Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Optimal Online Transmission Schedule for Remote State Estimation over a Hidden Markovian Channel ⋆

Bowen Sun, Xianghui Cao, Le Wang, Changyin Sun

School of Automation and School of Cyber Science and Engineering, Southeast University. (e-mails: {bwsun, xhcao}@seu.edu.cn)

Abstract: This paper investigates the optimal transmission scheduling problem in remote state estimation systems over an unreliable wireless channel where the channel state evolves as a Markov chain. However, due to inaccurate observations of the channel state, the wireless channel is modeled as a hidden Markov chain. We propose a prediction algorithm based on the Viterbi algorithm to estimate the channel state. To save the wireless sensor's energy, we consider scheduling the transmission of sensor transmissions while balancing between estimation performance and sensor energy expenditure. By jointly considering performance and energy, we formulate the scheduling problem as a Markov decision process. We prove the existence of the optimal transmission policy and derive a threshold structure of the optimal strategy. Finally, the performance of the proposed method is evaluated through simulations.

Keywords: Cyber-physical Systems, Markov decision process, Hidden Markov model

1. INTRODUCTION

Due to the rapid development of networking, communication technology, computing and modern control, cyber-physical systems (CPS) are onto a fast track of growth and focused by academy and industry widely (Ding et al. (2018)). The characteristics like high flexibility, rapid deployment and high fault tolerance in CPS bring many applications including medical, aerospace, automobile, etc. However, an increasing number of these applications demand remote estimation through the unreliable wireless network; thus, sensors need to make more flexible scheduling owing to energy constraints.

In this paper, we focus on the optimal transmission policy problem for the case where the optimal local estimation packet can be transmitted through an unreliable channel and the channel state is unknown. In the past decade, the packet loss process of wireless channels in the system has attracted considerable interest. Still, many kinds of research have focused on the distribution of channel packet loss conditions (Sinopoli et al. (2004)). The basic idea is that the packet loss modeled by the independent and identical distribution (i.i.d.) by Schenato et al. (2007) and to consider the impact of packet loss on the system estimation performance. Further, Garone et al. (2011) expanded the case to multiple channels. In order to improve system performance, Huang and Dey (2007) introduced Kalman filter to decrease the impact of system error and studied the estimation when Markov process is adopted to model the packet loss process. In Cao et al. (2014), the channel estimation performance was studied by cognitive radio, and the channel state model was proposed instead of the packet loss state model. Then, people extended their research objectives to the whole control or multiple systems. In Lin et al. (2019), state estimation and system performance analysis over non-acknowledgment networks with Markovian packet dropouts were studied. Marelli et al. (2019) considered the stability of the Kalman filter when its measurements are randomly lost, which means the measurement matrix and the measurement error covariance are random. Xu et al. (2019) studied the consensus communication problem of multi-agent systems through Markovian packet loss channels. Some researches are on the direction of attacks to cause packet loss, DoS attack was introduced in Qin et al. (2018) thus optimal attack scheduling needs to be designed. In practice, the channel state can not always be observed accurately, Barnes and Maharaj (2011) applied the hidden Markov model to cognitive radio to model channel occupancy, further Senthilkumar et al. (2018) proposed a hidden Markov framework to select the optimal channel for the users.

Motivated by the aforementioned works, we first consider the case where the channel has the memory about the old channel state which is very common in wireless communication that the assumption of independent packet loss not suitable. In the previous literature, the model process and the channel state for packet loss are widely studied such as i.i.d. or Markov process, while these can not describe the actual application fully thus we introduce the hidden Markov model to study the channel packet loss process and predict the channel state. As far as the channel state be concerned, it only exists two possibilities, being occupied to transmit packets (busy) and not occupied

⋆ This work was supported in part by the National Natural Science Foundation of China under Grants 61573103, 61520106009 and 61921004, in part by the Outstanding Young Scholarship of Jiangsu Province under Grant BK20180012, in part by the Southeast University "Zhongying Young Scholars" Project, and in part by the Key Laboratory of System Control and Information Processing, Ministry of Education, China (Corresponding author: Xianghui Cao).

(idle). Packet loss only occurs in the idle state where the channel is transmitting data, which means we should transmit packet when the channel is idle. However, our observation channel state is not satisfactory. Thus we use hidden Markov to model the channel state. In addition, wireless sensors are often powered by battery or have energy constraints, which means the number of times that the sensor can send packets in a certain period is limited. Not only that, for an online control system, the packet perceived by the sensor is also time-sensitive that we need to make the decision quickly on whether to transmit the packet. Our purpose is to design the optimal transmission policy to minimize energy consumption and maintain the optimal estimation performance under our channel model. The main contributions of this paper are summarized as follows:

(1) We formulate an optimization problem about the system estimation performance and transmission energy consumption comprehensively.
(2) We model the channel state as a hidden Markov process and formulate the sensor scheduling problem of state estimation as a Markov decision process (MDP) problem.
(3) We obtain the existence of the optimal transmission policy and prove that it is deterministic and stationary. Further, we get the threshold structure and analyze the equilibrium relationship between the average estimation error and energy.

The reminder of this paper is organized as follows. Section 2 provides the problem formulation. In Section 3, the design of the algorithm and the main results about the MDP are addressed. Section 4 presents a numerical example and shows how the model parameters affect the threshold structure and the equilibrium relationship. Finally, Section 5 shows some conclusions.

Notations: In the following, we present system model and adopt these notations: $\mathbb{R}$ is the set of real numbers. $\mathbb{Z}$ is the sets of integers. $\mathbb{R}^n$ represents the n-dimensional Euclidean space and $\mathbb{R}^{n\times m}$ represents the set of real matrices of dimension $n \times m$. $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$ denote the expectation and probability of a random variable, respectively. $\mathrm{Tr}(\cdot)$ and $\rho(\cdot)$ denote the trace and spectral radius of a square matrix, respectively. $\mathbb{S}_+^n$ is the set of positive semi-definite matrices of dimension $n \times n$, such that $X \geq 0$ if $X \in \mathbb{S}_+^n$. $\forall X, Y \in \mathbb{S}_+^n, X \geq Y$ if $X - Y \in \mathbb{S}_+^n$.

2. PROBLEM SETUP

The system architecture is described in Fig. 1. The states of physical process can be observed by a local sensor using Kalman filter and transmitted to the remote estimator through a wireless packet loss communication channel, besides, the sensor can also observe the state of wireless channels. Detailed description of the system consists of the following.

2.1 System Models

As shown in Fig. 1, we consider the following discrete linear time-invariant dynamic system

$$x_{k+1} = Ax_k + \omega_k \tag{1}$$

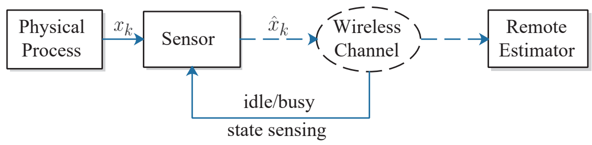

Fig. 1. System architecture

where $k \in \mathbb{Z}$ denotes the index of each time step, $x_k \in \mathbb{R}^n$ is the state vector, $\omega_k \in \mathbb{R}^n$ is the independent and identically distributed(i.i.d.) white Gaussian noise with zero mean and covariance $Q > 0$.

Besides, in Fig. 1, the system's measurement $y_k \in \mathbb{R}^l$ is monitored by the sensor and modeled by

$$y_k = Cx_k + v_k \tag{2}$$

where $y_k \in \mathbb{R}^l$ is the observation vector and v_k is also the i.i.d. white Gaussian noise with zero mean and covariance $R > 0$.

In the above, $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{l\times n}$ are all system matrices. We assume the system is stabilizable that the pair $(A, \sqrt{Q})$ is controllable and (A, C) is observable.

Channel Modeling Since the open-access wireless channel may be occupied by many other wireless devices, we assume that the state of the unreliable wireless channel suffers from the Markov process which includes two states, busy and idle, respectively. At each step k, the sensor first scans the wireless channel to judge whether its state is busy or idle and transmits the packet. We define a variable $\gamma_k = \{0, 1\}$ to represent whether the packet is transmitted to remote estimator successfully, specifically, $\gamma_k = 1$ means the transmission of packets is successful and $\gamma_k = 0$ means packet loss. The scanning time is assumed small enough that the channel state will not change during it. Denote $s_k \in \{0, 1\}$ and $o_k \in \{0, 1\}$ as the channel state and the observation result, respectively, i.e., $s_k = 0$ if the channel is idle and $s_k = 1$ otherwise; $o_k = 0$ if the channel is observed idle and $o_k = 1$ otherwise. Define the probability transition matrix Φ about the channel state that $\Pr(s_{k+1} = i|s_k = j)|_{i,j=\{0,1\}}$ as follows.

Proposition 1. $\{s_k\}$ complies with a homogeneous Markov chain with following transition probability matrix

$$\Phi = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{3}$$

where $\alpha = \Pr(s_k = 1|s_{k-1} = 0)$ and $\beta = \Pr(s_k = 0|s_{k-1} = 1)$.

Further, we assume the packet loss as an i.i.d. Bernoulli process when the channel is idle that the packet loss rate is denoted as ℓ that

$$\Pr(\gamma_k = 0) = \begin{cases} \ell, & \text{if } s_k = 0 \\ 1, & \text{if } s_k = 1 \end{cases} \tag{4}$$

Actually, the channel state can not be accurately observed thus the observation channel state has the incorrect possibility. Therefore, the sequence $\{s_k, o_k\}$ constitutes a hidden Markov model. Almost no consideration has been given in the previous literature.

Remark 2. The Markov process can be assumed that $0 < \alpha + \beta < 1$, if there's nothing about the past information at the certain step k, $\Pr(\gamma_k = 0)$ and $\Pr(\gamma_k = 1)$ are always the same, specially, $\Pr(\gamma_0 = 0) = \beta/(\alpha+\beta)$ and $\Pr(\gamma_0 = 1) = \alpha/(\alpha+\beta)$.

The sensor transmits its local estimates to the remote estimator. The local estimate and its corresponding estimation error covariance matrix are denoted as $\hat{x}_k$ and P_k, respectively:

$$\begin{aligned}\hat{x}_k =& \mathbb{E}[x_k|y_0,\cdots,y_k]\\ P_k =& \mathbb{E}[(x_k-\hat{x}_k)(x_k-\hat{x}_k)'|y_0,\cdots,y_k]\end{aligned}$$

Standard Kalman filter runs in the local estimator and the steady-state value of $\{P_k\}$ as $\bar{P}$ like Li et al. (2015).

2.2 Problem Formulation

The corresponding estimation error covariance P_k is described by the following equation due to packet loss of channel:

$$P_k = \begin{cases} h(P_{k-1}), & \text{if } \gamma_k = 0 \\ \bar{P}, & \text{if } \gamma_k = 1 \end{cases} \tag{5}$$

where $h(X) = AXA' + Q$.

It is apparent that the value of P_k is in the following infinitely countable set, that is

$$\{\bar{P}, h(\bar{P}), h^2(\bar{P}), \cdots\} \tag{6}$$

And then we adopt the expectation of P_k to describe it,

$$\mathbb{E}[P_k] = \Pr(\gamma_k = 1)\cdot\bar{P} + \Pr(\gamma_k = 0)\cdot h(\mathbb{E}[P_{k-1}]) \tag{7}$$

Furthermore, we can't ignore the energy constraint problem when the sensor prepares to transmit the packets; otherwise we can choose to transmit the packet at each time step regardless of whether or not the packet losses for optimal estimation. Therefore, we need to design an optimal transmission policy to maintain estimation performance under energy constraints. Considering the following cost function about estimation performance, we construct the optimal packet transmission strategy under limited energy such that the estimation performance and energy-related transmission get balanced.

The average expectation of estimation error covariance can be defined to describe the estimation performance:

$$J^1_{\boldsymbol{\theta}} = \limsup_{T\to\infty}\frac{1}{T+1}\mathbb{E}\left[\sum_{k=0}^{T}\mathrm{Tr}(P_k)\right] \tag{8}$$

where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots\}$ is the sequence of whether the sensor transmits the packet at each time.

The transmission energy is related to whether the sensor chooses to send the packet or not. We assume that the energy consumption of the packet transmitted at each time is the same, that the total energy consumption can be denoted by the exception of packet transmission:

$$J^2_{\boldsymbol{\theta}} = \limsup_{T\to\infty}\frac{1}{T+1}\mathbb{E}\left[\sum_{k=0}^{T}a_k\right] \tag{9}$$

Our purpose is to find the optimal packet transmission strategy under limited energy to minimize the balance between the average estimation error and the total energy, that is

Problem 3.

$$\min_{\boldsymbol{\theta}} \quad J^1_{\boldsymbol{\theta}} + \eta J^2_{\boldsymbol{\theta}} \tag{10}$$

where η is the variable to measure the importance of estimation performance and energy.

3. OPTIMAL TRANSMISSION ENERGY STRATEGY

In this section, we formulate Problem 3 as a Markov Decision Process (MDP) problem. Some preparation works is necessary about the system state.

3.1 The Hidden Markov Process

We can't directly sense the state of the wireless channel, only through the sensor to observe. The true channel state is hidden behind the observation channel states that they form a hidden Markov model. Nevertheless, during the long-term operation of the system, we can obtain all the observation states in the past and use these to predict the possibility of the channel state at the current moment.

We need to define some extra notations as follows: Define the correct detection and false sensing probabilities about the wireless channel, p_d, p_f, respectively. Moreover, define transition martix about the probability $\Pr(o_k|s_k)$ as $\Psi = \begin{bmatrix} p_d & 1-p_d \\ 1-p_f & p_f \end{bmatrix}$, where p_d is the correct detection of the idle state and p_f is the correct detection of the busy state. Define the probability about state of the wireless channel at time T as p_T for $s_T = 0$ and $1-p_T$ for $s_T = 1$, respectively. In order to get the prediction results of the current state, we assume that all the observation states in the past are known, define the sequence of old observations as $Y = \{y_1, y_2, \cdots, y_T\}$ and the initial probability of state s is $\Pi = \{\pi_1 = \frac{\beta}{\alpha+\beta}, \pi_2 = \frac{\alpha}{\alpha+\beta}\}$.

Algorithm 1 describes how to calculate the possibility of the current state of the wireless channel based on historical information, $U_{i,j}$ represents the maximum probability of state j at time i (while after iterating, the value of $U_{i,j}$ can't equal to probability directly). In step 11-13, we can find the maximum likelihood of each hidden state s at time T by iterating from the initial time 2 to the final time T. In step 15, p calculate the current probability from $U_{i,j}$.

Algorithm 1 The predicted probability of channel state

```
Input: The channel state S = {s_1, s_2}
 1:  The observation channel state O = {o_1, o_2}
 2:  The initial probability of state Π = {π_1, π_2}
 3:  The sequence of old observations Y = {y_1, y_2, ···, y_T}
 4:  The transition matrix A = Φ
 5:  The emission matrix B = Ψ
Output: The probability p_T of hidden state s_k = 0 at T
 6: i ← 2
 7: for i ← 1, 2 do
 8:     U_(1,i) ← π_i · B_{i y_1}
 9: end for
10: for j ← 2 to T do
11:     for i ← 1, 2 do
12:         U_(j,i) ← max_{k=1,2} U_(j−1,k) · A_{ki} · B_{i y_j}
13:     end for
14: end for
15: p_T ← U_{T,1} / Σ_{i=1}^{2} U_{T,i}
```

3.2 MDP Formulation

Firstly, We define a random variable τ_k to denote the time duration from the last successful transmission time to time k as:

$$\tau_k = k - \max\{k^* : \gamma_{k^*} = 1, 0 \leq k^* \leq k\} \tag{11}$$

Then according to (6), we can rewrite P_k as

$$P_k = h^{\tau_k}(\bar{P}) \tag{12}$$

At each step k, we denote the state of the MDP process as $s_k = P_{k-1}$, $\mathbb{A} = \{0, 1\}$ as the transmission action set and the action of the MDP process as $a_k \in \mathbb{A}$, that is the sensor chooses action $a_k = a(s_k)$ based on the process s_k.

Therefore, at time k, the one-stage reward can be written as

$$r(P_{k-1}, a_k) = \mathrm{Tr}[P_{k-1} + \eta a_k I] \tag{13}$$

where I is identity matrix.

Remark 4. The value of state at time k is $s_k = P_{k-1}$, thus the value at time $k+1$ will have two values depending on γ_k: when $\gamma_k = 1$, $s_{k+1} = P_k = \bar{P}$, otherwise, $s_{k+1} = P_k = h(P_{k-1})$.

The state transition probability of the MDP can be written as

$$t(P_k|P_{k-1}, a_k) = \begin{cases} (1-\ell) \cdot p_k, & \text{if } P_k = \bar{P}, a_k = 1 \\ \ell \cdot p_k, & \text{if } P_k = h(P_{k-1}), a_k = 1 \\ 0, & \text{if } P_k = \bar{P}, a_k = 0 \\ 1 - p_k, & \text{if } P_k = h(P_{k-1}), a_k = 0 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Denote the transmission strategy of the sensor as $\boldsymbol{\theta} = \{a_t(s_k)\}_{k=1,\cdots,T}$ and all transmission policies as Θ, then the following function represent the the average of the expected sum of rewards r from the average expected cost criterion under the strategy $\boldsymbol{\theta} \in \Theta$ and the initial state $s(1) = s \in \mathbb{S}$, that is

$$J_{s,\boldsymbol{\theta}} = \limsup_{T \to \infty} \frac{1}{T+1} E\left[\sum_{k=0}^{T} r(s_k, a_k)\right] \tag{15}$$

and its optimal value is $J^*(s) = \arg\min_{\boldsymbol{\theta} \in \Theta} J(s, \boldsymbol{\theta})$.

From the about discussion, Problem 3 can be solved by Markov decision process. Define the following average value function as $V_{\boldsymbol{\theta}} : \mathbb{S} \to \mathbb{R}$, thus the optimal cost $J^*(s)$ can be rewritten as a Bellman equation (Puterman (2014)) as follows

$$\begin{aligned} J^*(s_k) + V_{\boldsymbol{\theta}}(s_k) = \min_{a \in \mathbb{A}} \Bigg\{ & r(s_k, a) \\ & + \sum_{s'_{k+1} \in \mathbb{S}} t(s'_{k+1}|s_k, a) V_{\boldsymbol{\theta}}(s'_{k+1}) \Bigg\} \end{aligned} \tag{16}$$

Therefore, the optimal transmission policy for the sensor is given by

$$\begin{aligned} a^*(s_k) = \arg\min_{a_k \in \mathbb{A}} \Bigg\{ & r(s_k, a_k) \\ & + \sum_{s'_{k+1} \in \mathbb{S}} t(s'_{k+1}|s_k, a_k) J^*(s'_{k+1}) \Bigg\} \end{aligned} \tag{17}$$

3.3 Structural Results for the Optimal Transmission Policy

We prove that the existence of optimal strategy, further deduce its deterministic and stationary.

Theorem 5. The optimal deterministic and stationary policy for $\boldsymbol{\theta}^* \in \Theta$ exists, that is

$$J(s, \boldsymbol{\theta}^*) \geq J(s, \boldsymbol{\theta}) \qquad \forall s \in \mathbf{S}, \boldsymbol{\theta} \in \Theta \tag{18}$$

The optimal value $\boldsymbol{\theta}^*$ can be solved by Bellman equation. The proof of Theorem 5 can be given in Appendix A based on Sennott (1986).

According to this theorem, we can drive the optimal transmission strategy of (15) from Bellman equation (16) and this strategy is deterministic and stationary, which help us analyze the structural characteristics of the optimal strategy further.

And then we present some lemmas for analyzing the structural properties of the optimal solution.

Lemma 6. (Shi et al. (2011)) If $1 \leq \tau_1 \leq \tau_2$, then $h^{\tau_1}(\bar{P}) \leq h^{\tau_2}(\bar{P})$, $h(\bar{P}) \neq \bar{P}$.

Lemma 7. (Puterman (2014)) For two partially ordered sets X, Y, define $g(x, y)$ as a real-valued function on $X \times Y$, $g(x, y)$ is so-called superadditive when all $x^+ \geq x^-$ in X and $y^+ \geq y^-$ in Y, that is

$$g(x^+, y^+) + g(x^-, y^-) \geq g(x^+, y^-) + g(x^-, y^+) \tag{19}$$

And then, we present the existence of special threshold structure for our optimal transmission policy to reduce our computational complexity.

Theorem 8. The optimal transmission policy $a(s_k)$ is nondecreasing in s and has the following threshold structure:

$$a(s_k) = \begin{cases} a_1 & \mathrm{Tr}(s_k) < \mathrm{Tr}(s^*) \\ a_2 & \mathrm{Tr}(s_k) \geq \mathrm{Tr}(s^*) \end{cases} \tag{20}$$

where s^* is the limit function which can be solved by iterating according to specific problems.

The proof of Theorem 8 is shown in Appendix B.

The monotonic transmission policy means that we can find a specific threshold structure to make decision according to the specific problems, take our problem as an example, $s^* = P^*$ likes a boundary thus if $\mathrm{Tr}(s_k) > \mathrm{Tr}(s)$, we choose action a_1 that we do not transmit packet.

4. NUMERICAL EXAMPLE

We present two numerical examples to show the validity of our algorithm and illustrate the optimal transmission policy about Problem 3.

Firstly we set some initial parameters, the transition matrix $\Phi = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$, the emission matrix $\Psi = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$ and the initial state $\Pi = \{0.5, 0.5\}$.

We set one possible channel state sequence through the given state transition matrix and the initial value, and then we calculate one possible observation channel state sequence. Next, we use our prediction algorithm to calculate the possibility of the possible channel state sequence based on the observation channel state sequence. Our clever initial setup is used to compared easily that we only

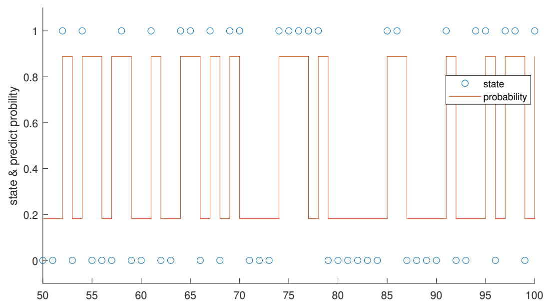

Fig. 2. Actual channel state and predicted probability

need to compare with 0.5. The final results are shown in Fig. 2 and we show the comparison of the predicted probability of busy channel state and the actual channel state from time 50 to time 100, our prediction accuracy is about 85%.

The simulation example above shows the reliability of our prediction algorithm, and then we can use fixed probability to simulate our Markov decision process. Firstly, we set some initial parameters in Table 1.

Table 1. Some initial parameters

Q	A	ℓ	p_k	α
$\begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 1.1 & 0.7 \\ 3.2 & 1.3 \end{bmatrix}$	0.5	0.5	0.95

And then, we consider τ_{k-1} to represent the state P_{k-1} which is the function of it. η is used to reflect the importance of transmission energy in the decision process. That is the larger η is, the more energy will be consumed during each transmission. So we can derive the optimal action decision in case of different τ_{k-1} and η like Fig. 3 by applying the policy iteration algorithm in Markov decision process toolbox of MATLAB.

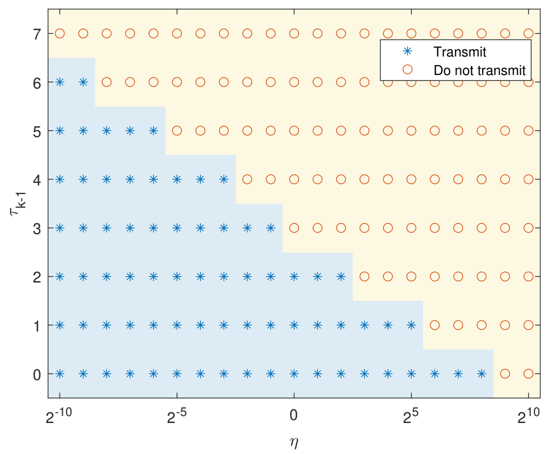

Fig. 3. Optimal transmission action a_k of τ_{k-1} and η

In Fig. 3 we can find some explicit conclusions. Firstly, the figure represents the threshold structure of the policy a_k in s_k intuitively which is described by Theorem 8. Furthermore, the figure shows that the policy also has the threshold structure in η explicitly. Secondly, when the transmission energy consumption increases, our policy tends to decrease the number of sending packets, another is when consecutive loss rate increases, our policy also tends to do nothing about sending packets to decrease the risk of packet loss and energy consumption.

And then, we fix $\eta = 1$ to study the influence of state prediction probability p_k on decision making, the initial parameters is like Fig. 3, the optimal transmission action policy is shown in Fig. 4. We can find that with the increase of probability that the system channel state is predicted as idle, our policy tends to send packets.

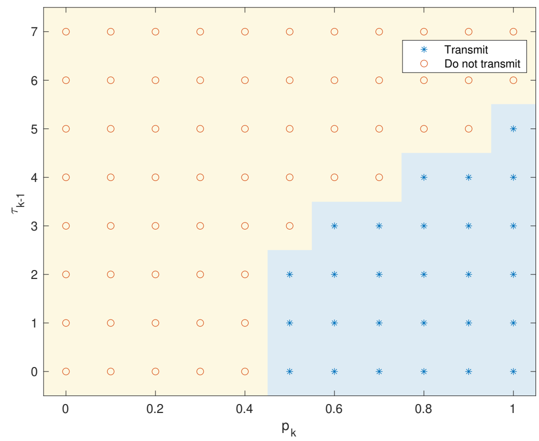

Fig. 4. Optimal transmission action a_k of τ_{k-1} and p_k

5. CONCLUSION

In this paper, we have investigated the optimal transmission policy that the sensor transmits or not under the energy constraint and incomplete channel cognition. We design an algorithm to predict the unreliable channel state and apply the MDP algorithm to schedule the optimal sensor transmission. We prove the existence of the policy which has the characteristics of deterministic, stationary and threshold structure. Our simulations discover that the threshold structure exists not only in the state s of the MDP process, but in the variable η and state prediction probability p. Our further work is to interrupt a unified decision probability model instead of a specific scheduling scheme for a given problem.

REFERENCES

Barnes, S. and Maharaj, B. (2011). Performance of a hidden markov channel occupancy model for cognitive radio. In *2011 IEEE Africon Conference*, 1–6. IEEE.

Cao, X., Cheng, P., Chen, J., Ge, S.S., Cheng, Y., and Sun, Y. (2014). Cognitive radio based state estimation in cyber-physical systems. *IEEE Journal on Selected Areas in Communications*, 32(3), 489–502.

Ding, D., Han, Q.L., Xiang, Y., Ge, X., and Zhang, X.M. (2018). A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing*, 275, 1674–1683.

Garone, E., Sinopoli, B., Goldsmith, A., and Casavola, A. (2011). Lqg control for mimo systems over multiple erasure channels with perfect acknowledgment. *IEEE Transactions on Automatic Control*, 57(2), 450–456.

Huang, M. and Dey, S. (2007). Stability of kalman filtering with markovian packet losses. *Automatica*, 43(4), 598–607.

Li, Y., Shi, L., Cheng, P., Chen, J., and Quevedo, D.E. (2015). Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach. *IEEE Transactions on Automatic Control*, 60(10), 2831–2836.

Lin, H., Lam, J., Wang, Z., and Lam, H.K. (2019). State estimation over non-acknowledgment networks with markovian packet dropouts. *Automatica*, 109, 108484.

Marelli, D.E., Sui, T., Rohr, E.R., and Fu, M. (2019). Stability of kalman filtering with a random measurement equation: Application to sensor scheduling with intermittent observations. *Automatica*, 99, 390–402.

Peng, L., Shi, L., Cao, X., and Sun, C. (2017). Optimal attack energy allocation against remote state estimation. *IEEE Transactions on Automatic Control*, 63(7), 2199–2205.

Puterman, M.L. (2014). *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

Qin, J., Li, M., Shi, L., and Yu, X. (2018). Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks. *IEEE Transactions on Automatic Control*, 63(6), 1648–1663.

Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K., and Sastry, S.S. (2007). Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1), 163–187.

Sennott, L.I. (1986). A new condition for the existence of optimal stationary policies in average cost markov decision processes. *Operations Research Letters*, 5(1), 17–23.

Senthilkumar, S. et al. (2018). Hidden markov model based channel selection framework for cognitive radio network. *Computers & Electrical Engineering*, 65, 516–526.

Shi, L., Cheng, P., and Chen, J. (2011). Sensor data scheduling for optimal state estimation with communication energy constraint. *Automatica*, 47(8), 1693–1698.

Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M.I., and Sastry, S.S. (2004). Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9), 1453–1464.

Xu, L., Mo, Y., and Xie, L. (2019). Distributed consensus over markovian packet loss channels. *IEEE Transactions on Automatic Control*.

Appendix A. PROOF OF THEOREM 5

First we introduce a lemma which is important to the following proof.

Lemma 9. (Sennott (1986)) If $J^*(s) = \arg\min_{\boldsymbol{\theta}} J(s, \boldsymbol{\theta})$ is the infimum over all policies of the expected cost incurred when the process starts in channel state s, then the $J^*(s)$ satisfies the optimality equation

$$J^*(s) = \min_{a\in\mathbb{A}(s)} \left\{ r(s,a) + \alpha \sum_{s\in\mathbb{S}} t(s'|s,a) J^*(s') \right\} \tag{A.1}$$

where α is the discount factor and the average cost optimal policy can use discounted cost policy to expression. Following this lemma the theorem 1 is straightforward.

According to the results in Sennott (1986), we prove this theorem. Due to the action set $\mathbb{A} = \{0, 1\}$ is a finite set, the state space $\mathbb{S} = \{\bar{P}, h(\bar{P}), h^2(\bar{P}), \cdots\}$ in our problem is denumerable, base on the Assumption 1, 2 in Sennott (1986), we only to verify the following two conditions:

(1) There exists a bounded solution to
$$|J^*(s) - J^*(0)| < N \tag{A.2}$$
for all s and non-negative N.

(2) There exist a non-negative function M_s and non-negative integers B that
$$\sum_{j\in\mathbb{S}} t(j|s,\boldsymbol{\theta}) M_j < B \tag{A.3}$$

(3) There exists that
$$\sum_{j\in\mathbb{S}} t(j|s,\boldsymbol{\theta}) M_j + r(s,\boldsymbol{\theta}) \le M_s \tag{A.4}$$

First, we prove condition (1) holds. From (13), $J^*(s)$ is additive and $r(s_k, a_k)$ is increasing, thus $J^*(s)$ is increasing, the left half of (A.2) holds. In addition, we find the right half of it is coincident with condition (2) cleverly.

Next, consider the policy $\boldsymbol{\theta}$ in our transmission, define the following function like Peng et al. (2017):

$$f(X) = AXA + \varphi I \tag{A.5}$$

where we can find a positive constant φ that $\bar{P} \le \varphi I$ and $Q \le \varphi I$. Therefore, we can get

$$h^{\tau_k}(\bar{P}) \le f^{\tau_k}(\bar{P}) \le f^{\tau_k}(\varphi I) \le \varphi \sum_{j=0}^{\tau_k} A^j (A')^j \tag{A.6}$$

When $k = 0$, $h^{\tau_k}(\bar{P}) = \bar{P} \le \varphi I$, based on the idea of the average iterated expected cost function like (15) that the cost function $V(s)$, the reward function $r(s_k, a_k)$ and (A.6) that $\mathrm{Tr}(h^{\tau_k}(\bar{P}) + \eta a_k) \le \mathrm{Tr}(h^{\tau_k}(\bar{P}) + \eta) \le \varphi \sum_{j=0}^{\tau_k} (\rho^{2j}(A) + \eta)$ and $s_k = (h^{\tau_{k-1}}(\bar{P}))$, we can define

$$M_s = \phi \sum_{j=0}^{\tau_{k-1}} (\rho^{2j}(A) + \eta) \tag{A.7}$$

where define $\phi > \varphi$ for the sake of satisfying condition (3) formally. For τ_k is finite and then

$$\begin{aligned}\sum_{j\in\mathbb{S}} t(j|s,\boldsymbol{\theta}) M_j =& (1-\ell)p_k\phi + \ell p_k \phi \sum_{j=0}^{\tau_{k-1}+1} (\rho^{2j}(A) + \eta) \\ &+ (1-p_k)\phi \sum_{j=0}^{\tau_{k-1}+1} (\rho^{2j}(A) + \eta) \\ =& \phi((1 - p_k + \ell p_k) \sum_{j=0}^{\tau_{k-1}+1} \rho^{2j}(A) \\ &+ \eta(\tau_{k-1}+1)(1 - p_k + \ell p_k) + (1-\ell)p_k)\end{aligned} \tag{A.8}$$

Due to $\rho(A)$ is bounded for all matrix A thus in finite horizon, we can find a non-negative integers B greater than (A.8), so we finish the proof of condition (2) and the right half of (A.2).

Then, we prove condition (3),

$$
\begin{aligned}
&\sum_{j\in\mathbb{S}} t(j|s,\boldsymbol{\theta})M_j + r(s,\boldsymbol{\theta}) - M_s \\
=&\phi(\ell p_k + (1-p_k)) \sum_{j=0}^{\tau_{k-1}+1} (\rho^{2j}(A)+\eta) \\
&+ \mathrm{Tr}(h^{\tau_{k-1}}(\bar{P}) + \eta a_k) - \phi\sum_{j=0}^{\tau_{k-1}} (\rho^{2j}(A)+\eta) \\
&+ (1-\ell)p_k\phi \\
\leq&(\varphi - \phi(1-(1-(1-\ell)p_k)\rho^2(A))) \sum_{j=0}^{\tau_{k-1}} (\rho^{2j}(A)+\eta) \\
&+ \phi(1-\ell)p_k
\end{aligned} \tag{A.9}
$$

When $\rho(A) \leq 1$, (A.9) can be expressed as

$$(\varphi - \phi(1-\ell)p_k)k(1+\eta) + \phi(1-\ell)p_k \leq 0$$

therefore

$$\phi \geq \frac{k(1+\eta)}{(k(1+\eta)-1)(1-\ell)p_k}\varphi \tag{A.10}$$

when $k \to \infty$, $\phi \geq \frac{\varphi}{(1-\ell)p_k}$ so as to make (A.9) nonpositive. Otherwise, when $\rho(A) > 1$, (A.9) is expressed as

$$\phi \geq \frac{-k(\rho^{2k}(A)+\eta)}{(1-\ell)p_k + k(\rho^{2k}(A)+\eta)((\ell p_k - p_k + 1)\rho^2(A) - 1)}\varphi$$

when $k \to \infty$, $\phi \geq \frac{\varphi}{1-(1-(1-\ell)p_k)\rho^2(A)}$ for the sake of making (A.9) nonpositive.And then, we finish the proof of condition (3).

Appendix B. PROOF OF THEOREM 8

In order to prove the existence of threshold structure, we need to verify the following four conditions depending on Puterman (2014).

(1) $r(s,a)$ is nondecreasing in s for all $a \in \mathbb{A}$.
(2) $\sum_{j\succeq s'_k} t(j|s_k,a_k)$ is nondecreasing, $\forall s_k \in \mathbb{S}$ in s for all $a \in \mathbb{A}$.
(3) $r(s,a)$ is superadditive function on $\mathbb{S}\times\mathbb{A}$.
(4) $\sum_{j\succeq s'_k} t(j|s_k,a_k)$ is is superadditive function on $\mathbb{S}\times\mathbb{A}$, $\forall s'_k \in \mathbb{S}$.

Firstly we prove that condition (1) and (3) exist below.

We choose the $s = h^{\tau_{k-1}}(\bar{P})$ and $j = h^{\tau_k}(\bar{P})$ at time k, based on Lemma 6, for $a = 0$ and 1, $r(s,a) = \mathrm{Tr}[P_{k-1} + \eta a_k I] = \begin{cases} \mathrm{Tr}[h^{\tau_{k-1}}(\bar{P})] & a_k = 0 \\ \mathrm{Tr}[h^{\tau_{k-1}}(\bar{P})] + n\eta & a_k = 1 \end{cases}$ is nondecreasing in s.

In addition, Lemma 6 also derive the following for $\tau^+ \geq \tau^-$

$$h^{\tau^-}(\bar{P}) \geq h^{\tau^+}(\bar{P}) \tag{B.1}$$

For $a^+ \geq a^-$ and fixed a, we can get

$$r(s^+,a^+) - r(s^+,a^-) \geq r(s^-,a^+) - r(s^-,a^-) \tag{B.2}$$

and then $r(s,a)$ is superadditive function, condition (2) is satisfied.

And then we prove that condition (2) and (4) exist below.

We fix a_k for all $a_k \in \mathbb{A}$ and define $s_k^- = h^{\tau_{k-1}}(\bar{P})$, $s_k^+ = h^{\tau_k}(\bar{P})$, which is $s_k^- < s_k^+$.

Let $k = 1$ as an example,

- If $s'_k = h^0(\bar{P}) = \bar{P}$, that is $s'_k \leq s_k^- < s_k^+$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k) = \sum_{j\succeq s'_k} t(j|s_k^-,a_k) = (1-\ell)p_k$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h(\bar{P})$, that is $s_k^- < s'_k \leq s_k^+$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k) = \sum_{j\succeq s'_k} t(j|s_k^-,a_k) = \ell p_k$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h^2(\bar{P})$, that is $s_k^- < s_k^+ < s'_k$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k) = \ell p_k > \sum_{j\succeq s'_k} t(j|s_k^-,a_k) = 0$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h^2(\bar{P})$, that is $s_k^- < s_k^+ < s'_k$ and $\tau'_k - \tau_k^+ \geq 2$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k) = \sum_{j\succeq s'_k} t(j|s_k^-,a_k) = 0$$

Then we can get $\sum_{j\succeq s'_k} t(j|s_k,a_k)$ is nondecreasing aforementioned.

Furthermore, we prove that is $\sum_{j\succeq s'_k} t(j|s_k,a_k)$ superadditive function. We make the same classification discussion like above and fix a_k.

- If $s'_k = h^0(\bar{P}) = \bar{P}$, that is $s'_k \leq s_k^- < s_k^+$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k^+) = \sum_{j\succeq s'_k} t(j|s_k^-,a_k^+) = (1-\ell)p_k$$
$$\sum_{j\succeq s'_k} t(j|s_k^-,a_k^-) = 1-p_k > \sum_{j\succeq s'_k} t(j|s_k^+,a_k^-) = 0$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h(\bar{P})$, that is $s_k^- < s'_k \leq s_k^+$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k^+) = \ell p_k > \sum_{j\succeq s'_k} t(j|s_k^-,a_k^+) = 0$$
$$\sum_{j\succeq s'_k} t(j|s_k^-,a_k^-) = \sum_{j\succeq s'_k} t(j|s_k^+,a_k^-) = 1-p_k$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h^2(\bar{P})$, that is $s_k^- < s_k^+ < s'_k$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k^+) = \ell p_k > \sum_{j\succeq s'_k} t(j|s_k^-,a_k^+) = 0$$
$$\sum_{j\succeq s'_k} t(j|s_k^-,a_k^-) = \sum_{j\succeq s'_k} t(j|s_k^+,a_k^-) = 0$$
- If $s'_k = h^{\tau_{k-1}}(\bar{P}) = h^2(\bar{P})$, that is $s_k^- < s_k^+ < s'_k$ and $\tau'_k - \tau_k^+ \geq 2$,
$$\sum_{j\succeq s'_k} t(j|s_k^+,a_k^+) = \sum_{j\succeq s'_k} t(j|s_k^-,a_k^+) = 0$$
$$\sum_{j\succeq s'_k} t(j|s_k^-,a_k^-) = 1-p_k > \sum_{j\succeq s'_k} t(j|s_k^+,a_k^-) = 0$$

Thus we can get $\sum_{j\succeq s'_k} t(j|s_k,a_k)$ is superadditive function.

To summarize, we verify that our model satisfies the four conditions above, that is existing a limsup average optimal stationary policy with the property that the optimal policy is nondecreasing in s.

Thus, we can derive that the optimal policy $a(s_k)$ has the special threshold structure form.