ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

The multi-action transmission scheduling problem for remote state estimation under global energy constraint

Xu Fanghui

School of Mathematics and Statistics, Guangxi Normal University, Guilin, China
fanghuixu@stu.gxnu.edu.cn

Abstract: *This paper proposes a multi-action transmission scheduling problem for remote state estimation in wireless networked physical systems under global energy constraints. Firstly, the system state estimation is obtained through Kalman filtering. Then, the estimated values are transmitted from sensors to a remote estimator through randomly fading channel. Unlike traditional transmission scheduling problems, the proposed system model allows sensors to choose from multiple power levels at each decision node. Additionally, the study considers global energy constraint and formulates the system model as a constrained Markov decision process to obtain the optimal transmission strategy that minimizes the average estimation error covariance at the remote estimator over an infinite time horizon. Through model transformation, a important conclusion is derived. As the average estimation error covariance at the remote estimator increases, the optimal transmission power exhibits an increasing trend. This conclusion extends the threshold structure of the optimal strategy to a monotonic increasing structure. Finally, the theoretical result is verified through numerical simulations.*

Keywords: *Optimal transmission scheduling; Estimation error covariance (EEC); Constrained Markov decision process (CMDP)*

1. Introduction

In recent years, wireless communication has completely changed the way information is exchanged. We are able to use smartphones, tablets, and other devices to effortlessly access the internet, make phone calls, send messages, and remotely control devices. Therefore, research on transmission scheduling in wireless communication is of great significance. Efficient transmission scheduling can ensure balanced resource allocation and enhance the service quality and reliability of wireless communication systems by minimizing wait times and optimizing performance objectives. To achieve these communication goals, many scholars have extensively researched optimal transmission scheduling. Zhang et al. [1-3] studied the optimal transmission strategies for sensors, but overlooked the energy consumption that may occur due to the mobility of communication devices; Yuan et al. [4-6] investigated energy allocation problems in time-varying channels within a limited time period; Knorn S, Leong A S [7-8] incorporated energy harvesting devices; Qi Y, Wei J et al. [9-12] obtained threshold strategies for the estimation error covariance (EEC) within a finite time range; Leong A S et al. [13] further derived a dual-threshold strategy for the optimal transmission scheduling regarding EEC and channel state over an infinite time span. In addition, Salh A [14] considered the existence of eavesdropping, aiming to minimize the EEC at the remote estimator while keeping the EEC at the eavesdropping location above a certain level.

However, these references only considered transmission scheduling problems with two actions, where sensors only decide whether to transmit data, limiting the flexibility of the system and overlooking more diverse optimal solutions. In practice, transmitters may have multiple transmission power levels; based on this, this study considers transmission scheduling problems with multiple actions. Furthermore, this study also takes into account global energy constraints, formulating the transmission scheduling problem as a constrained Markov decision process (CMDP) problem. A fundamental method to solve CMDP problems is the method of Lagrange multipliers, which can transform the CMDP problem into an unconstrained Markov decision problem and then solve it by finding the optimal Lagrange multipliers and MDP policy [15-17]. In order to find the optimal Lagrange multipliers, the gradient descent (GD) algorithm was proposed in [18-19]; however, the computational complexity of this algorithm is too high for the structural nature of the optimal strategy for the transmission scheduling model in this study. Therefore, this study proposes a new algorithm to compute the optimal Lagrange multipliers, significantly reducing the number of iterations and computational complexity.

2. Model And Problem

2.1. System Model

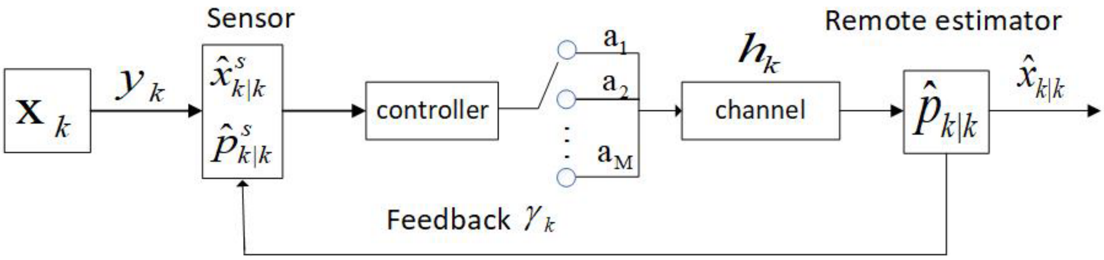

Fig. 1: Remote state estimation flow chart

As shown in Figure 1, we consider a discrete time process formulated as follows.

$$\begin{cases} x_{k+1} = Ax_k + \omega_k \\ y_k = Cx_k + v_k \end{cases}, \tag{1}$$

where $x_k \in R^{nx}$ denotes the state of the system at time k , and ω_k is Gaussian random vector with zero-mean and variance $Q \geq 0$. In this system, the sensor performs the measurements, and $y_k \in R^{ny}$ denotes the system measurement at moment k , v_k is subject to a Gaussian distribution with zero-mean and variance $U > 0$. We further assume that the two noises $\{\omega_k\}$ and $\{v_k\}$ are independent of each other. Assuming that A is unstable, (A,C) is detectable and $(A,Q^{1/2})$ is controllable [20].

1) Local Estimation: The sensor is equipped with Kalman filtering capabilities, allowing for the calculation of the local state estimate and EEC directly on the sensor using the following standard Kalman filtering equations:

$$\begin{cases} \hat{x}^s_{k|k-1} \triangleq \mathbb{E}[x_k \mid y_0,\cdots,y_{k-1}]; \hat{x}^s_{k|k} \triangleq \mathbb{E}[x_k \mid y_0,\cdots y_k] \\ \hat{p}^s_{k|k-1} \triangleq \mathbb{E}[(x_k - \hat{x}^s_{k|k-1})(x_k - \hat{x}^s_{k|k-1})^T \mid y_0,\cdots,y_{k-1}] \\ \hat{p}^s_{k|k} \triangleq \mathbb{E}[(x_k - \hat{x}^s_{k|k})(x_k - \hat{x}^s_{k|k})^T \mid y_0,\cdots,y_k] \end{cases}. \tag{2}$$

Under the detectability assumption, the steady state value of $\hat{p}^s_{k|k}$ exists, and we denote $\overline{p} = \lim_{k\to\infty} \hat{p}^s_{k|k}$ in this article.

2) Markov Channel: We consider a flat-fading channel, and adopt a discrete-time Markov chain with finite state space $\mathcal{H}$ to model the fading channel. The state transition probability matrix is denoted by $P = \left[\mathrm{p}\left(\mathrm{h}' \mid \mathrm{h}\right)\right]$, where $p(h' \mid h) \triangleq p(h_{k+1} = h' \mid h_k = h)$ denotes the transition probability from state h at time k to state h' at time $k+1$ for $\forall h, h' \in \mathcal{H}$. For the channel state $h_k \in \mathcal{H}$ at time k , its corresponding channel gain is denoted by g_k , i.e., $g_k \triangleq g(h_k)$, $k = 0,1,2,\cdots$.

3) Data transmission: Let $\gamma_k = 1$ denote successful detection of sensor's packet at time k ,and $\gamma_k = 0$ denote that the packet is failed to be transmitted. We also assume that each packet has the same

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

size which is denoted by R. Let π_k be the transmission power of the sensor at time k, and we assume that π_k is able to take M values, which are given in

$$\mathcal{A} = \{a_1, a_2, \ldots, a_M\}. \tag{3}$$

Given the channel state $h \in \mathcal{H}$, and the transmission power $a \in \mathcal{A}$ is chosen, then the probability that the remote estimator can successfully receive the packet from the sensor is denoted by

$$\tilde{p}(h,a) \triangleq p(r(h,a) \geq R), \tag{4}$$

where $r(h,a)$ means the channel capacity, and it can be obtained by

$$r(h,a) = B\log_2\left[1 + \frac{g(h)a}{\Gamma N_0 B}\right], \tag{5}$$

in which B represents the channel bandwidth, Γ denotes the signal-noise ratio (SNR) gap, and N_0 is the power spectral density of the Gaussian noise.

4) Remote Estimation: The optimal remote estimator, in terms of minimum mean-square error(MMSE), is defined as follows:

$$\hat{x}_{k|k} = \begin{cases} A\hat{x}_{k-1|k-1} & \gamma_k = 0 \\ \hat{x}^s_{k|k} & \gamma_k = 1 \end{cases}, \quad \hat{p}_{k|k} = \begin{cases} f(\hat{p}_{k-1|k-1}) & \gamma_k = 0 \\ \overline{p} & \gamma_k = 1 \end{cases}, \tag{6}$$

where $f(X) \triangleq AXA^T + Q$. Assuming that γ_k will be feed back to the sensor by the remote estimator before the start of next time $k+1$, then $\hat{P}_{k|k}$ will be update at the sensor. To simplify the notation, in the following $\hat{P}_{k|k}$ is denoted as p_k. Let $\mathcal{P}$ be the state space of all possible values of p_k at the remote estimator.In this article, we assume that $\mathcal{P}$ is a finite state set. Obviously, this assumption is reasonable, since in practical application, after a finite number of transmissions, the packet can be received by the remote estimator eventually. Thus, $\mathcal{P}$ can be denoted as

$$\mathcal{P} \triangleq \left\{\overline{p}, f(\overline{p}), f^2(\overline{p}), \ldots, f^{\kappa}(\overline{p})\right\}, \tag{7}$$

where $\kappa < \infty$ is a given large enough positive integer, and we define

$$f^k(X) = \begin{cases} f(f^{k-1}(X)) & k > 0 \\ X & k = 0 \end{cases}. \tag{8}$$

The elements in $\mathcal{P}$ have the following size relationship [20]

$$\overline{p} \leq f(\overline{p}) \leq f^2(\overline{p}) \leq \cdots \leq f^{\kappa}(\overline{p}), \tag{9}$$

and this inequality also indicates

$$Tr(\overline{p}) \leq Tr(f(\overline{p})) \leq Tr(f^2(\overline{p})) \leq \cdots \leq Tr(f^{\kappa}(\overline{p})), \tag{10}$$

where $Tr(f^n(\overline{p}))$ denotes the trace of the matrix $f^n(\overline{p})$ for $n = 0,1,2,\cdots,\kappa$.

2.2. Problem formulation

In the following, a new transmission scheduling problem, i.e., multi-action transmission scheduling problem under global energy constraint, is developed, and then it is transformed into a CMDP.The more detail is given as follows.

1) State space: Let $s_k = [h_k, p_k]$ be the state of CMDP at time k, $k = 0,1,2,\cdots$, and the state

space is denote by $\mathcal{S} \triangleq \mathcal{H} \times \mathcal{P}$.

2) Action space: Let A be the action space given in (2). Furthermore, we define

$$\pi \triangleq \{\pi_k, k = 0,1,2,\ldots\} \tag{11}$$

as the transmission scheduling policy, where $\pi_k \triangleq \pi(s_k) \in \mathcal{A}$ denotes the rule of action at time k when the state is s_k. Also, define Φ as the set of all π and $\Phi_D \subset \Phi$ as the set of all deterministic policies.

3) Transition probabilities: The transition probability from state $s = [h, p]$ to $s' = [h', p']$ when taking action $a \in \mathcal{A}$ is defined as

$$\begin{aligned} &p(s' \mid s,a) = p([h',p'] \mid [h,p],a) \triangleq p(h' \mid h) \\ &\{\tilde{p}(h',a) I_{\{p'=\overline{p}\}} + (1-\tilde{p}(h',a)) I_{\{p'=f(p)\}}\} \end{aligned}, \tag{12}$$

where $I_{\{\cdot\}}$ is an indicator function, and $\tilde{p}(\cdot,\cdot)$ is defined in (4).

4) Cost function: For $\forall s = [h, p] \in \mathcal{S}, a \in \mathcal{A}$, define

$$t(s,a) \triangleq Tr(p), c(s,a) \triangleq R \cdot a\ , \tag{13}$$

where $t(s,a)$ denotes the trace of the EEC matrix p , and $c(s,a)$ is the energy consumption of the sensor during each data packet transmission when the transmission power a is taken. We call $t(s,a)$ and $c(s,a)$ as instantaneous target cost and instantaneous constrained cost, respectively.

5) Average target cost function: For $\forall s_0 \in \mathcal{S}$, $\forall\ \pi \in \Phi$, define average target cost function as

$$T^{s_0}(\pi) \triangleq \mathbb{E}_\pi^{s_0}\Big[\limsup_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} t(s_k,\pi_k)\Big]. \tag{14}$$

6) Average constrained cost function: Similarly, the average constrained cost is

$$C^{s_0}(\pi) = \mathbb{E}_\pi^{s_0}\Big[\limsup_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} c(s_k,\pi_k)\Big]. \tag{15}$$

One of our primary objectives is to determine the optimal transmission scheduling policy $\pi^* \in \Phi$ that minimizes the target cost function (14) while the average constrained power consumption (15) is not more than some given threshold value $\tilde{c}$. That is, we mainly investigate the following optimal transmission scheduling problem: finding an optimal $\pi^* \in \Phi$, such that

$$T^{s_0}(\pi^*) = \inf_{\pi\in\Phi}\{T^{s_0}(\pi)\}, s.t. C^{s_0}(\pi^*) \le \tilde{c}\ . \tag{16}$$

2.3. Lagrangian cost problem

The average target cost function and the average constrained cost function under the policy $\pi^*(\tilde{c})$ are denoted by $T(\pi^*(\tilde{c}))$ and $C(\pi^*(\tilde{c}))$, respectively. Now, we solve the CMDP model by transforming it into unconstrained MDP with Lagrangian. We first recall the following lemma.

Lemma 2 [21] If the CMDP with only one global constraint is feasible, then there exist $\lambda_1, \lambda_2 \ge 0$, and $0 \le q \le 1$, so that the optimal policy π^* can be obtained by a combination of two optimal deterministic policies $\pi^*(\lambda_1)$ and $\pi^*(\lambda_2)$, in which the probability of $\pi^*(\lambda_1)$ being adopted is q

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

and $\pi^*(\lambda_2)$ being adopted is $1-q$. Furthermore, there exists at most one state that $\pi^*(s;\lambda_1) \neq \pi^*(s;\lambda_2)$.

For the unconstrained average MDP with Lagrangian multiplier λ, the cost function $V(s;\lambda)$ of the optimal deterministic policy $\pi^*(\lambda)$ can be obtained from

$$J(\pi^*(\lambda);\lambda) + V(s;\lambda) = \min_{a\in\mathcal{A}} \{t(s,a;\lambda) + \sum_{s'\in\mathcal{S}} p(s' \mid s,a)V(s';\lambda)\}, \tag{17}$$

where $V(s;\lambda)=0$ for a certain reference state $s\in\mathcal{S}$, and $J(\pi^*(\lambda),\lambda)$ is the optimal average Lagrangian cost. The above equation (17) is also called Bellman equation. The function $V(s;\lambda)$ in (17) can be obtained by Relative Value Iteration (RVI) algorithm, in which a sequence of estimates $\{V^m(s;\lambda)\}_{m=1}^{\infty}$ with relationship

$$V^{m+1}(s;\lambda) = \min_{a\in\mathcal{A}} \{t(s,a;\lambda) + \sum_{s'\in\mathcal{S}} p(s' \mid s,a)V^m(s';\lambda)\}, \tag{18}$$

then we can obtain the value function $V(s;\lambda) = \lim_{m\to\infty}\sup V^m(s;\lambda)$. Besides, the Lagrange cost function is defined when action a is taken in state s as

$$V(s,a;\lambda) = t(s,a;\lambda) + \sum_{s'\in\mathcal{S}} p(s' \mid s,a)V(s';\lambda), \tag{19}$$

then there are

$$V(s;\lambda) = \min_{a\in\mathcal{A}} V(s,a;\lambda); \pi^*(s;\lambda) = \arg\min_a V(s,a;\lambda). \tag{20}$$

3. Structural result of transmission scheduling

We will establish the monotonic structure of scheduling policy. Toward that, we begin by introducing the following definition lemma.

Definition 1(Submodularity)[22] A function f is said to be supermodular in (a,p), if given $h\in\mathcal{H}$, for $\forall a' \geq a$ and $\forall p' \geq p$,it holds that

$$f(a',p';h) - f(a,p';h) \geq f(a',p;h) - f(a,p;h). \tag{21}$$

Lemma 1 [22] If the function f is a submodular function with respect to $(x,\pi(x))$, it follows that $\pi(x)$ is increasing with respect to x.

Lemma 2 Given $h\in\mathcal{H}$, $\lambda>0$, the value function $V([h,p];\lambda)$ is increasing function of $p\in\mathcal{P}$.

Proof. From (18), we only need to prove that $V^m([h,p];\lambda)$ is creasing function with respect to p for all m. It is apparently true for $m=0$, then the increasing property of

$$\begin{aligned} V^{m+1}([h,p];\lambda) = \min_{a\in\mathcal{A}} \Big\{ & t([h,p],a;\lambda) + \sum_{h'\in\mathcal{H}} p(h' \mid h) \\ & [\tilde{p}(h',a)V^m([h',\overline{p}];\lambda) + (1-\tilde{p}(h',a))V^m([h',f(p)];\lambda)] \Big\} \end{aligned} \tag{22}$$

is obtained easily.

According to above discussion, we give the main result in the following.

Theorem 1 Given a feasible transmission cost constraint $\tilde{c}>0$,for $\forall s=[h,p]\in\mathcal{S}$, the optimal policy $\pi^*(s,\tilde{c})$ is monotonic increasing with p.

Proof. Since $f(p') \geq f(p)$ for any $p' \geq p$, we have obtained from Lemma 2 that

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

$$V([h',f(p')];\lambda)-V([h',f(p)];\lambda)\geq 0. \tag{23}$$

Based on (4), for $\forall a'\geq a$, $\tilde{p}(h',a')\geq \tilde{p}(h',a)$, then the following inequality holds:

$$\begin{aligned}&\tilde{p}(h',a')V([h',f(p')];\lambda)-\tilde{p}(h',a')V([h',f(p)];\lambda)\\ &\geq \tilde{p}(h',a)V([h',f(p')];\lambda)-\tilde{p}(h',a)V([h',f(p)];\lambda)\end{aligned}, \tag{24}$$

based on (24), we further get

$$\begin{aligned}&[\tilde{p}(h',a')V([h',\overline{p}];\lambda)+(1-\tilde{p}(h',a'))V([h',f(p')];\lambda)]\\ &-[\tilde{p}(h',a')V([h',\overline{p}];\lambda)+(1-\tilde{p}(h',a'))V([h',f(p)];\lambda)]\\ &\leq [\tilde{p}(h',a)V([h',\overline{p}];\lambda)+(1-\tilde{p}(h',a))V([h',f(p')];\lambda)]\\ &-[\tilde{p}(h',a)V([h',\overline{p}];\lambda)+(1-\tilde{p}(h',a))V([h',f(p)];\lambda)]\end{aligned}. \tag{25}$$

This indicates that

$$\begin{aligned}&p(s^{'}\mid s,a)V(s^{'};\lambda)=p(h'\mid h)\\ &[\tilde{p}(h',a)V([h',\overline{p}];\lambda)+(1-\tilde{p}(h',a))V([h',f(p)];\lambda)]\end{aligned} \tag{26}$$

is a submodular function of (p,a) from definition 1. Noting that the positive weighting of the submodular function remains the submodular function, then we obtain that

$$V_1([h,p],a;\lambda)\triangleq \sum_{s^{'}\in S} p(s^{'}\mid s,a)V(s^{'};\lambda) \tag{27}$$

is also a submodular function of (p,a). It is easy to see that $t([h,p],a;\lambda)$ is a submodular function of (p,a), then

$$V([h,p],a;\lambda)=t([h,p],a;\lambda)+V_1([h,p],a;\lambda) \tag{28}$$

is a submodular function of (p,a) .Therefore,the monotonically increasing structure of the deterministic policies $\pi^{*}(s;\lambda)$ with respect to p is obtained from Lemma 3. This completes the proof.

4. Simulation results

In this section we perform an empirical analysis of the optimal transmission scheduling problem under the framework of the CMDP model.

4.1. Simulation Setup

Consider the following linear time-invariant system with parameters

$$\mathrm{A}=\begin{bmatrix}1.2 & 0.2\\ 0.2 & 0.7\end{bmatrix},\mathrm{C}=\begin{bmatrix}1 & 1\end{bmatrix},Q=I_{2\times 2},U=1.$$

From these parameters we can obtain

$$\overline{p}=\begin{bmatrix}1.3634 & -0.8347\\ -0.8347 & 1.0809\end{bmatrix}.$$

The channel state can be represented by three values: 1, 2, and 3, corresponding to the bad, fair, and good states of the channel, respectively. And the channel gains of state 1,2,3 are $0, 10\times 10^{-13}$ and 15×10^{-13} respectively. The transition probability matrix of the channel fading process is assumed to be

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

$$P = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

Assume that the packet size for each transmission task is $R = 140$ bits, system bandwidth $B = 5$ MHz, noise power spectral density at Remote Estimator $N_0 = 10^{-17}$, and SNR gap $\Gamma = 5.48$, the corresponding symbol error rate is 10^{-4}. We let $K = 10$ in our simulations.

4.2. Numerical analysis

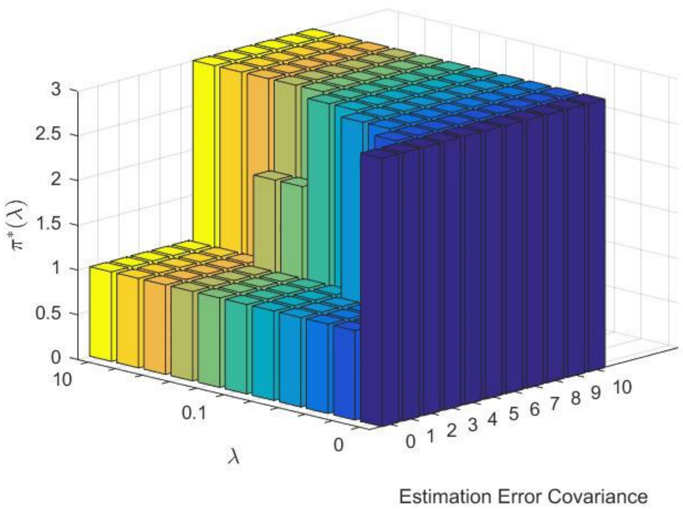

Fig. 2: Optimal policy with respect to EEC when $h = 1$ *and* $\lambda \in [0, 0.2]$

Figure 2 shows the variation of optimal deterministic policy $\pi^*(\lambda)$ with EEC when the channel state is 1 and $\lambda \in [0, 0.2]$. From Figure 2 we can intuitively obtain that, given the channel state 1, for any $\lambda \in [0, 0.2]$, when the EEC increases, the transmission power chosen by the optimal deterministic policy becomes larger, which is consistent with the conclusion of Theorem 1. Another observation is that, given the channel state and λ, the optimal transmission power is piece-wise function of the EEC. That is, when the performance index EEC reaches a certain level, the dynamic system will select the same transmission power to optimize the performance index. Confirming the monotonic property of the optimal deterministic policy.

Now, an average cost constraint $\tilde{c}$ will be arbitrarily given and then the optimal transmission policy will be computed. We take $\tilde{c} = 150$ as an example and set $\epsilon = 10^{-4}$. With the RVI algorithm in [18-19] and for finding the optimal Lagrange multiplier λ^*, we can find it with only 15 explorations in following table 1.

Table1: The process of exploration of λ^*

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
1.00000	0.50000	0.25000	0.12500	0.18750	0.15625	0.14063	0.14844
λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	
0.15234	0.15039	0.15137	0.15186	0.15161	0.15173	0.15167	

After the explorations in the table 1 above, we can get $\lambda^* = 0.15167$, then $\lambda^- = 0.1516$ and $\lambda^+ = 0.1517$. Thus we can obtain two deterministic policies $\pi^*(\lambda^-)$ and $\pi^*(\lambda^+)$. By calculating

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

we can get the weighting factor $q = 0.3571$ and $T(\pi^*(\tilde{c})) = 16.9281$.

5. Conclusions

This article focuses on the transmission scheduling problem for multi-action remote state estimation under a global constraint in CPS. The objective is to determine the optimal power allocation at discrete time intervals to minimize the average EEC at the remote estimator, while ensuring that the average energy consumption remains below a given threshold. We address this problem using the framework of CMDP and derive structural results for transmission scheduling. The theoretical findings are validated through simulations. In our future work, we plan to explore the optimal scheduling problem considering multiple global constraints. We will continue our investigation into the structural results. It is believed that this study has important theoretical and practical significance.

References

[1] Zhang K, Shi Y, Karnouskos S, et al. Advancements in industrial cyber-physical systems: an overview and perspectives[J]. IEEE Transactions on Industrial Informatics, 2022.

[2] Pivoto D G S, de Almeida L F F, da Rosa Righi R, et al. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review[J]. Journal of manufacturing systems, 2021, 58: 176-192.

[3] Xing W, Zhao X, Basar T, et al. Optimal transmission scheduling for remote state estimation in CPSs with energy harvesting two-hop relay networks[J]. Automatica, 2023, 152: 110963.

[4] Yuan F, Tang S, Liu D. AoI-Based Transmission Scheduling for Cyber-Physical Systems Over Fading Channel Against Eavesdropping[J]. IEEE Internet of Things Journal, 2023.

[5] Ho C K, Zhang R. Optimal energy allocation for wireless communications with energy harvesting constraints[J]. IEEE Transactions on Signal Processing, 2012, 60(9): 4808-4818.

[6] Ozel O, Tutuncuoglu K, Yang J, et al. Transmission with energy harvesting nodes in fading wireless channels: Optimal policies[J]. IEEE Journal on selected areas in communications, 2011, 29(8): 1732-1743.

[7] Knorn S, Dey S. Optimal energy allocation for linear control with packet loss under energy harvesting constraints[J]. Automatica, 2017, 77: 259-267.

[8] Leong A S, Dey S, Quevedo D E. Sensor scheduling in variance based event triggered estimation with packet drops[J]. IEEE Transactions on Automatic Control, 2016, 62(4): 1880-1895.

[9] Qi Y, Cheng P, Chen J. Optimal sensor data scheduling for remote estimation over a time-varying channel[J]. IEEE Transactions on Automatic Control, 2016, 62(9): 4611-4617.

[10] Wei J, Ye D. On two sensors scheduling for remote state estimation with a shared memory channel in a cyber-physical system environment[J]. IEEE Transactions on Cybernetics, 2021.

[11] Wang J, Ren X, Dey S, et al. Remote state estimation with usage-dependent Markovian packet losses[J]. Automatica, 2021, 123: 109342.

[12] Wei J, Ye D. Double Threshold Structure of Sensor Scheduling Policy Over a Finite-State Markov Channel[J]. IEEE Transactions on Cybernetics, 2022.

[13] Leong A S, Quevedo D E, Dolz D, et al. Transmission scheduling for remote state estimation over packet dropping links in the presence of an eavesdropper[J]. IEEE Transactions on Automatic Control, 2018, 64(9): 3732-3739.

[14] Salh A, Audah L, Alhartomi M A, et al. Smart packet transmission scheduling in cognitive IoT systems: DDQN based approach[J]. IEEE Access, 2022, 10: 50023-50036.

[15] Djonin D V, Krishnamurthy V. Q -learning algorithms for constrained Markov decision processes with randomized monotone policies: Application to MIMO transmission control[J]. IEEE Transactions on Signal Processing, 2007, 55(5): 2170-2181.

[16] Altman E. Constrained Markov decision processes: stochastic modeling[M]. Routledge, 1999.

[17] Le H, Voloshin C, Yue Y. Batch policy learning under constraints[C]//International Conference on Machine Learning. PMLR, 2019: 3703-3712.

[18] Zang W, Miao F, Gravina R, et al. CMDP-based intelligent transmission for wireless body area network in remote health monitoring[J]. Neural computing and applications, 2020, 32: 829-837.

[19] Sun C, Stevens-Navarro E, Shah-Mansouri V, et al. A constrained MDP-based vertical handoff decision algorithm for 4G heterogeneous wireless networks[J]. Wireless Networks, 2011, 17: 1063-1081.

[20] Miryoosefi S, Brantley K, Daume III H, et al. Reinforcement learning with convex constraints[J].

ISSN 2616-5775 Vol. 6, Issue 13: 20-28, DOI: 10.25236/AJCIS.2023.061304

Advances in neural information processing systems, 2019, 32.
[21] Yang L, Rao H, Lin M, et al. Optimal sensor scheduling for remote state estimation with limited bandwidth: A deep reinforcement learning approach[J]. Information Sciences, 2022, 588: 279-292.
[22] Zhang J, Sun J, Lin H. Optimal DoS attack schedules on remote state estimation under multi-sensor round-robin protocol[J]. Automatica, 2021, 127: 109517.