noise Ratio) exceeds a certain threshold, only the high frequency band is used, otherwise the data is transmitted by two bands simultaneously. The large capacity of the high frequency band and the reliability of the low frequency are combined in this hybrid transmission strategy. It motivates us to develop a new auxiliary transmission scheme for remote estimation, where the redundant transmission channel is considered to improve the reliability of transmission. In addition, the optimal multiple sensors scheduling to balance the estimation quality and the power consumption is worthy of research.

Motivated by the above discussion, this paper focuses on the remote estimation problem by scheduling multiple sensors. Most existing works used to deal with bandwidth constraint and they assumed that all the channels possess the same communication characteristic. It is interesting to schedule the sensors from the perspective of redundant transmission channels, which can be regarded as an effective method to improve the reliability of transmission.

Notations: $\mathbb{N}$ denotes the set of non-negative integers. $\mathbb{R}^n$, $\mathbb{R}^{n\times n}$ represent n-dimensional real vectors, and $n \times n$-dimensional real matrices, respectively. $M \succ (\succeq) 0$ is a positive definite (semidefinite) matrix. $\rho(M)$, M^T, and $\mathrm{Tr}(M)$ indicate the spectral radius, transpose, and trace of a matrix

This work was supported in part by Key Area Research and Development Program of Guangdong Province (2021B0101410005)the Natural Science Foundation of Guangdong Province, China (2021B1515420008) , the National Natural Science Foundation of China under Grants (62121004, U22A2044, 62206063), and he China Postdoctoral Science Foundation (2022TQ0084).

Z. Li, Y. Jia, Z. Huang, W. Lv, Y. Xu are with the Guangdong Provincial Key Laboratory of Intelligent Decision and Cooperative Control, School of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: xuyong809@163.com).

L. Yang is with the School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, QLD 4000, Australia (Email:yangl25@qut.edu.au).

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

M. $\mathbb{P}[x]$ and $\mathbb{E}[x]$ denote the probability and the expectation of a random variable x, respectively. $\mathcal{N}(0,\Sigma)$ is a Gaussian distribution with zero-mean and covariance Σ.

II. Problem Formulation

A. System Model

Consider the remote estimation problem for the following discrete linear time-invariant (LTI) systems:

$$x_{i,k+1} = A_i x_{i,k} + w_{i,k}, \tag{1}$$

$$y_{i,k} = C_i x_{i,k} + v_{i,k}, \tag{2}$$

where $i = 1,\dots,N$ denotes the system index. $x_{i,k} \in \mathbb{R}^{n_i}$ represents the state of the i-th system, and $y_{i,k} \in \mathbb{R}^{m_i}$ denotes the measurement of the i-th smart sensor. The noise of the system and the measurement are represented by $w_{i,k}$ and $v_{i,k}$, which are mutually independent random variables, and obey the Gaussian distributions $w_{i,k} \sim \mathcal{N}(0,Q_i)$, $v_{i,k} \sim \mathcal{N}(0,R_i)$. Assume that the noise covariances Q_i and R_i are positive semi-definite and positive definite, respectively. In addition, the initial system state $x_{i,0}$ is a zero-mean Gaussian variable with covariance $P_{i,0} \succeq 0$. Hypothetically, $(A_i,\sqrt{Q_i})$ is controllable and (A_i,C_i) is observable.

Each smart sensor can calculate a local state estimate $\hat{x}^s_{i,k}$ and a corresponding error covariance $P^s_{i,k}$ by a local Kalman filter, respectively, which are defined as follows:

$$\hat{x}^s_{i,k} \triangleq \mathbb{E}[x_{i,k}|y_1,\dots,y_k], \tag{3}$$

$$P^s_{i,k} \triangleq \mathbb{E}[(x_{i,k}-\hat{x}^s_{i,k}(x_{i,k}-\hat{x}^s_{i,k})^T|y_1,\dots,y_k]. \tag{4}$$

More in detail, they are computed based on the following Kalman filtering [24]:

$$\begin{aligned}
&\hat{x}^s_{i,k+1|k} = A_i\hat{x}^s_{i,k},\ P^s_{i,k+1|k} = A_i P^s_{i,k} A_i^T + Q_i,\\
&\hat{x}^s_{i,k+1} = \hat{x}^s_{i,k+1|k} + \check{K}_{i,k+1}(y_{i,k+1} - C_i\hat{x}^s_{i,k+1|k}),\\
&P^s_{i,k+1} = (I - \check{K}_{i,k+1}C_i)P^s_{i,k+1|k},
\end{aligned}$$

where $\check{K}_{i,k+1} = P^s_{i,k+1|k}C_i^T(C_iP^s_{i,k+1|k}C_i^T + R_i)^{-1}$ represents the local Kalman gain. $\hat{x}^s_{i,k+1|k}$ and $P^s_{i,k+1|k}$ denote the prediction and the predicted error covariance, respectively. Based on the assumption of controllability and observability, the local Kalman filter exponentially attains a steady state. In other words, the error covariance satisfies

$$\lim_{k\to\infty} P^s_{i,k} = \overline{P}_i,\ i = 1,\dots,N, \tag{5}$$

where $\overline{P}_i$ is a constant positive definite matrix. Since this work focuses on the performance over infinite time horizon, we assume that $P^s_{i,k}$ has achieved $\overline{P}_i$.

B. Auxiliary Transmission Scheme

As illustrated in Fig. 1, the sensors transmit their local state estimate to a remote estimator. To meet the demand for the bandwidth capacity, the high frequency wave is used for the sensors to transmit data, such as millimetre wave (mmWave) with a frequency range from 24.25 GHz to 52.6 GHz. The range is divided into N frequency bands for the sensors, which are denoted as $HF_1,\dots,HF_N$. However, the packet dropouts more likely occur at high frequency due to the signal attenuation. $\gamma_{i,k} = 1$ indicates that the local state estimate is successfully transmitted from the i-th sensor to the remote estimator, and $\gamma_{i,k} = 0$, otherwise. The packet dropout probability is

$$\mathbb{E}[\gamma_{i,k} = 0] = \lambda_i. \tag{6}$$

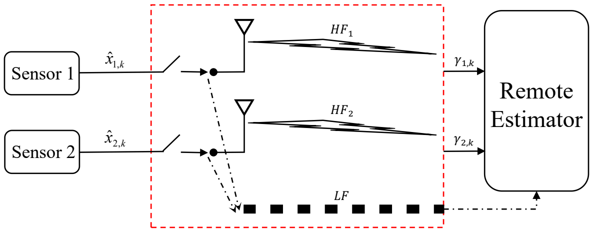

Fig. 1. System model.

To improve the transmission reliability, an auxiliary transmission scheme is proposed in this work, which utilizes the low frequency wave to transmit a duplication of the local state estimate. The low frequency wave has a more reliable transmission, whose packet dropout probability is $\lambda^* < \min_i \lambda_i$. However, due to the limited bandwidth of the low frequency range, only one sensor is allowed to use the auxiliary channel, which is denoted as LF. Besides, each transmission of the auxiliary channel consumes E energy cost of the sensor. Therefore, in view of the energy conservation and the channel occupancy, we need to determine 1) whether to use the auxiliary channel, and 2) which sensor to transmit a duplicate state estimate over the auxiliary channel. Define $u_{i,k}$ as the decision variable for i-th sensor, i.e., $u_{i,k} = 1$ if the i-th sensor uses both HF_i and LF channels to transmit data, and $u_{i,k} = 0$, otherwise, which satisfies $\sum_{i=1}^N u_{i,k} \le 1$.

C. Remote Estimator

The value of $\gamma_{i,k}$ indicates whether the remote estimator receives the local state estimate $x^s_{i,k}$. If the remote estimator unsuccessfully receives the data packet, it only predicts the state estimate. Thus, $\hat{x}_{i,k+1}$ and $P_{i,k+1}$ is updated by

$$(\hat{x}_{i,k},P_{i,k}) = \begin{cases} (\hat{x}^s_{i,k},\overline{P}_i) & \text{if } \gamma_{i,k} = 1,\\ (A_i\hat{x}_{i,k-1},h_i(P_{i,k-1})) & \text{if } \gamma_{i,k} = 0. \end{cases} \tag{7}$$

where $h_i(X) = A_iXA_i^T + Q_i$ represents the one-step prediction error covariance. Based on the information of whether or not the remote estimator receives the data packet, we can obtain the number of consecutive packet dropouts of each system. Here, let $\tau_{i,k}$ denote the stopping time of each system, which is the number of consecutive transmission failures up to time k,

$$\tau_{i,k} \triangleq k - \max\{0 \le t \le k : \gamma_{i,t} = 1\}, \tag{8}$$

where the stopping times of the systems are mutually independent. According to the previous stopping time, $\tau_{i,k}$ can be calculated by

$$\tau_{i,k} = \begin{cases} \tau_{i,k-1} + 1, & \text{if } \gamma_{i,k} = 0,\\ 0, & \text{if } \gamma_{i,k} = 1. \end{cases} \tag{9}$$

Based on the stopping time $\tau_{i,k}$, the error covariance at the remote estimator can be computed by $\tau_{i,k}$ as

$$P_{i,k} = h_i^{\tau_{i,k}}(\overline{P}_i).$$

D. Problem of Interest

In this article, we consider the auxiliary transmission scheme, where the *LF* channel is used to transmit the same information to improve the remote estimation performance. Define a policy as $\mathbf{u} \triangleq \{u_{i,0}, u_{i,1}, \ldots\}$. The objective is to determine $u_{i,k}$ at each time to minimize the combination of the expected average error covariance and energy consumption, i.e.,

$$\min_{\mathbf{u}\in\mathbb{U}} J(\mathbf{u}), \tag{10}$$

with

$$J(\mathbf{u}) \triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \sum_{i=1}^{N} \mathbb{E}\left[\mathrm{Tr}(P_{i,k}) + \beta u_{i,k} E\right], \tag{11}$$

where $\mathbb{U}$ is the set of all policies, and β is a constant weight. Taking into account the objective (10), one needs to achieve the tradeoff between the energy consumption and the remote estimation performance.

III. Markov Decision Process

During the transmission process, the remote estimator can detect whether the data packet is successfully received, i.e., $\gamma_{i,k} = 1$ or not. Hence, the stopping time $\tau_{i,k}$ can be computed by equation (9). Here, we denote a quadruple $\{\mathscr{S}, \mathscr{A}, \mathscr{P}, \mathscr{C}\}$ as an MDP.

State space: the state space is defined as $\mathscr{S} \triangleq \mathbb{N}^N$ and $s_k \triangleq (\tau_{1,k-1}, \ldots, \tau_{N,k-1}) \in \mathscr{S}$, which consists of the stopping time of each sensor at time instant k.

Action space: the action space is defined as $\mathscr{A} \triangleq \{0,1\}^N$, in which $a_k \triangleq (u_{i,k}, \ldots, u_{N,k})$ indicates the use of the auxiliary channel for each sensor. Note that the action a_k satisfies $\sum_{i=1}^N u_{i,k} \leq 1$ for all k.

Transition probability: under the condition that the packet dropout probability of each channel is λ_i, we have the transition probability as follows,

$$\mathscr{P}(s_{k+1}|s_k, a_k) \triangleq \prod_{i=1}^{N} \mathbb{P}(\tau_{i,k}|\tau_{i,k-1}, u_{i,k}), \tag{12}$$

where

$$\mathbb{P}(\tau_{i,k}|\tau_{i,k-1}, u_{i,k}) = \begin{cases} \lambda_i & \text{if } u_{i,k} = 0 \text{ and } \tau_{i,k} = \tau_{i,k-1} + 1, \\ 1-\lambda_i & \text{if } u_{i,k} = 0 \text{ and } \tau_{i,k} = 0, \\ \lambda_i \lambda^* & \text{if } u_{i,k} = 1 \text{ and } \tau_{i,k} = \tau_{i,k-1} + 1, \\ 1-\lambda_i\lambda^* & \text{if } u_{i,k} = 1 \text{ and } \tau_{i,k} = 0. \end{cases} \tag{13}$$

Cost function: With the state and the action, the immediate cost value is obtained, and the state transfers to the next state as well. According to the choice of different actions, define the cost function $\mathscr{C}(\cdot,\cdot)$ as

$$\mathscr{C}(s_k, a_k) \triangleq \sum_{i=1}^{N} \mathbb{E}\left[\mathrm{Tr}(P_{i,k}) + \beta u_{i,k} E\right]. \tag{14}$$

If the auxiliary channel is idle, the cost function is

$$\mathscr{C}(s_k, a_k) = \sum_{i=1}^{N} \mathrm{Tr}(\lambda_i h_i^{\tau_{i,k-1}+1}(\overline{P}_i) + (1-\lambda_i)\overline{P}_i).$$

If the auxiliary channel is used by the i-the sensor, the cost function is computed by

$$\begin{aligned} \mathscr{C}(s_k, a_k) &\triangleq \mathrm{Tr}(\lambda_i\lambda^* h_i^{\tau_{i,k-1}+1}(\overline{P}_i) + (1-\lambda_i\lambda^*)\overline{P}_i + \beta E) \\ &+ \sum_{j=1, j\neq i}^{N} \mathrm{Tr}(\lambda_j h_j^{\tau_{j,k-1}+1}(\overline{P}_j) + (1-\lambda_j)\overline{P}_j), \end{aligned}$$

where the application of the auxiliary transmission can improve the transmission efficiency of the corresponding system, but it brings a cost of the additional computation E to the system.

Define a policy $\pi \triangleq \{\pi_k\}_{k=0}^{\infty}$ as a sequence of mappings from the state to an action. According to the above, we can acquire different values of cost function under different policies π. Therefore, the objective of the MDP is to obtain an optimal policy π^* to minimize the expected average cost over an infinite time horizon, that is,

$$J(s,\pi) = \limsup_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\mathscr{C}(s_k, a_k)], \tag{15}$$

$$J(s,\pi^*) = \min_{\pi\in\Pi} J(s,\pi), \tag{16}$$

where s is the initial state, and the action is $a_k = \pi_k(s_k)$.

IV. Optimal Scheduling For Auxiliary Transmission Scheme

A. Existence of Optimal Deterministic and Stationary Policy

According to the established MDP model, whether there is an optimal strategy is a primary problem in this subsection. To this end, the following assumption is required:

Assumption 1: The spectral radius of each system and the packet dropout probabilities satisfy

$$\max_i \rho^2(A_i)\lambda_i\lambda^* < 1. \tag{17}$$

For ease of presentation, we assume $N = 2$ in the following analysis, i.e., there are only two systems. Without loss of generality, let the packet dropout probabilities satisfy $\lambda_1 < \lambda_2$. The existence of an optimal deterministic stationary policy is given as follows.

Theorem 1: If Assumption 1 holds, there exists a constant ρ^* and a function $V(\cdot)$ to solve the average cost optimality (Bellman) equation, and ρ^* is the optimal value for the problem (15), that is,

$$\rho^* + V(s) = \min_{a\in\mathscr{A}} \{\mathscr{C}(s,a) + \sum_{s'\in\mathscr{S}} \mathscr{P}(s'|s,a)V(s')\}. \tag{18}$$

where s' stands for the next state of the systems. In addition, there exists an optimal deterministic stationary policy π^* to solve the Bellman optimality equation (18), i.e.,

$$a^* = \pi^*(s) = \arg\min_{a\in\mathscr{A}} \{\mathscr{C}(s,a) + \sum_{s'\in\mathscr{S}} \mathscr{P}(s'|s,a)V(s')\}. \tag{19}$$

Proof: First, we denote the discounted total cost under a policy as:

$$V_\sigma(s) \triangleq \sum_{k=0}^{\infty} \sigma^k \mathbb{E}[\mathscr{C}(s_k, \pi(s_k))|s_0 = s], \tag{20}$$

$$V_\sigma^*(s) = \inf_{\pi \in \Pi} V_\sigma(s). \tag{21}$$

Then, the action space $\mathscr{A}$ is finite. According to the established MDP model and Abelian theorem [25], we have

$$\begin{aligned}&\liminf_{T \longrightarrow \infty} \frac{1}{T+1} \sum_{k=0}^{T} \mathscr{C}_k \leq \liminf_{\sigma \uparrow 1} (1-\sigma) \sum_{k=0}^{\infty} \sigma^k \mathscr{C}_k \\ &\leq \limsup_{\sigma \uparrow 1} (1-\sigma) \sum_{k=0}^{\infty} \sigma^k \mathscr{C}_k \leq \limsup_{T \longrightarrow \infty} \frac{1}{T+1} \sum_{k=0}^{T} \mathscr{C}_k.\end{aligned}$$

The one-stage cost $\mathscr{C}_k$ is nonnegative, continuous, and the set $\{a \in \mathbb{A} | \mathscr{C}_k(s,a) < \theta\}$ is compact for any $\theta \in \mathbb{R}$. In the process of state transition, the corresponding probability $\mathscr{P}(s_{k+1}|s_k, a_k)$ is strongly continuous in a_k. Since the average cost is bounded, if we always select the action $a = (0,1)$, there exists a state $b \in \mathscr{S}$ to satisfy $(1-\sigma)V_\sigma^*(b) \leq \overline{M}$, where $\sigma \in [\underline{\sigma}, 1)$, $0 < \underline{\sigma} < 1$ and $\overline{M}$ is a nonnegative number. Besides, we pick the state $z = (0,0)$ and let $N = \inf_k \geq 0\{k : \tau_i = 0, \tau_j = 0\}$. Then, we can obtain:

$$\begin{aligned}V_\sigma^*(s) &\leq \mathbb{E}\{\sum_{k \geq 0}^{N-1} \sigma^k \mathscr{C}_k(s_k, a_k)|s_0 = s\} + \mathbb{E}\{\sigma^N | s_0 = s\} V_\sigma^*(z) \\ &\leq \mathbb{E}\{\sum_{k \geq 0}^{N-1} \mathscr{C}_k(s_k, a_k)|s_0 = s\} + V_\sigma^*(z),\end{aligned}$$

with $s \neq z$. If Assumption 1 holds, $\mathbb{E}\{\sum_{k\geq 0}^{N-1} \mathscr{C}_k(s_k,a_k)|s_0 = s\}$ is always bounded for any state s. However, we can set $f(s) = \mathbb{E}\{\sum_{k\geq 0}^{N-1} \mathscr{C}_k(s_k,a_k)|s_0 = s\}$ to satisfy $-\underline{M} \leq V_\sigma^*(s) - V_\sigma^*(z) \leq f(s)$ in the above inequality, where $\underline{M}$ is a constant and $f(s)$ is a nonnegative function. Based on finite states and discrete actions, the function $f(s)$ is measurable, and $\{V_{\sigma(n)}^*(s) - V_{\sigma(n)}^*(z)\}$ is equicontinuous. In summary, the above conditions have been verified. To find an optimal deterministic stationary policy, consider the vanishing discount approach in [25], and it suffices to satisfy

$$\limsup_{T \to \infty} \frac{1}{T+1} \sum_{k=0}^{T} \mathbb{E}[\mathscr{C}(s_k, a_k)] < \infty. \tag{22}$$

Because the energy cost of the auxiliary channel is bounded, and the action space is finite, we have $\mathbb{E}[\mathrm{Tr}(P_{i,k}) + \beta u_{i,k} E] < \infty$ for $k \in \mathbb{N}$, which completes the proof. ■

B. Structural Results

The above results show that the optimal policy is deterministic and stationary, which greatly reduces the search space of the optimal policy. Nevertheless, due to the countable state space, solving the Bellman equation is still an intractable issue. In this section, a structural result of the optimal policy is obtained to reduce the computational complexity, which is essential for the MDP. We first give the following definition:

Definition 1: (Submodularity). Define a function $f(\cdot,\cdot) : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ as a submodular function, which satisfies

$$f(x^+, y^+) + f(x^-, y^-) \leq f(x^+, y^-) + f(x^-, y^+), \tag{23}$$

where $x^+ \geq x^-$ and $y^+ \geq y^-$.

Theorem 2: Given the stopping time of each system $\tau_j^f, j \neq i$, the corresponding optimal action $u^*(\tau_i)$ of the i-th system is non-decreasing in τ_i.

Proof: We define the function $Q(\cdot,\cdot)$ as

$$Q(\tau_i, u_i) \triangleq \mathscr{C}(s_{(i)}, a_{(i)}) + \sum_{s'_{(i)} \in \mathscr{S}} \mathbb{P}(s'_{(i)}|\tau_i, u_i) V(s'_{(i)}), \tag{24}$$

where $s_{(i)} \triangleq (\tau_1^f, \ldots, \tau_i, \ldots, \tau_N^f)$, $a_{(i)} \triangleq (0, \ldots, u_i, \ldots, 0)$ and the next state is $s'_{(i)} \triangleq (\tau_1^f, \ldots, \tau'_i, \ldots, \tau_N^f)$ with fixed $\tau_j^f, j \neq i$, and $\tau'_i = 0$ or $\tau_i + 1$. Based on Theorem 1 in [19], given fixed $\tau_j^f, j \neq i$, to prove that $u^*(\tau_i)$ is non-decreasing in τ_i, it is sufficient to show that $Q(\tau_i, u_i)$ is submodular in (τ_i, u_i), i.e., $Q(\tau_i, u_i)$ satisfies the inequality

$$Q(\tau_i^+, u_i^+) - Q(\tau_i^-, u_i^+) \leq Q(\tau_i^+, u_i^-) - Q(\tau_i^-, u_i^-), \tag{25}$$

where $\tau_i^- \leq \tau_i^+$ and $u_i^- \leq u_i^+$. According to the equations (24) and (25), we have

$$\begin{aligned}&Q(\tau_i^+, u_i^+) - Q(\tau_i^-, u_i^+) \\ =&\lambda_1 \lambda^* \mathrm{Tr}(h_i^{\tau_i^+ + 1}(\overline{P}_i)) + (1 - \lambda_1 \lambda^*) \mathrm{Tr}(\overline{P}_i) \\ &- \lambda_1 \lambda^* \mathrm{Tr}(h_i^{\tau_i^- + 1}(\overline{P}_i)) - (1 - \lambda_1 \lambda^*) \mathrm{Tr}(\overline{P}_i) \\ &+ \mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^+] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^+] \\ =&\lambda_1 \lambda^* (\mathrm{Tr}(h_i^{\tau_i^+ + 1}(\overline{P}_i)) - \mathrm{Tr}(h_i^{\tau_i^- + 1}(\overline{P}_i))) \\ &+ \mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^+] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^+],\end{aligned}$$

and

$$\begin{aligned}&Q(\tau_i^+, u_i^-) - Q(\tau_i^-, u_i^-) \\ =&\lambda_1 \mathrm{Tr}(h_i^{\tau_i^+ + 1}(\overline{P}_i)) + (1 - \lambda_1) \mathrm{Tr}(\overline{P}_i) \\ &- \lambda_1 \mathrm{Tr}(h_i^{\tau_i^- + 1}(\overline{P}_i)) - (1 - \lambda_1) \mathrm{Tr}(\overline{P}_i) \\ &+ \mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^-] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^-] \\ =&\lambda_1 (\mathrm{Tr}(h_i^{\tau_i^+ + 1}(\overline{P}_i)) - \mathrm{Tr}(h_i^{\tau_i^- + 1}(\overline{P}_i))) \\ &+ \mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^-] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^-].\end{aligned}$$

Obviously, we have $\lambda_1 \lambda^* (\mathrm{Tr}\{h_i^{\tau_i^+ + 1}(\overline{P}_i)\} - \mathrm{Tr}\{h_i^{\tau_i^- + 1}(\overline{P}_i)\}) \leq \lambda_1 (\mathrm{Tr}\{h_i^{\tau_i^+ + 1}(\overline{P}_i)\} - \mathrm{Tr}\{h_i^{\tau_i^- + 1}(\overline{P}_i)\})$. If the formula (25) is true, then one only has to guarantee that $\mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^+] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^+] \leq \mathbb{E}[V(\tau_i^{+'})|\tau_i^+, u_i^-] - \mathbb{E}[V(\tau_i^{-'})|\tau_i^-, u_i^-]$.

Therefore, we need to compute $V(s)$ by the relative value iteration (Bellman equation), that is,

$$V_{n+1}(s) = \min_{a \in \mathscr{A}} \{\mathscr{C}(s,a) + \sum_{s' \in \mathscr{S}} \mathscr{P}(s'|s,a) V_n(s')\} - V_n(s_\alpha),$$

where s_α is an arbitrary state, $s_\alpha \neq s$. In the following, we define $Q_n(\tau_i, u_i) = \mathscr{C}(s_{(i)}, a_{(i)}) + \sum\limits_{s'_{(i)} \in \mathscr{S}} \mathscr{P}(s'_{(i)}|s_{(i)}, a_{(i)}) V_n(s'_{(i)})$.

Without loss of generality, assume $N = 2$. We fix τ_2 to describe the state $s^- \triangleq (\tau_1^-, \tau_2)$, $s^+ \triangleq (\tau_1^+, \tau_2)$, and denote the action space as $a \in \mathscr{A} \triangleq \{(0,0),(1,0),(0,1)\}$, where $a_0 = (0,0)$ indicates that the auxiliary channel is idle, and $a_1 = (1,0)$, $a_2 = (0,1)$ indicate that the sensor 1 and 2 adopt the auxiliary channel to transmit data, respectively. Therefore, the value function of states $V(s)$ is

$$\begin{aligned}
&V_1(s^-)\\
=&\min_{a\in\mathscr{A}}\{\mathscr{C}(s^-,a)+\sum_{s^{-'}\in\mathscr{S}}\mathscr{P}(s^{-'}|s^-,a)V_0(s^{-'})\}-V_0(s_\alpha)\\
=&\min_{u_1\in\{0,1\}}\{Q_0(\tau_1^-,u_1)\}-V_0(s_\alpha)\\
\leq&\min_{u_1\in\{0,1\}}\{Q_0(\tau_1^+,u_1)\}-V_0(s_\alpha)=V_1(s^+),
\end{aligned}$$

and the action-value function $Q_0(\tau_1^-, u_1)$ is

$$\begin{aligned}
&Q_0(\tau_1^-,u_1)=Q_0(\tau_1^-,0)\\
=&\bar{\lambda}_1\mathrm{Tr}(\overline{P}_1)+\lambda_1\mathrm{Tr}(h_i^{\tau_1^-+1}(\overline{P}_1))+\bar{\lambda}_1\mathrm{Tr}(\overline{P}_2)+\lambda_1\mathrm{Tr}(h_i^{\tau_2+1}(\overline{P}_2))\\
\leq&\bar{\lambda}_1\mathrm{Tr}(\overline{P}_1)+\lambda_1\mathrm{Tr}(h_i^{\tau_1^++1}(\overline{P}_1))+\bar{\lambda}_1\mathrm{Tr}(\overline{P}_2)+\lambda_1\mathrm{Tr}(h_i^{\tau_2+1}(\overline{P}_2))\\
=&Q_0(\tau_1^+,0)=Q_0(\tau^+,u_1),
\end{aligned}$$

and

$$\begin{aligned}
&Q_0(\tau_1^-,u_1)=Q_0(\tau_1^-,1)\\
=&\bar{\lambda}_1\mathrm{Tr}(\overline{P}_1)+\lambda_1\mathrm{Tr}(h_i^{\tau_1^-+1}(\overline{P}_1))\\
&+(1-\lambda_1\lambda^*)\mathrm{Tr}(\overline{P}_2)+\lambda_1\lambda^*\mathrm{Tr}(h_i^{\tau_2+1}(\overline{P}_2))\\
\leq&\bar{\lambda}_1\mathrm{Tr}(\overline{P}_1)+\lambda_1\mathrm{Tr}(h_i^{\tau_1^++1}(\overline{P}_1))\\
&+(1-\lambda_1\lambda^*)\mathrm{Tr}(\overline{P}_2)+\lambda_1\lambda^*\mathrm{Tr}(h_i^{\tau_2+1}(\overline{P}_2))\\
=&Q_0(\tau_1^+,1)=Q_0(\tau^+,u_1),
\end{aligned}$$

where $\bar{\lambda}_1 = (1-\lambda_1)$. Assume that $V_n(s^-) \leq V_n(s^+)$. This inequality still holds according to $Q_{n-1}(\tau_1^-,0) = (1-\lambda_1)\mathrm{Tr}\{\overline{P}_1\} + \lambda_1\mathrm{Tr}\{h_i^{\tau_1^-+1}(\overline{P}_1)\} + (1-\lambda_1)\mathrm{Tr}\{\overline{P}_2\} + \lambda_1\mathrm{Tr}\{h_i^{\tau_2+1}(\overline{P}_2)\} \leq (1-\lambda_1)\mathrm{Tr}\{\overline{P}_1\} + \lambda_1\mathrm{Tr}\{h_i^{\tau_1^++1}(\overline{P}_1)\} + (1-\lambda_1)\mathrm{Tr}\{\overline{P}_2\} + \lambda_1\mathrm{Tr}\{h_i^{\tau_2+1}(\overline{P}_2)\} = Q_{n-1}(\tau_1^+,0)$, and $Q_{n-1}(\tau_1^-,1) \leq Q_{n-1}(\tau_1^+,1)$. The next value function of state satisfies

$$\begin{aligned}
&V_{n+1}(s^-)\\
=&\min_{a\in\mathscr{A}}\{\mathscr{C}(s^-,a)+\sum_{s^{-'}\in\mathscr{S}}\mathscr{P}(s^{-'}|s^-,a)V_n(s^{-'})\}-V_n(s_\alpha)\\
=&\min_{u\in\{0,1\}}\{Q_n(\tau_1^-,u_1)\}-V_n(s_\alpha)\\
\leq&\min_{u\in\{0,1\}}\{Q_n(\tau_1^+,u_1)\}-V_n(s_\alpha)=V_{n+1}(s^+),
\end{aligned}$$

which proves the monotonicity of the state value function $V_n(s)$, it is shown that $Q(\tau_i,u_i)$ is submodular function. Hence, $u^*(\tau_i) = arg\min_{u_i} Q(\tau_i,u_i)$ is non-decreasing in τ_i. The proof is completed. ■

Remark 1: Utilizing the threshold structure, the optimal policy can be efficiently determined through the relative value iteration algorithm (RVIA) with slight adjustments. To elaborate, for a given state s, if the optimal action is a^* for a state $s^* \leq s$, it suffices to apply RVIA solely to actions $a \geq a^*$. This approach minimizes superfluous computations, thus reducing computational complexity.

Remark 2: The assumption that $N = 2$ can be extended to a general N. For example, when $N = 3$, the state becomes $s = (\tau_1,\tau_2,\tau_3)$, and the action is $a \in \{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$. Building upon the evidence presented in the proof of Theorem 2, it is evident that the Q function within the context of relative value iteration continues to exhibit submodularity. Consequently, the threshold structure also persists.

V. Illustrative Example

This section illustrates the threshold structure of the optimal policy by an example. In addition, an optimal policy is approximated by the deep reinforcement learning in the simulation environment. Adopt two LTI systems with the following parameters:

$$A_1 = \begin{bmatrix}1.12 & 0.8\\ 0 & 0.1\end{bmatrix},\ Q_1 = \begin{bmatrix}1 & 0\\ 0 & 0.5\end{bmatrix},\ C_1 = [1,0.9],\ R_1 = 1.2,$$

$$A_2 = \begin{bmatrix}1.09 & 0.7\\ 0 & 0.2\end{bmatrix},\ Q_2 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ C_2 = [1,1],\ R_2 = 1.$$

The above parameters are selected by experience. Two smart sensors obtain their local state estimates, and transmit them to a remote estimator over unreliable channels. While practical applications typically involve a greater number of sensors than just two, we employ a pair of sensors to visually elucidate the threshold structure of the optimal policy. The packet arrival probabilities of the channels are $\lambda_1 = 0.4$ and $\lambda_2 = 0.5$, respectively. To improve the estimation performance, an auxiliary channel with high reliability is deployed to transmit a duplication when necessary, whose packet arrival probability is assumed to be $\lambda_3 = 0.7$. The energy cost of using the auxiliary channel is $E = 30$. The optimal scheduling policy for this auxiliary channel is shown in Fig. 2. One finds that when the holding times of two systems are both small, the auxiliary channel is unnecessary to be activated for energy conservation.

Finally, the deep reinforcement learning algorithm, i.e., dueling double deep Q-network (D3QN) is employed to approximate an optimal policy. Please refer to [26] for the details of the algorithm. The parameters are given as follows. The maximum capacity of the replay buffer is 10^6, and the numbers of nodes in the neural network are 128, 128, and 64, respectively. The target network is synchronized with an online network every 100 steps. The learning rate is 0.0003. The discount factor $\sigma = 0.99$. We train the network over 100 episodes, and each of them has 100 time steps. The learning process and the comparison with other policies are depicted in Fig. 3, where the greedy policy selects the sensor with the largest holding time to use the auxiliary channel. It is evident that the D3QN significantly reduces the average cost, i.e., the estimation error covariance and the energy consumption.

VI. Conclusion

This paper has proposed an auxiliary transmission scheme to improve the remote estimation performance. To implement

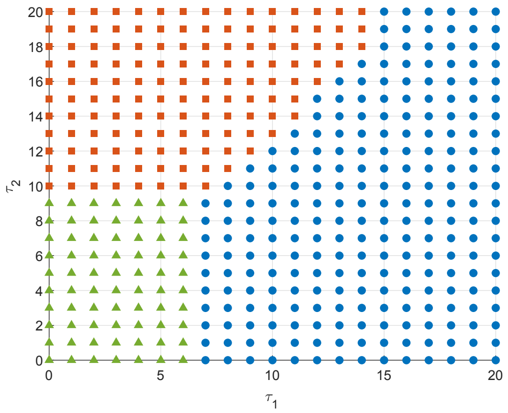

Fig. 2. The optimal scheduling policy for the auxiliary channel. Blue circles, red squares and green triangles correspond to the action $(1,0)$, $(0,1)$ and $(0,0)$, respectively.

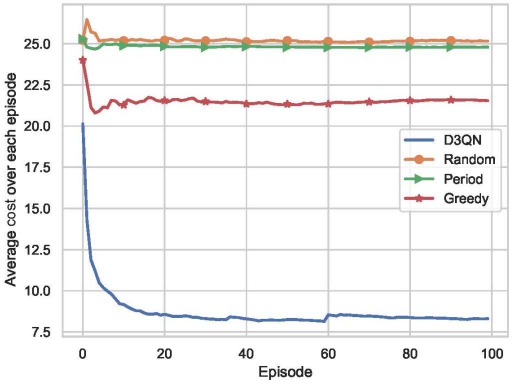

Fig. 3. The learning process of the optimal scheduling policy.

the optimal scheduling for the auxiliary channel, an MDP model has been established, and the existence of an optimal deterministic and stationary policy has also been presented. Besides, the optimal policy has been verified to have a threshold structure. The D3QN algorithm has been employed to obtain an optimal policy for the optimal scheduling. Finally, the simulation results have visually shown the threshold structure of the optimal policy, and the better performance of the DRL-based policy.

References

[1] M. Klügel, M. Mamduhi, O. Ayan, M. Vilgelm, K. H. Johansson, S. Hirche, and W. Kellerer, "Joint cross-layer optimization in real-time networked control systems," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 4, pp. 1903–1915, 2020.

[2] E. Mousavinejad, F. Yang, Q.-L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE transactions on cybernetics*, vol. 48, no. 11, pp. 3254–3264, 2018.

[3] B. Tavassoli, "Stability of nonlinear networked control systems over multiple communication links with asynchronous sampling," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 511–515, 2013.

[4] Z. Wang, J. Sun, and Y. Bai, "Stability analysis of event-triggered networked control systems with time-varying delay and packet loss," *Journal of Systems Science and Complexity*, vol. 34, no. 1, pp. 265–280, 2021.

[5] Y. Ni, Z. Guo, Y. Mo, and L. Shi, "On the performance analysis of reset attack in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 1, pp. 419–425, 2019.

[6] M.-C. Chiu, C.-D. Tsai, and T.-L. Li, "An integrative machine learning method to improve fault detection and productivity performance in a cyber-physical system," *Journal of Computing and Information Science in Engineering*, vol. 20, no. 2, p. 021009, 2020.

[7] A. Emadi, S. S. Williamson, and A. Khaligh, "Power electronics intensive solutions for advanced electric, hybrid electric, and fuel cell vehicular power systems," *IEEE Transactions on Power Electronics*, vol. 21, no. 3, pp. 567–577, 2006.

[8] C. E. Chronaki, D. G. Katehakis, X. C. Zabulis, M. Tsiknakis, and S. C. Orphanoudakis, "Weboncoll: medical collaboration in regional healthcare networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 1, no. 4, pp. 257–269, 1997.

[9] H. Yang, M. Huang, Y. Li, S. Dey, and L. Shi, "Joint power allocation for remote state estimation with swipt," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1434–1447, 2022.

[10] T. Iwaki, J. Wu, Y. Wu, H. Sandberg, and K. H. Johansson, "Multi-hop sensor network scheduling for optimal remote estimation," *Automatica*, vol. 127, p. 109498, 2021.

[11] L. Yang, Y. Xu, H. Rao, Y.-H. Liu, and C.-Y. Su, "Efficient measurement scheduling for remote state estimation with limited bandwidth," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[12] X. Hu, X. Zhan, J. Wu, and H. Yan, "Performance limits of network delay systems based on bandwidth and packet loss constraints," *Asian Journal of Control*, vol. 24, no. 6, pp. 3466–3474, 2022.

[13] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.

[14] L. Mo, X. Cao, Y. Song, and A. Kritikakou, "Distributed node coordination for real-time energy-constrained control in wireless sensor and actuator networks," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4151–4163, 2018.

[15] E. Kung, J. Wang, J. Wu, D. Shi, and L. Shi, "On the nonexistence of event triggers that preserve gaussian state in presence of packet-drop," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4302–4307, 2019.

[16] J. Ding, S. Sun, J. Ma, and N. Li, "Fusion estimation for multi-sensor networked systems with packet loss compensation," *Information Fusion*, vol. 45, pp. 138–149, 2019.

[17] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 1041–1046, 2012.

[18] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2661–2675, 2015.

[19] S. Wu, X. Ren, Q.-S. Jia, K. H. Johansson, and L. Shi, "Learning optimal scheduling policy for remote state estimation under uncertain channel condition," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 2, pp. 579–591, 2019.

[20] S. Wu, K. Ding, P. Cheng, and L. Shi, "Optimal scheduling of multiple sensors over lossy and bandwidth limited channels," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 3, pp. 1188–1200, 2020.

[21] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 119–125, 2018.

[22] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements," *IEEE Network*, vol. 32, no. 2, pp. 8–15, 2018.

[23] P. U. Adamu, M. Lopez-Benitez, and J. Zhang, "Hybrid transmission scheme for improving link reliability in mmwave urllc communications," *IEEE Transactions on Wireless Communications*, 2023.

[24] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[25] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov control processes: basic optimality criteria*, vol. 30. Springer Science & Business Media, 2012.

[26] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International Conference on Machine Learning*, pp. 1995–2003, PMLR, 2016.