Automatica 96 (2018) 22–31

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Optimal scheduling of multiple sensors over shared channels with packet transmission constraint ☆

Shuang Wu [a], Xiaoqiang Ren [b,*], Subhrakanti Dey [c], Ling Shi [a]

[a] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[c] ITR, University of South Australia, Mawson Lakes, SA 5095, Australia

ARTICLE INFO

Article history:
Received 9 September 2017
Received in revised form 25 February 2018
Accepted 14 May 2018

Keywords:
Kalman filters
State estimation
Sensor scheduling
Packet length
Markov decision process
Switching type policy

ABSTRACT

In this work, we consider the optimal sensory data scheduling of multiple process. A remote estimator is deployed to monitor S independent linear time-invariant processes. Each process is measured by a sensor, which is capable of computing a local estimate and sending its local state estimate wrapped up in packets to the remote estimator. The lengths of the packets are different due to different dynamics of each process. Consequently, it takes different time durations for the sensors to send the local estimates. In addition, only a portion of all the sensors are allowed to transmit at each time due to bandwidth limitation. We are interested in minimizing the sum of the average estimation error covariance of each process at the remote estimator under such packet transmission and bandwidth constraints. We formulate the problem as an average cost Markov decision process (MDP) over an infinite horizon. We first study the special case when $S = 1$ and find that the optimal scheduling policy always aims to complete transmitting the current estimate. We also derive a sufficient condition for boundedness of the average remote estimation error. We then study the case for general S. We establish the existence of a deterministic and stationary policy for the optimal scheduling problem. We find that the optimal policy has a consistent property among the sensors and a switching type structure. A stochastic algorithm is designed to utilize the structure of the policy to reduce computation complexity. Numerical examples are provided to illustrate the theoretical results.

1. Introduction

Networked control systems are control systems which incorporate communication systems. With the wireless communication technology, control systems can be remotely operated and set up in a distributed configuration. Applications of networked control systems are numerous, including industrial automation, habitat monitoring, smart grid and autonomous traffic management (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002). A salient feature of the networked controls systems is that the control signal and sensory data are transmitted via packets through a communication network. Although the wireless communication channel facilitates remote operation, algorithm design for the whole control system becomes more difficult as the communication channels have limited capacity to deliver information perfectly. The packets may be dropped, delayed, out-of-sequence, etc. (You & Xie, 2013; Zhang, Gao, & Kaynak, 2013).

As state estimation is crucial to feedback control systems, efficient processing of sensory data under limited resources is crucial. The accuracy of the state estimation can be improved by designing customized algorithms of sensory data scheduling. These algorithms tradeoff the estimation performance and constraints on the available resources, e.g., channel bandwidth, energy budget, etc. Shi, Cheng, and Chen (2011) proposed a periodic optimal sensory data scheduling policy for a single sensor under a limited energy budget. Other works on optimal allocation of energy for sensors can be found in Nourian, Leong, and Dey (2014) and Wu, Li, Quevedo, Lau, and Shi (2015) and references therein. To maximize lifetime of the network, Mo, Shi, Ambrosino, and Sinopoli (2009) provided a sensor selection algorithm based on convex optimization. As communication is expensive in wireless sensor network, efficient utilization of online information to reduce communication rate is another research focus. Event-triggered transmission strategies

☆ The work by S. Wu and L. Shi is supported by a Hong Kong RGC General Research Fund 16222716. The material in this paper was presented at the 20th World Congress of the International Federation of Automatic Control, July 9–14, 2017, Toulouse, France. This paper was recommended for publication in revised form by Associate Editor Hyeong Soo Chang under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: swuak@ust.hk (S. Wu), xren@ntu.edu.sg (X. Ren), subhrakanti.dey@angstrom.uu.se (S. Dey), eesling@ust.hk (L. Shi).

https://doi.org/10.1016/j.automatica.2018.06.019
0005-1098/

were proposed to deal with this issue (Ding, Wang, Ho, & Wei, 2017; Ren, Wu, Johansson, Shi, & Shi, 2018; Shi, Chen, & Shi, 2014; Wu, Jia, Johansson, & Shi, 2013). The bandwidth of wireless communication channel can be limited, and only a few sensors are allowed to be activated in each time slot because the sensors can interfere with each other. Gupta, Chung, Hassibi, and Murray (2006) analyzed the performance of a stochastic sensor selection algorithm. Other researches related to the optimal scheduling of multiple sensors under channel bandwidth constraints can be found in Han, Wu, Zhang, and Shi (2017), Hovareshti, Gupta, and Baras (2007), Mo, Garone, and Sinopoli (2014), Ren, Wu, Dey, and Shi (2018) and Zhao, Zhang, Hu, Abate, and Tomlin (2014). Mo, Garone et al. (2014) and Zhao et al. (2014) proved that the optimal sensor scheduling scheme can be approximated arbitrarily by a periodic schedule over an infinite horizon. As the sensors are nowadays equipped with storage buffer and on-board computation unit, pre-processing can be done on the sensor. Hovareshti et al. (2007) showed that the estimation quality can be improved for such type of smart sensors. Han et al. (2017) proved that the optimal scheduling policy over an infinite horizon can also be arbitrarily approximated by a periodic schedule if smart sensors are used.

Many previous works assumed that it takes the same time duration for transmission of all packets and the transmissions are done in one time step. This is based on the assumption that the data packets have the same length. To improve control system performance by efficiently utilizing the communication resources, the packet length can be chosen to be different. Pioneering work was done in Tamboli and Manikopoulos (1995). Optimal allocation of packet length has been studied in wireless communication community, e.g., Dong et al. (2014). It was shown in Mori, Ishii, and Ogose (2011) that both latency and throughput performance of the communication channel can be improved. In some state-of-the-art communication protocols, e.g., Time Slotted Channel Hopping (TSCH, IEEE 802.15.4e), the packet length of different packets can be different and the data transmission scheme can be designed accordingly. Standards of TSCH application can found in Watteyne, Palattella, and Grieco (2015). Besides the communication protocols, some applications, e.g., underwater vehicles, can only bear a few bits in one transmission (Cui, Kong, Gerla, & Zhou, 2006). In those cases, the local estimate should be split into more than one packet. Zhao, Kim, Shi, and Liu (2011) studied how to tradeoff the quality of control and the quality of service in the communication channel through optimal packet length allocation.

Different from previous works on sensory data scheduling, our work focuses on the constraint of packet length. We study the sensory data scheduling for remote state estimation of multiple linear time-invariant stochastic processes, each driven by white Gaussian noises. Every process is measured by a sensor, which is able to compute the local estimate of the process and send the local estimate to a remote estimator. Because the dynamics of different processes are different, the packet lengths of each process may be different. Consequently, it costs different time durations for each sensor to complete one transmission of an estimate. Moreover, not all the sensors can transmit data to the remote estimator at the same time step due to bandwidth limitations. We are interested in minimizing the average estimation error at the remote estimator over an infinite time horizon.

Some preliminary results have been reported in our conference paper (Wu, Ren, Dey, & Shi, 2017), in which we formulated the optimal scheduling of sensors under the packet length and bandwidth constraints as an infinite time horizon Markov decision process (MDP) with an average cost criteria. We proved that there exists a deterministic and stationary optimal policy for the MDP. Moreover, we showed that the optimal policy has nice properties, i.e., consistency and switching type structure. The consistency means that once a sensor is chosen to schedule, the transmission of the current estimate should not be interrupted by selecting other sensors. The switching type structure stands that the policy on the state space is separated by curves (2-dimension) or hyperplanes (3-dimension or higher).

The problem considered in this work is challenging because scheduling multiple sensors monitoring multiple processes can be classified as a restless bandit problem, which is proven to be computationally intractable (Papadimitriou & Tsitsiklis, 1999). Moreover, as we consider the packet length constraint and lossy transmission, the state space of the MDP formulation is embedded in a high dimensional space. The analysis of the corresponding MDP is thus complex as well.

Compared with the conference paper, our novel contributions are as follows.

(1) Different from the perfect channel assumption in the conference paper, the problem formulation of this work includes the case when the communication channel is lossy. The erasure phenomenon can make the remote estimation error unbounded (e.g., Sinopoli et al. (2004)). Suppose a local estimate, which might consist of several packets) has been selected to be transmitted. If the starting packet has been received by the remote estimator and the ending packet has not been received by the remote estimator, the estimate is defined as in-transmission. When a packet loss occurs, transmitting the most updated or the current in-transmission estimate is to be studied. In this work, we show that it is optimal to continue transmitting the current in-transmission estimate, which can be viewed as consistency within one process for the optimal policy.

(2) In our conference work, the proof of the existence of a deterministic and stationary policy applies only to the unstable processes in a perfect channel. In this work, we adopt another framework and establish the existence result by showing a set of conditions in Sennott (1996).

(3) We extend the results of switching policy to the lossy channel case. By utilizing the switching structure, we develop a stochastic numerical algorithm to compute the switching curve of the optimal policy. This reduces computation overhead compared with directly applying a value iteration algorithm to the original MDP problem.

The remainder of this paper is organized as follows. In Section 2, we provide the mathematical formulation of the problem of interest. The main results, which consist of the MDP formulation, the existence of a deterministic and stationary policy, the structure of the optimal policy, and a stochastic algorithm for computing the switching curve are given in Section 3. In Section 4, a numerical example is provided to illustrate the main results. We summarize the paper in Section 5.

Notation: Denote $\mathbb{N}$ as the set of integers greater than zero. For a matrix X, let $\mathrm{Tr}(X)$, $X^\top$ and $\rho(X)$ represent the trace, the transpose and the spectral radius of X, respectively. The identity matrix is I, and its size is determined from the context. For a square matrix $X \in \mathbb{R}^{n\times n}$, $X > (\geq)\ 0$ stands for X is positive (semi-)definite. Let $P(\cdot)$ and $P(\cdot|\cdot)$ stand for the probability and conditional probability, and $\mathbb{E}[\cdot]$ stands for the expectation of a random variable.

2. System setup and problem formulation

Consider the following S independent discrete-time linear time-invariant dynamic processes in Fig. 1:

$$x_{k+1}^{(i)} = A_i x_k^{(i)} + w_k^{(i)},$$

where $i \in N \triangleq \{1, \ldots, S\}$, $x_k^{(i)} \in \mathbb{R}^{n_i}$ is the state of the ith process at time k, A_i is the system matrix, and $w_k^{(i)}$ is the state disturbance noises, which is Gaussian distributed with mean zero and covariance $Q_i \geq 0$. The initial state $x_0^{(i)}$ is also a zero-mean

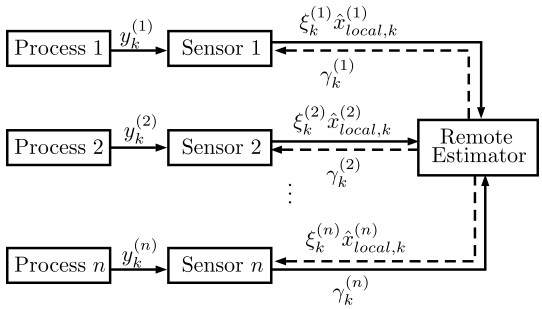


Fig. 1. System architecture.

Gaussian random variable with covariance $\Pi_i > 0$. The states of each process are measured by a sensor with noise:

$$y_k^{(i)} = C_i x_k^{(i)} + v_k^{(i)},$$

where $y_k^{(i)} \in \mathbb{R}^{m_i}$ is the measurement at time k, C_i is the observation matrix, and the measurement noise $v_k^{(i)}$ is zero-mean Gaussian with covariance $R_i > 0$. Moreover, $x_0^{(i)}$, $w_k^{(i)}$, and $v_k^{(i)}$ are uncorrelated with each other. We further assume that $(A_i, \sqrt{Q_i})$ is stabilizable and (A_i, C_i) is detectable for each process.

The sensors are assumed to have sufficient storage and computation capacity. After obtaining the measurement $y_k^{(i)}$, the sensor i runs a Kalman filter to compute the minimum mean square error (MMSE) estimate of $x_k^{(i)}$:

$$\begin{aligned}
\hat{x}_{local,k}^{(i)-} &= A_i \hat{x}_{local,k-1}^{(i)}, \\
P_{local,k}^{(i)-} &= A_i P_{local,k-1}^{(i)} A_i^\top + Q_i, \\
K_{local,k}^{(i)} &= P_{local,k}^{(i)-} C_i^\top (C_i P_{local,k}^{(i)-} C_i^\top + R_i)^{-1}, \\
\hat{x}_{local,k}^{(i)} &= \hat{x}_{local,k}^{(i)-} + K_{local,k}^{(i)} (y_k^{(i)} - C_i \hat{x}_{local,k}^{(i)-}), \\
P_{local,k}^{(i)} &= (I_{n_i} - K_{local,k}^{(i)} C_i) P_{local,k}^{(i)-},
\end{aligned}$$

where $P_{local,k}^{(i)-}$ is the *a priori* estimation error covariance, $P_{local,k}^{(i)}$ is for the *a posteriori* estimation error covariance, and $K_{local,k}^{(i)}$ is the optimal filter gain. The iteration algorithm starts with $\hat{x}_{local,0}^{(i)-} = 0$ and $P_{local,0}^{(i)-} = \Pi_i$. Since we assume that the initial error covariance matrix $\Pi_i > 0$, $\forall i$, $(A_i, \sqrt{Q_i})$ is controllable and (A_i, C_i) is observable for each process, according to Anderson and Moore (1979), the above iteration of the error covariance $P_{local,k}^{(i)}$ converges exponentially to a steady value $\overline{P}^{(i)}$. Without loss of generality, we assume that the local estimation covariances are in the steady state.

As the bandwidth of the wireless communication channel would be limited, only L out of the S sensors can transmit their local estimates, $\hat{x}_{local,k}^{(i)}$, to the remote estimator at each time. Let $\gamma_k^{(i)} \in \{0, 1\}$ denote the data request decision by the remote estimator whether the ith sensor transmits its estimate at time k. If the remote estimator ask sensor i for data at time k, $\gamma_k^{(i)} = 1$; otherwise, $\gamma_k^{(i)} = 0$. Let $\theta = \{\gamma_k^{(i)} : i = 1, 2, \ldots, S; k \geq 0\}$ be the scheduling policy which allocates decision variable, $\gamma_k^{(i)}$, of the remote estimation system. In this work, we assume that the request can be received by the sensors via feedback channels with probability one.

The packet transmission from the sensor to the remote estimator is lossy as the transmission power of the sensors are limited. We assume that the packet arrival follows a time-homogeneous Bernoulli process. Let $\xi_k^{(i)} = 1$ denote that the transmission at time k is successful for the ith sensor and $\xi_k^{(i)} = 0$ otherwise. The arrival rate is $\mathbb{E}[\xi_k^{(i)}] = \lambda_i$, $\forall k \geq 0$. The packet of the ith sensor can be received by the remote estimator at time k if $\gamma_k^{(i)} \xi_k^{(i)} = 1$.

Furthermore, as the scales, n_i, of each dynamic process are different, so are the data packet lengths of the local estimates to be transmitted. To decode the transmitted local estimates, $\hat{x}_{local,k}^{(i)}$, the remote estimator should receive all the relevant packets of the estimate. As the local estimate of a higher dimension state may be split into more pieces than a lower dimension process, it may take more time steps to transmit the whole packet for the higher dimensional process. Let d_i denote the total time steps for the ith sensor to transmit its each local estimate. For example, suppose there are two processes. One of them is a scalar process, the dimension of which is one, and the dimension of the other is two. Accordingly, the packet lengths of the two processes are $d_1 = 1$ and $d_2 = 2$, respectively. As a result, it takes one time step for the first process to transmit its local estimate, and two time steps for the second one.

Let $\eta_{k,\ell}^{(i)} = 1$ stand for the arrival of the sensor i's local estimate of time ℓ, $\hat{x}_{local,l}^{(i)}$, at time k (note that it may take more than one step to transmit a single local estimate) and $\eta_{k,\ell}^{(i)} = 0$ otherwise. At the remote estimator, define the time elapsed from the last complete transmission of the local estimate for the ith sensor at time k as

$$\tau_k^{(i)} = k - \max_{k'}\{\ell : \eta_{k',\ell}^{(i)} = 1, 0 \leq k' \leq k\}. \tag{1}$$

Based on the above settings, the remote estimator updates its estimation of the states as follows:

$$\hat{x}_k^{(i)} = \begin{cases} A_i^{\tau_k^{(i)}} \hat{x}_{local,k-\tau_k^{(i)}}^{(i)}, & \text{if } \eta_{k,\ell}^{(i)} = 1, \\ A_i \hat{x}_{k-1}^{(i)}, & \text{if } \eta_{k,\ell}^{(i)} = 0. \end{cases}$$

As the local error covariances are assumed to be in their steady states, the estimation covariances of the remote estimator are as follows:

$$P_k^{(i)} = \begin{cases} h_i^{\tau_k^{(i)}}(\overline{P}^{(i)}), & \text{if } \eta_k^{(i)} = 1, \\ h_i(P_{k-1}^{(i)}), & \text{if } \eta_k^{(i)} = 0, \end{cases}$$

where the affine mapping of symmetric matrices $h_i^\ell(\cdot)$ and $h_i(\cdot)$ are defined as

$$\begin{aligned}
h_i^0(X) &\triangleq X, \\
h_i(X) &\triangleq A_i X A_i^\top + Q_i, \\
h_i^\ell(X) &\triangleq \underbrace{h_i \circ \cdots \circ h_i}_{\ell}(X) = A_i^\ell X (A_i^\top)^\ell + \sum_{t=0}^{\ell-1} A_i^t Q_i (A_i^\top)^t,
\end{aligned}$$

where $\circ$ denotes a function composition. The following properties of $h_i(\cdot)$ will be useful in later sections.

Lemma 1 (*Lemma 3.1 in Shi and Zhang (2012)*). *The Lyapunov-like operator $h_i^\ell(X)$ is monotonic with respect to ℓ, i.e., $\forall i \in N$, if $\ell_1 \leq \ell_2$ for $\ell_1, \ell_2 \in \mathbb{Z}_+$, $h_i^{\ell_1}(\overline{P}^{(i)}) \leq h_i^{\ell_2}(\overline{P}^{(i)})$. Consequently, $\forall \ell \in \mathbb{Z}_+$, $\mathrm{Tr}(\overline{P}^{(i)}) < \mathrm{Tr}(h(\overline{P}^{(i)})) < \cdots < \mathrm{Tr}(h^\ell(\overline{P}^{(i)}))$.*

Given any initial states $(\tau_0^{(1)}, \ldots, \tau_0^{(n)})$ of the system, the average per-stage cost of a scheduling policy is defined as

$$J\left(\theta, (\tau_0^{(1)}, \ldots, \tau_0^{(n)})\right) \triangleq \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\sum_{k=0}^{T-1} \sum_{i=1}^{S} \mathrm{Tr}(P_k^{(i)})\right].$$

With the above definition, the optimal scheduling policy is a feasible policy minimizing the total cost:

Problem 1.

$$\min_{\theta} J\Big(\theta, (\tau_0^{(1)}, \ldots, \tau_0^{(S)})\Big)$$

$$s.t. \sum_{i=1}^{S} \gamma_k^{(i)} = L, \ \forall k \geq 0.$$

This scheduling index considered stands for the asymptotic level of uncertainty in the remote state estimate, which is the main focus in many works on sensor scheduling problems, e.g., Han et al. (2017), Hovareshti et al. (2007), Mo, Garone et al. (2014) and Zhao et al. (2014). The following questions relating to the optimal scheduling policy are of interest.

(1) If a transmission fails, is it optimal to continue with the in-transmission estimate or start the most recent estimate? Moreover, if a transmission fails, is it optimal to keep scheduling the current selected sensor or others?
(2) Is there a deterministic and stationary optimal policy?
(3) If there is a deterministic and stationary optimal policy, are there any special structures of the optimal policy?

Remark 1. In this work, the scheduling policy is implemented in a centralized manner and the information available to the central controller is based on the time elapsed since last completed transmission. It would be interesting to consider event-triggered protocol (e.g. Ding et al. (2017)), where the transmission decisions are decentralized and based on information of the current state, with the packet length constraint.

3. Main results

This section focuses on solving Problem 1. We formulate it as an infinite time horizon Markov decision process with average cost criterion. We first study the some properties of the case when $S = 1$. We show that the optimal policy always continues the in-transmission packets until completion and derive the condition of boundedness of the estimation error covariance. We then prove, for the general S, that there exists a deterministic and stationary (independent of time index k) optimal scheduling policy. We further show that the optimal policy has a consistency and switching structure. By exploiting the consistency property and a switching structure, we develop a stochastic algorithm to compute this switching type policy with reduced computation overhead.

3.1. MDP formulation

To simplify notations, the following discussions are restricted to the case when $S = 2$ and $L = 1$. In Remark 3, we show that the structural results can be extended to general S.

To represent Problem 1 as a discrete time Markov decision process, we define the following quadruplet $(\mathbb{S}, \mathbb{A}, P(\cdot|\cdot,\cdot), c(\cdot,\cdot))$. Each item is elaborated as follows.

(1) The state space $\mathbb{S}$ at time $k \geq 0$ is defined as $s_k \triangleq (\tau_k^{(1)}, \tau_k^{(2)}, q_k^{(1)}, q_k^{(2)}, r_k^{(1)}, r_k^{(2)}) \in \mathbb{N}^6$. The estimate holding time $\tau_k^{(i)}$ is defined in (1). The packet quality index $q_k^{(i)} \in \mathbb{N}$ indicates the quality of the current in-transmission estimate. In other words, $q_k^{(i)}$ stands for the time elapsed since starting transmitting the in-transmission packet for process i. If the transmission of an estimate of sensor i is complete at time $k+1$, $\tau_{k+1}^{(i)} = q_k^{(i)}$. The remaining packet number $r_k^{(i)} \in \{1, 2, \ldots, d_i\}$ stands for how many packets are left to be sent to complete the transmission of the estimate of the ith sensor. For example, assume the estimate of the first sensor consists of three packets. Then $r_k^{(1)} = 2$ means that one packet has been transmitted to the remote estimator and there are still two packets to be transmitted.

(2) The action a_k is in the action space $\mathbb{A} \triangleq \mathbb{I} \times \mathbb{B}$, where $\mathbb{I} \triangleq \{1, 2\}$ and $\mathbb{B} \triangleq \{0, 1\}$. An action is denoted by $a_k = (i, b)$, where i stands for which sensor to be scheduled and $b = 0$ stands for start transmitting $\hat{x}_{local,k}^{(i)}$ and $b = 1$ for continuing the current in-transmission estimate. If no estimate is in-transmission, $b = 1$ and $b = 0$ lead to same results. We later show that an optimal policy never allows $b = 0$ and the action space $\mathbb{A}$ can be reduced to $\mathbb{I}$.

(3) The state transition, $P(s'|s, a)$, characterizes the state transition probability to s' when current state is s and action is a. As the transition is homogeneous, we write $s = (\tau_1, \tau_2, q_1, q_2, r_1, r_2)$ and $s' = (\tau_1', \tau_2', q_1', q_2', r_1', r_2')$ by dropping the time index. The state transition law is as follows

$$P(s'|s, a = (i, \cdot)) = \begin{cases} \lambda_i, & \text{if } a = (i, b),\ s' = s_s, \\ 1 - \lambda_i, & \text{if } a = (i, b),\ s' = s_f, \\ 0, & \text{others}, \end{cases}$$

where the next state $s' = s_s$ and $s' = s_f$ indicates whether a transmission succeeded or failed. We elaborate the notation by the following cases.

(i) If $a = (1, 1)$ and the transmission is successful, i.e., $\gamma_k^{(1)} \xi_k^{(1)} = 1$, the next state is

$$s_s = \begin{cases} (0, \tau_2 + 1, 0, 0, d_1, d_2), \\ \quad \text{if } d_1 = 1, \\ (\tau_1 + 1, \tau_2 + 1, 1, 0, d_1 - 1, d_2), \\ \quad \text{if } d_1 > 1 \text{ and } r_1 = 0, \\ (\tau_1 + 1, \tau_2 + 1, q_1 + 1, 0, r_1 - 1, d_2), \\ \quad \text{if } d_1 > 1 \text{ and } r_1 - 1 > 0, \\ (q_1, \tau_2 + 1, 0, 0, d_1, d_2), \\ \quad \text{if } d_1 > 1 \text{ and } r_1 - 1 = 0. \end{cases} \tag{2}$$

If the transmission fails, i.e., $\gamma_k^{(1)} \xi_k^{(1)} = 0$, the next state is $s_f = (\tau_1 + 1, \tau_2 + 1, q_1 + 1, 0, r_1, d_2)$.

(ii) If $a = (1, 0)$ and the transmission is successful, i.e., $\gamma_k^{(1)} \xi_k^{(1)} = 1$, the next state is

$$s_s = \begin{cases} (0, \tau_2 + 1, 0, 0, d_1, d_2), \\ \quad \text{if } d_1 = 1, \\ (\tau_1 + 1, \tau_2 + 1, 1, 0, d_1 - 1, d_2), \\ \quad \text{if } d_1 > 1 \text{ and } r_1 = 0. \end{cases} \tag{3}$$

If the transmission fails, i.e., $\gamma_k^{(1)} \xi_k^{(1)} = 0$, the next state is $s_f = (\tau_1 + 1, \tau_2 + 1, 0, 0, d_1, d_2)$.

(iii) and (iv) The case for $a = (2, 1)$ and $(2, 0)$ are similar to (i) and (ii) and omitted for brevity.

(4) The one-stage cost, independent of action, is a function of the current state

$$c(s, a) \triangleq \mathrm{Tr}(h_1^{\tau_1}(\overline{P}^{(1)})) + \mathrm{Tr}(h_2^{\tau_2}(\overline{P}^{(2)})).$$

Problem 1 can be considered as a stochastic optimal control problem. The MDP framework can completely represent the problem as long as we correctly identify the variables with the elements in an MDP model. As we can see that the state transition defined in our MDP model is Markovian, and this completely characterizes the dynamic system considered in this work.

Define $H_k = (s_0, a_0, \ldots, s_{k-1}, a_{k-1}, s_k)$ be the history of the states and actions up to time k, and $\mathbb{H}_k$ be the class of all possible histories. Moreover, denote $\pi = (\pi_1, \ldots, \pi_k, \ldots)$ as a feasible policy, where π_k is a stochastic kernel from $H_k \in \mathbb{H}_k$ to $\mathbb{A}$.[1] Let Π be the set of all such feasible policies. The long term

[1] A function $\pi_k(\cdot|\cdot)$ is stochastic kernel from H_k to $\mathbb{A}$ if (a) $\pi_k(\cdot|H_k)$ is a probability distribution on $\mathbb{A}$ for fixed $H_k \in \mathbb{H}_k$ and (b) $\pi_k(A_k|\cdot)$ is a measurable function on the σ-algebra generated by $\mathbb{H}_k$ for fixed A_k, which is an element in the σ-algebra generated by $\mathbb{A}$.

average cost associated with a policy π and an initial state $s_0 = (\tau_0^{(1)}, \tau_0^{(2)}, 0, 0, d_1, d_2)$ is defined to be

$$J(\pi, s_0) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_{s_0}^{\pi}\left[\sum_{k=0}^{T-1} c(s_k, \pi_k)\right].$$

One may see that the Problem 1 is equivalent to the following problem.

Problem 2. Find an optimal policy $\pi^* \in \Pi$ to minimize the long term average cost:

$$J(\pi^*, s_0) = \inf_{\pi\in\Pi} J(\pi, s_0).$$

We are interested in the set of deterministic and stationary policies Π^{DS}. A policy π is deterministic if, given the history H_k, the decision is concentrated at one action. A policy π is stationary if the stochastic kernel π_k time-invariant and depends only on the current state. If there exists such an optimal policy, it can be obtained from the following average cost optimal equation (ACOE) (Hernández-Lerma & Lasserre, 1996)

$$\rho^* + V(s) = \min_{\pi\in\Pi^{DS}}\left[c(s, \pi(s)) + \mathbb{E}_s^{\pi}[V]\right], \tag{4}$$

where ρ^* is a constant, $V(\cdot)$ is a function on $\mathbb{S}$ and $\mathbb{E}_s^{\pi}[V]$ is the conditional expectation of the value of the next state under policy π, i.e., $\mathbb{E}_s^{\pi}[V] = \sum_{s'\in\mathbb{S}} V(s')P(s'|s,\pi)$, which reduces to $V(s')$ for $\lambda_i = 1$. Note that ρ^* is the optimal cost, i.e., $\rho^* = \min_{\pi\in\Pi} J(\pi, s_0)$, and $V(s)$ is the relative value function, which is unique up to a shift, i.e., if both $V(s)$ and $W(s)$ satisfy the ACOE, then there exists a constant C such that $W(s) - V(s) = C, \forall s \in \mathbb{S}$.

3.2. Single process analysis

We first study the case when $S = 1$. The following issues are to be addressed. (1) Is it optimal to continue with the in-transmission estimate or start transmitting a new estimate? (2) What is the condition for boundedness of expectation of the average estimation error covariance?

We first show that it is necessary for the optimal policy, if exists uniquely (which will be proven later), to continue in-transmission estimate. Next, we show the condition of the boundedness of the estimation error covariance under this type policy. As we only analyze one process, the process index i is omitted. A state is represented by $s = (\tau, q, r)$, where τ stands for estimate holding time, q for packet quality, and r for number of remaining packets. Moreover, the action space consists only two elements, starting new estimates $a = (1, 0)$ or continuing the in-transmission estimate $a = (1, 1)$.

The following proposition states that it is necessary for an optimal policy to continue the in-transmission estimate.

Theorem 1. *If $r > 1$, $\pi^*\Big((\tau, q, r)\Big) = (1, 1)$.*

Proof. If $d = 1$, start a new transmission is the same as continue in-transmission in each time step. Therefore, we consider $d > 1$. Note that if the first packet of the new estimate is transmitted, the state will only transit from (τ, q, r) to $(\tau + 1, 1, d - 1)$ with probability λ and $(\tau + 1, 0, d)$ with probability $(1 - \lambda)$. By the ACOE (4), this statement is equivalent to

$$\begin{aligned}&\lambda V((\tau+1, q+1, r-1)) + (1-\lambda)V((\tau+1, q+1, r))\\ &\le \lambda V((\tau+1, 1, d-1)) + (1-\lambda)V((\tau+1, 0, d)),\ \forall r < d,\end{aligned} \tag{5}$$

if $r - 1 > 0$, and

$$\begin{aligned}&\lambda V((q, 0, d)) + (1-\lambda)V((\tau+1, q+1, r))\\ &\le \lambda V((\tau+1, 1, d-1)) + (1-\lambda)V((\tau+1, 0, d)),\ \forall r < d,\end{aligned} \tag{6}$$

if $r - 1 > 0$. To meet these two inequalities, it suffices to prove $V((\tau+1, q+1, r-1)) \le V((\tau+1, 1, d-1))$, $V((\tau+1, q+1, r)) \le V((\tau + 1, 0, d))$ and $V((q, 0, d)) \le V((\tau + 1, 1, d - 1))$. These inequalities hold if the relative value function is

(1) monotonic increasing with respect to τ;
(2) monotonic decreasing with respect to q;
(3) monotonic increasing with respect to r.

These can be proved by induction because the solution to the ACOE is unique under some conditions (B1, B2 after Theorem 2). Specifically, The ACOE can be compactly written as

$$\rho^\star + V(s) = \mathcal{T}[V(s)],$$

where $\mathcal{T}[V(s)] = \min_{\pi}\Big[c(s, \pi(s)) + \mathbb{E}_s^{\pi}[V]\Big]$. We assume that the above condition holds, and we verify if the operator $\mathcal{T}$ preserves such structure.[2]

By the induction assumption, we can obtain $V(\tau, 0, d) \ge V(\tau, 1, d - 1) \ge V(\tau, q, r)$ for any $q \ge 1$ and $r \le d - 1$. Therefore, it suffices to verify that continuing in-transmission preserve the above monotonicity. The three monotone properties can be proven using identical method. We only show the proof for monotonicity with respect to τ. Let $\tau \ge \tau'$. If $r - 1 > 0$,

$$\begin{aligned}&V((\tau, q, r)) - V((\tau', q, r))\\ &= \lambda[V((\tau+1, q+1, r-1)) - V((\tau'+1, q+1, r-1))]\\ &\quad + (1-\lambda)[V((\tau+1, q+1, r)) - V((\tau'+1, q+1, r))]\\ &\ge 0,\end{aligned}$$

where the inequality is due to the induction assumption. If $r - 1 = 0$, we can also obtain

$$\begin{aligned}&V((\tau, q, r)) - V((\tau', q, r))\\ &= (1-\lambda)[V((\tau+1, q+1, r)) - V((\tau'+1, q+1, r))]\\ &\ge 0.\end{aligned}$$

As the monotonicity of the relative value function is proven, the proof is complete. ■

The optimal scheduling policy is then as follows. (1) Start transmitting a new estimate if there is no in-transmission estimate. (2) Continue the in-transmission if there is one. This result reveals that the action space can be reduced to $\mathbb{I}$ without losing optimality. These can be written compactly as

$$\pi^*(s) = \begin{cases}(1, 0), & \text{if } q = 0 \text{ and } r = d,\\ (1, 1), & \text{if } q > 0 \text{ and } 0 < r < d.\end{cases} \tag{7}$$

We are now ready to derive the stability condition for $S = 1$. The stability metric is chosen as the boundedness of the average of the trace of the estimation error covariance. The average estimation error can be obtained by computing the average between two completed transmissions of two estimates. When a completed transmission occurs, the time elapsed is $q = d + F$, where d is the packet length and F is transmission failure times. As the transmission of one single packet follows a Bernoulli distribution, the random variable F follows the negative binomial distribution

$$P(F = k) = \binom{k+d-1}{d-1}\lambda^d(1-\lambda)^k. \tag{8}$$

The expectation of the average estimation error covariance is a function of two binomial distribution and its boundedness condition is given by the following theorem.

[2] Note that the induction is not on the value function but on the operator $\mathcal{T}$. In fact, this idea was also used in Leong, Dey, and Quevedo (2017) and Zhou, Cui, and Tao (2017).

Theorem 2. *If $(1-\lambda)\rho^2(A) < 1$, the average estimation error covariance at the remote estimator is bounded under the policy (7).*

Proof. Since for $\rho(A) < 1$, the boundedness is straightforward. In the following proof, we only discuss the case when $\rho(A) \geq 1$. Note that $\mathrm{Tr}(h^t(\overline{P})) = \mathrm{Tr}(A^t\overline{P}(A^\top)^t) + \sum_{k=0}^{t-1}\mathrm{Tr}(A^kQ(A^\top)^k) \leq \|\overline{P}\|_2\rho^{2t}(A) + \|Q\|_2\frac{1-\rho^{2t}(A)}{1-\rho^2(A)} \leq C\rho^{2t}(A)$ for some constant C. Therefore, it suffices to prove the time average of $(\rho^2(A))^\tau$, where τ is the random variable standing for the time elapsed since last completely received estimate.

If the average cost is bounded, the associated Markov chain is positive recurrent and hence ergodic. Therefore, we can compute the long term average of $(\rho^2(A))^\tau$ by computing the average of it between two completely received estimate.

Recall that it takes $F + d$ steps to completely transmit an estimate, where F is the times of failures. Let F_1 and F_2 denote two neighboring times of failures of the completely transmitted estimate. We can compute the expected time average of $(\rho^2(A))^\tau$ by

$$\mathbb{E}\Big[\frac{1}{F_2+d}\sum_{t=F_1+d-1}^{F_1+F_2+2d-2}\rho^{2t}(A)\Big],$$

where the expectation is taken with respect to F_1 and F_2.

Note that the binomial coefficient is upper bounded by

$$\binom{k+d-1}{d-1} = \frac{\prod_{d-2}^{j=0}(k+d-1-j)}{(d-1)!} \leq \frac{(k+d-1)^{d-1}}{(d-1)!}.$$

We can compute

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{F_2+d}\sum_{t=F_1+d-1}^{F_1+F_2+2d-2}\rho^{2t}(A)\Big]\\
&\leq \sum_{F_1=0}^{\infty}\sum_{F_2=0}^{\infty}\Bigg[\Bigg(\frac{1}{F_2+d}\sum_{t=F_1+d-1}^{F_1+F_2+2d-2}\rho^{2t}(A)\Bigg)\\
&\quad\Bigg(\frac{(F_1+d-1)^{d-1}}{(d-1)!}\frac{(F_2+d-1)^{d-1}}{(d-1)!}\lambda^{2d}(1-\lambda)^{F_1+F_2}\Bigg)\Bigg]\\
&= \frac{\lambda^{2d}}{[(d-1)!]^2}\sum_{F_1=0}^{\infty}\sum_{F_2=0}^{\infty}\Bigg[\frac{[(F_1+d-1)(F_2+d-1)]^{d-1}}{F_2+d}\\
&\quad\Bigg(\sum_{t=F_1+d-1}^{F_1+F_2+2d-2}(1-\lambda)^{F_1+F_2}\rho^{2t}(A)\Bigg)\Bigg].
\end{aligned}$$

We can compute

$$\begin{aligned}
&\sum_{t=F_1+d-1}^{F_1+F_2+2d-2}(1-\lambda)^{F_1+F_2}\rho^{2t}(A)\\
&= (1-\lambda)^{F_1+F_2}\frac{\rho^{2(F_1+d-1)}(A)(\rho^{2(F_2+d)}(A)-1)}{\rho^2(A)-1}\\
&\leq \frac{(1-\lambda)^{F_1+F_2}\rho^{2(F_1+F_2+2d)}(A)}{\rho^2(A)-1}\\
&= \frac{\rho^{4d}(A)}{\rho^2(A)-1}\Big((1-\lambda)\rho^2(A)\Big)^{F_1+F_2}
\end{aligned}$$

which is an exponential function of F_1 and F_2 with $(1-\lambda)\rho^2(A)$ as its basis. Note that

$$\frac{[(F_1+d-1)(F_2+d-1)]^{d-1}}{F_2+d}$$

is a power function of F_1 and F_2. By ratio test (Rudin, 1964, pp. 65), the convergence of the summation with respect to F_1 and F_2 only relies on the growing rate of the exponential function. Therefore, the right hand side is bounded if $(1-\lambda)\rho^2(A) < 1$. This completes the proof. ■

This result is same as the stability condition when $d = 1$ in Schenato (2008). This result reveals that the lengths of the packets do not affect the stability condition. We can prove the existence of an optimal deterministic and stationary policy satisfies the ACOE (4) by verifying the following two conditions (Sennott, 1996).

(B1) There exists a stationary policy π inducing an irreducible and positive recurrent Markov chain on the state space and satisfying the average cost J_π is finite.

(B2) There exists $\epsilon > 0$ such that $B = \{s|\exists a \text{ s.t. } c(s,a) < J_\pi + \epsilon\}$ is finite.

As the one-stage cost is monotone with respect to τ, the positive recurrence condition is satisfied if the average cost is bounded. This is because we can choose the one-stage cost as the Lyapunov function to satisfy the Foster drift condition (Meyn & Tweedie, 2012) for positive recurrence. The average cost of the single process under the "always continues in-transmission" policy is bounded if $(1-\lambda)\rho^2(A) < 1$ as shown in the last theorem. Therefore, the first condition is thus verified. The second condition is trivial as the one-stage cost is bounded below and monotone increasing with respect to τ.

3.3. Structural policy for multiple processes

We now study the case for multiple processes. We first show that the optimal policy for the MDP is deterministic and stationary. After that, two structural results, the consistency (Theorem 4) and the switching structure (Theorem 5), of the optimal policy are given.

The existence of a deterministic and stationary optimal policy is as follows.

Theorem 3 (*Existence*). *If $\max_i\{\rho^2(A_i)\}\max_i\{1-\lambda_i\} < 1$, there exists a constant ρ^*, a function $V(\cdot)$ on $\mathbb{S}$ and a deterministic and stationary policy $\pi^* \in \Pi^{DS}$ that satisfies the ACOE (4).*

Proof. The proof still relies on verifying (B1) and (B2). By Theorem 2, the boundedness of the average estimation error covariance is independent of the packet length under the policy which always continue in-transmission. Moreover, by Mesquita, Hespanha, and Nair (2012), the boundedness is also independent of the number of the processes under a L-drop triggered protocol. Therefore, the condition (B1) is verified. The condition (B2) is trivial as the one-stage cost is bounded below and monotonic increasing. ■

Remark 2. The boundedness condition may be anticipated to be $\max_i\{\rho^2(A_i)(1-\lambda_i)\} < 1$. However, this boundedness condition for every single process, is necessary but not sufficient. Our proposed condition provides a uniform bound, which is sufficient to guarantee the boundedness.

In fact, the stability of scheduling multiple processes relies on the transmission protocol we use. Therefore, we can only provide a sufficient condition. Meanwhile, the existence of a stationary and deterministic optimal policy depends on whether it is feasible to stabilize the error covariance. The relation between the stability condition and the optimal scheduling policy is intertwined. Therefore, we can only relax the stability condition and present a sufficient one.

The consistency property of the optimal policy is as follows.

Theorem 4 (*Consistency*). *If $r_i < d_i$, then $a^* = i$ for $i = 1, 2$.*

Theorem 4 is consistent with intuition. Note that only when the last scheduling action is scheduling sensor i can $r_i < d_i$ happens. Assume only one sensor is allowed to schedule in each time step, Theorem 4 implies that once a sensor is chosen to schedule and the first transmission is successful, another sensor can be chosen to schedule only after the incomplete transmission is finished. In other words, if the first packet is successfully delivered, the transmission of one estimate should be consecutively executed until the packets have been all delivered.

By combining Theorems 1 and 4, the optimal scheduling policy can be summarized as follows. If the transmission of one estimate is complete, select one sensor to schedule. If the first packet of an estimate is successfully received, the remote estimator will keep scheduling the same sensor until the estimate is complete. Otherwise, repeat selecting which sensor to schedule.

Next, we show the switching type policy of selecting the sensors. The switching policy is important from two perspectives. First, we can develop efficient algorithms to obtain the optimal policy by utilizing the structure as we will discuss in the next subsection. Second, the switching policy can facilitate online implementation of the scheduling policy. Note that we only consider the case when $r_1 = r_2 = 1$. As it has been shown that the scheduling policy should be consistent, the decision of which sensor to schedule is only made when the packet(s) of one local estimate have been all transmitted to the remote estimator.

Theorem 5. *Switching type policy.*

(1) *If* $\pi^*((\tau_1, \tau_2, 0, 0, d_1, d_2)) = 1$, *then* $\pi^*((\tau_1 + z, \tau_2, 0, 0, d_1, d_2)) = 1$, *where* z *is any positive integer;*
(2) *If* $\pi^*((\tau_1, \tau_2, 0, 0, d_1, d_2)) = 2$, *then* $\pi^*((\tau_1, \tau_2 + z, 0, 0, d_1, d_2)) = 2$, *where* z *is any positive integer.*

Remark 3. The switching structure can be extended to general S. Let $r_i = 1, \forall i \in [1, S]$. For $1 \le i \le n$, define $\tau_i^- = (\tau_1, \ldots, \tau_{i-1}, \tau_{i+1}, \ldots, \tau_n)$. There exists measurable functions $\phi_i : \mathbb{N}^{n-1} \mapsto \mathbb{N}$ and the optimal policy is as follows:

(1) Choose the ith sensor if $\phi_i(\tau_i^-) \le \tau_i$;
(2) Do not choose the ith sensor if $\phi_i(\tau_i^-) > \tau_i$.

Since these statements hold for all sensors, the functions $\phi_i()$'s are nondecreasing with respect to each of its component.

Note that the implementation of the optimal policy for a perfect channel is different from a lossy channel. If there are no packet drops, the transmission sequence can be allocated beforehand. If we take packet drop into consideration, the implementation requires an update of system states for the remote estimator and the sensors. One possible implementation is done by letting the remote estimator request the target sensors to send their local estimates. The channel from the remote estimator to the sensor is more reliable than the channel in the opposite direction as the remote estimator are capable of sending information with greater power.[3]

The switching policy saves storage space for online implementation. Rather than storing the action on each state of the Markov chain, one only needs to store the states on the switching boundary for implementing the scheduling scheme. When it is time to schedule a sensor, only the comparison between the current state and the boundary is needed. This reduces the storage space required in online implementation. Furthermore, if the optimal policy has a specific structure, special algorithms can be developed to reduce off-line computation. In the following, we shall exploit this feature by devising an algorithm for finding the optimal switching curve.

[3] The packet arrival acknowledgments from the remote estimator to the sensors are assumed to be reliable, which is common assumption in the literature, e.g., Mo, Chabukswar, and Sinopoli (2014), Shi et al. (2014) and You and Xie (2011).

3.4. Computation overhead reduction

We have found that the optimal policy has several nice structures. The dependence of the optimal policy and the total cost, however, is inexplicit and it is thus impossible to analytically calculate the switching policy. We need to resort to numerical methods to obtain a specific policy. Fortunately, with the structural policy, the off-line computation overhead of the optimal policy can be reduced. Theorem 1 helps reduce the action space and Theorem 4 helps reduce the states whose values are to be computed, and the switching structure reduces the policies to be evaluated.

Note that the consistency structure (Theorem 4) can be utilized to reduce the number of states. In a relative value iteration algorithm (Puterman, 1994, pp. 373), the value of all states need to be computed. Because of consistency of the optimal policy, the policy on the states whose last two components are greater than one is determined by the policy when the corresponding components are both one. As a result, the state space can be reduced by a factor of $\prod_{i=1}^{S} d_i$. In the new state space, we can conduct either policy iteration or value iteration to compute the optimal policy.

Meanwhile, the search space of the policies can also be reduced by utilizing the switching structure. In the relative value iteration, all feasible policies are evaluated. Inspired by Nourian et al. (2014), we develop the following simultaneous perturbation stochastic approximation (SPSA) algorithm to accelerate each iteration by only considering switching type policies. For the kth iteration, define the value function as follows

$$J^{(k)} \triangleq c(s, \pi(s)) + \mathbb{E}_s^{\pi}[V].$$

Then the steps in Algorithm 1 are carried in the kth iteration of the value iteration.

Algorithm 1 Switching Type SPSA (Spall, 2005) Based Value Iteration for Sensor Scheduling

---

Choose initial switching policy $\hat{\vartheta}_0$
repeat
In the nth iteration of $J^{(k)}$,
1. generate simultaneous perturbation vector:

$$P(\delta_n = k) = \begin{cases} 0.5, & k = -1, \\ 0.5 & k = 1; \end{cases}$$

2. approximate the gradient:

$$\hat{\nabla}_{\hat{\vartheta}} J_n^{(k)} = \frac{J_n^{(k)}(\hat{\vartheta}_n + \mu_n \delta_n) - J_n^{(k)}(\hat{\vartheta}_n - \mu_n \delta_n)}{2\mu_n} \delta_n,$$

where $\mu_n = \mu/(1+n)^{\gamma}$;
3. update the switching curve estimate:

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n - \varepsilon_{n+1} \hat{\nabla}_{\hat{\vartheta}} J_n^{(k)}, \tag{9}$$

where $\varepsilon_n = \varepsilon/(n+1+\epsilon)^{\kappa}$. **until** convergence

---

In the algorithm, the parameters $\mu, \gamma, \varepsilon, \epsilon$ and κ may be determined according to empirical guidelines provided in Section 7.5 in Spall (2005). Typically, $\gamma = 0.602$ and $\kappa = 0.101$. By properly choosing parameters, Algorithm 1 converges and we have the following proposition.

Proposition 1 (*Theorem 7.1 in Spall (2005)*)**.** *The estimate sequence* ε_n *of the switching policy generated by (9) converges as* $n \to \infty$ *to a local minimum of* $J^{(k)}$ *almost surely.*

Note that since the algorithm converges to a local minimum, one needs to try several different initial switching policies, and pick the best solution.

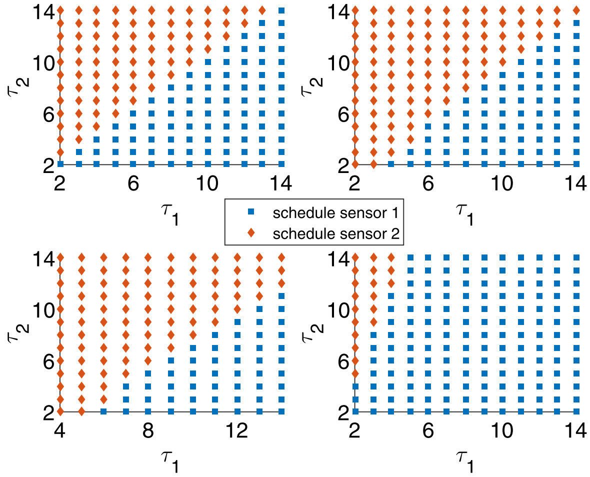


Fig. 2. Visualization of the switching type policy for $n = 2$.

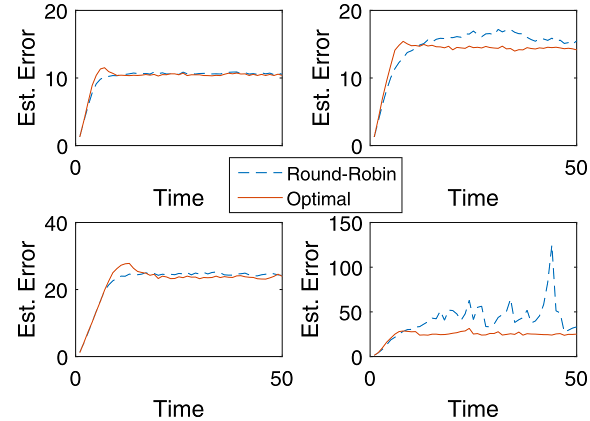


Fig. 3. Performance comparison of the optimal policy with Round-Robin protocol for $n = 2$.

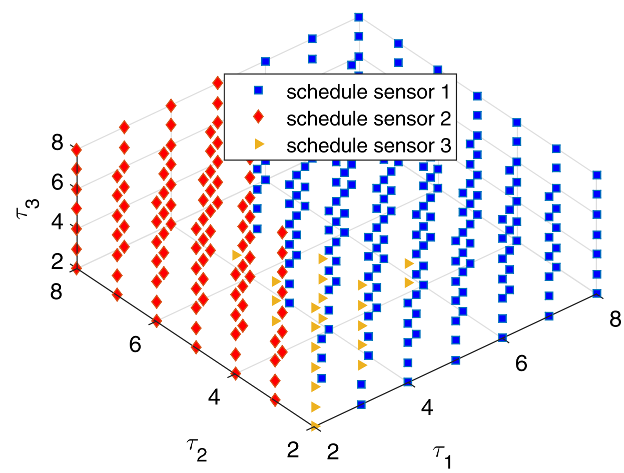


Fig. 4. Visualization of the switching type policy for $n = 3$.

4. Numerical examples

In this section, we provide two numerical simulations to illustrate our results on the switching curve type policy. In each example, only one process is allowed to schedule in each time step.

In the first example, we show the effect on the geometry of the switching curve due to system parameters. There are two processes. The first case depicts the benchmark system dynamics, which is as follows.

$$x_{k+1}^{(1)} = 1.1x_k^{(1)} + w_k^{(1)}, y_k^{(1)} = x_k^{(1)} + v_k^{(1)},$$
$$x_{k+1}^{(2)} = 1.1x_k^{(2)} + w_k^{(2)}, y_k^{(2)} = x_k^{(2)} + v_k^{(2)},$$

where $\mathbb{E}[(w_k^{(i)})^2] = \mathbb{E}[(v_k^{(i)})^2] = 1, i = 1, 2$. The benchmark packet arrival rate and packet length are 0.9 and two for both processes, respectively. The switching curve of the benchmark case is presented in the left upper corner of Fig. 2. It can be seen that the switching curve went through the straight line $\tau_1 = \tau_2$. The second case (on the right upper corner of Fig. 2) presents the result when the packet arrival rate of the first process decreases from 0.9 to 0.7. The third case (on the left lower corner of Fig. 2) shows the result when the packet length of the first process increases from two to four. The last case (on the right lower corner of Fig. 2) illustrates the result when the eigenvalue of the system matrix of the first processes deviates increases from 1.1 to 1.4. We can observe that the optimal scheduling policy "prefers" processes with high packet arrival rates, shorter packet lengths and greater spectral radii of their system matrices.

To investigate the actual performance, we run our policy with the above example and compare its performance in terms of the average estimation error covariance with a revised Round-Robin protocol. Both policies follow the consistency property of the optimal policy (Theorem 4). The only difference lies in how to select which sensor to schedule after the last completed transmission. The sensor completing the last transmission passes its transmission token to the other one, while the optimal policy determines which sensor to schedule according to the switching curve. We ran 1000 Monte Carlo simulations, and the comparisons of average performance in terms of the scheduling index of the two protocols in the four cases described above are shown in Fig. 3. It can be seen that the optimal policy determined by the switching curve outperforms the revised Round-Robin protocol.

In the second example, there are three dynamic processes as follows

$$x_{k+1}^{(1)} = 1.3x_k^{(1)} + w_k^{(1)}, y_k^{(1)} = x_k^{(1)} + v_k^{(1)},$$
$$x_{k+1}^{(2)} = 1.2x_k^{(2)} + w_k^{(2)}, y_k^{(2)} = x_k^{(2)} + v_k^{(2)},$$
$$x_{k+1}^{(3)} = 1.1x_k^{(3)} + w_k^{(3)}, y_k^{(3)} = x_k^{(3)} + v_k^{(3)},$$

where $\mathbb{E}[(w_k^{(i)})^2] = \mathbb{E}[(v_k^{(i)})^2] = 1, i = 1, 2, 3$. The packet arrival rates of each channel are $\lambda_1 = 0.7, \lambda_2 = 0.8$ and $\lambda_3 = 0.9$ and the packet length of each process are two, which means it costs two successful transmission to complete transmitting an estimate. The switching curve policy in this case is presented in Fig. 4. It is clear that there are two hypersurfaces splitting the actions.

5. Conclusion

We have considered minimization of the average error covariance over an infinite time horizon by designing an optimal sensor scheduling policy. In this problem, different sensors may occupy different time durations for transmission because the estimates consist of multiple packets and each packet takes one time step to transmit. Moreover, due to a bandwidth constraint, only a portion of all the sensors are allowed to transmit simultaneously. We have formulated the problem as an infinite time horizon Markov decision process with average cost criterion. We have found that there is a deterministic and stationary optimal policy for the problem. Furthermore, we have shown that the optimal policy has consistency and switching curve structures. The consistent behavior implies that once a sensor is chosen to be scheduled and the first transmission succeeds, the transmission of the current estimate should not be interrupted. The consistency, along with the switching type structure, has facilitated online implementation. By exploiting the structural results, we develop a stochastic algorithm to obtain the switching type policy, which reduces computational overhead of the generic value iteration algorithm. Numerical examples have been provided to illustrate our results.

Appendix A. Proof of Theorem 4

Without loss of generality, assume the state moves to $(\tau_1, \tau_2, q_1, 0, r_1, d_2)$ after the optimal action is taken. The state indicates that, the optimal action taken in the last time step is scheduling sensor 1. We need to prove that the optimal action in current state is continuing scheduling sensor 1. Because the one-stage cost only depends on τ_1 and τ_2, this requires

$$\begin{aligned} &\lambda_1 V((\tau_1+1, \tau_2+1, q_1+1, 0, r_1-1, d_2)) \\ &+(1-\lambda_1)V((\tau_1+1, \tau_2+1, q_1+1, 0, r_1, d_2)) \\ &\le \lambda_2 V((\tau_1+1, \tau_2+1, 0, 1, d_1, d_2-1)) \\ &+(1-\lambda_2)V((\tau_1+1, \tau_2+1, 0, 0, d_1, d_2)). \end{aligned} \tag{10}$$

Denote the left hand side of (10) as *LHS* and right hand side as *RHS*. To prove (10), we need to show that

$$\begin{aligned} LHS &\le \lambda_1 V((\tau_1, \tau_2, q_1, 0, r_1, d_2)) \\ &\quad + (1-\lambda_1)V((\tau_1, \tau_2, q_1, 0, r_1+1, d_2)) && (11a) \\ &\le \lambda_2 V((\tau_1, \tau_2, 0, 1, d_1, d_2-1)) \\ &\quad + (1-\lambda_2)V((\tau_1, \tau_2, 0, 0, d_1, d_2)) && (11b) \\ &\le RHS. && (11c) \end{aligned}$$

Similar to the idea we use to prove Theorem 3.4 in our previous work (Wu et al., 2017), we can use the ACOE to expand the right hand side of (11a) and obtain *LHS* being part of the terms in the expansion. Specifically, denote the first term of the right hand side in (11a) as $\lambda_1 V(s)$ and the two terms on the left hand side as $\lambda_1 V(s_s)$ and $(1-\lambda_1)V(s_f)$. We have $V(s) = c(s, a) + \lambda_1 V(s_s) + (1-\lambda_1)V(s_f)$ by setting $\rho^* = 0$ in the ACOE (4). As $c(s, a)$ is positive, we have $V(s) \ge LHS$. By the monotonicity discussed in the proof of Theorem 1, we have $V(s)$ is less than the second term of the right hand side in (11a). Therefore, (11a) holds. Note that only by taking $a = 1$ can the state $(\tau_1, \tau_2, q_1, 0, r_1, d_2)$ occurs. Therefore, (11b) holds. The inequality (11c) is due to monotonicity of the relative value function.

Note that the above discussion assumes $d_2 > 1$ and $r_1 > 1$. If $d_2 = 1$, $RHS = \lambda_2 V((\tau_1+1, 0, 0, 0, d_1, 1)) + (1-\lambda_2)V((\tau_1+1, \tau_2+1, 0, 0, d_1, 1))$ and the right (left) hand side of (11b) (11c) turns out to be $\lambda_2 V((\tau_1, 0, 0, 0, d_1, 1)) + (1-\lambda_2)V((\tau_1, \tau_2, 0, 0, d_1, 1))$. If $r_1 = 1$, *LHS* becomes $\lambda_1 V((q_1, \tau_2+1, 0, 0, d_1, d_2)) + (1-\lambda_1)V((\tau_1+1, \tau_2+1, q_1+1, 0, 1, d_2))$. The remaining part of the above proof can be naturally extended.

Appendix B. Proof of Theorem 5

The statement is equivalent to monotonicity of the optimal policy in τ_1 if τ_2 is fixed and in τ_2 if τ_1 is fixed. We prove the monotonicity by showing that the following conditions (Puterman, 1994, Theorem 8.11.3) holds.

(1) $c(s, a)$ is nondecreasing in s for all $a \in \mathbb{A}$;
(2) $c(s, a)$ is a superadditive function on $\mathbb{S} \times \mathbb{A}$;
(3) $Q(s'|s, a) = \sum_{i=s'}^{\infty} P(i|s, a)$ is nonincreasing in s for all $s' \in \mathbb{S}$ and $a \in \mathbb{A}$;
(4) $Q(s'|s, a)$ is a superadditive function on $\mathbb{S} \times \mathbb{A}$ for all $s' \in \mathbb{S}$.

Table 1
Table of next states.

Current state s	Action a	Tx succeeds	Tx fails
s^+	a^+	s_2^+	s_f^+
s^+	a^-	s_1^+	s_f^+
s^-	a^-	s_1^-	s_f^-
s^-	a^+	s_2^-	s_f^-

We first consider the monotonicity in τ_1 with τ_2 fixed. We can define an order relation satisfies: if $\tau_1 \ge \tau_1'$, $s \succcurlyeq s'$ with $s = (\tau_1, \tau_2, q_1, q_2, r_1, r_2)$ and $s' = (\tau_1', \tau_2, q_1, q_2, r_1, r_2)$. According to Lemma 1, condition (1) is immediate. As $c(s, a)$ in independent of a, condition (2) follows trivially.

Suppose the current state is $s = (\tau_1, \tau_2, 0, 0, d_1, d_2)$. With a slight abuse of notation, define $s_1 := (\tau_1+1, \tau_2+1, 1, 0, d_1-1, d_2)$, $s_2 := (\tau_1+1, \tau_2+1, 0, 1, d_1, d_2-1)$ and $s_f = (\tau_1+1, \tau_2+1, 0, 0, d_1, d_2)$. Based on the state transition law defined before, s transits to $s' = s_1$ with probability λ_1 and to $s' = s_f$ with probability $1-\lambda_1$ under $a = 1$. Similarly, s transits to $s' = s_2$ with probability λ_2 and $s' = s_f$ with probability $1-\lambda_2$ under $a = 2$. If $a = i$,

$$Q(s'|s, a) = \begin{cases} 1, & \text{if } s_i \succcurlyeq s', \\ 1-\lambda_i, & \text{if } s_f \succcurlyeq s' \succcurlyeq s_i, \\ 0, & \text{if } s' \succcurlyeq s_f. \end{cases}$$

Therefore, condition (3) is valid as long as $s_f \succcurlyeq s_i$, $i = 1, 2$, which is feasible.

To check condition (4), we need to verify if

$$Q(s'|s^+, a^+) + Q(s'|s^-, a^-) \ge Q(s'|s^+, a^-) + Q(s'|s^-, a^+), \tag{12}$$

where $s^+ \succcurlyeq s^-$ and $a^+ \ge a^-$. There are four cases for the next states, which are summarized in Table 1.

Note that to be consistent with previous three conditions, we require $s_f^+ \succcurlyeq s_1^+, s_2^+, s_f^- \succcurlyeq s_1^-, s_2^-$, and $s_*^+ \succcurlyeq s_*^-$, where the subscript $*$ stands for a fixed symbol. We are able to allow the following order relations hold: $s_f^+ \succcurlyeq s_f^- \succcurlyeq s_1^+ \succcurlyeq s_1^- \succcurlyeq s_2^+ \succcurlyeq s_2^-$. Consequently, (12) holds due to the monotonicity of $Q(s'|s, a)$ and condition (4) holds accordingly.

As the four conditions all hold, the optimal policy is monotone. According to our definition on the order of the states for $(\tau_1, \tau_2, 0, 0, d_1, d_2)$, the first part of the switching curve is proven. Similar arguments can be claimed for the second part. Therefore, the proof is complete.

References

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, *38*(4), 393–422.
Anderson, B., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs: Prentice-Hall.
Cui, J.-H., Kong, J., Gerla, M., & Zhou, S. (2006). The challenges of building mobile underwater wireless networks for aquatic applications. *IEEE Network*, *20*(3), 12–18.
Ding, D., Wang, Z., Ho, D. W., & Wei, G. (2017). Observer-based event-triggering consensus control for multiagent systems with lossy sensors and cyber-attacks. *IEEE Transactions on Cybernetics*.
Dong, W., Chen, C., Liu, X., He, Y., Liu, Y., Bu, J., et al. (2014). Dynamic packet length control in wireless sensor networks. *IEEE Transactions on Wireless Communications*, *13*(3), 1172–1181.
Gupta, V., Chung, T. H., Hassibi, B., & Murray, R. M. (2006). On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, *42*(2), 251–260.
Han, D., Wu, J., Zhang, H., & Shi, L. (2017). Optimal sensor scheduling for multiple linear dynamical systems. *Automatica*, *75*, 260–270.
Hernández-Lerma, O., & Lasserre, J. B. (1996). *Discrete-time Markov control processes: Basic optimality criteria, Vol. 30*. Springer Science & Business Media.

Hovareshti, P., Gupta, V., & Baras, J. S. (2007). Sensor scheduling using smart sensors. In *46th IEEE conference on decision and control* (pp. 494–499). IEEE.

Leong, A. S., Dey, S., & Quevedo, D. E. (2017). Sensor scheduling in variance based event triggered estimation with packet drops. *IEEE Transactions on Automatic Control*, *62*(4), 1880–1895.

Mesquita, A. R., Hespanha, J. P., & Nair, G. N. (2012). Redundant data transmission in control/estimation over lossy networks. *Automatica*, *48*(8), 1612–1620.

Meyn, S. P., & Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Mo, Y., Chabukswar, R., & Sinopoli, B. (2014). Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, *22*(4), 1396–1407.

Mo, Y., Garone, E., & Sinopoli, B. (2014). On infinite-horizon sensor scheduling. *Systems & Control Letters*, *67*, 65–70.

Mo, Y., Shi, L., Ambrosino, R., & Sinopoli, B. (2009). Network lifetime maximization via sensor selection. In *7th Asian control conference, 2009* (pp. 441–446). IEEE.

Mori, S., Ishii, K., & Ogose, S. (2011). Cross-layer designed adaptive packet length control for wireless networks. In *IEEE 22nd international symposium on personal, Indoor and mobile radio communications* (pp. 2354–2358). IEEE.

Nourian, M., Leong, A. S., & Dey, S. (2014). Optimal energy allocation for Kalman filtering over packet dropping links with imperfect acknowledgments and energy harvesting constraints. *IEEE Transactions on Automatic Control*, *59*(8), 2128–2143.

Papadimitriou, C. H., & Tsitsiklis, J. N. (1999). The complexity of optimal queuing network control. *Mathematics of Operations Research*, *24*(2), 293–305.

Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc.

Ren, X., Wu, J., Dey, S., & Shi, L. (2018). Attack allocation on remote state estimation in multi-systems: Structural results and asymptotic solution. *Automatica*, *87*, 184–194.

Ren, X., Wu, J., Johansson, K. H., Shi, G., & Shi, L. (2018). Infinite horizon optimal transmission power control for remote state estimation over fading channels. *IEEE Transactions on Automatic Control*, *63*(1), 85–100.

Rudin, W. (1964). *Principles of mathematical analysis, Vol. 3*. McGraw-hill New York.

Schenato, L. (2008). Optimal estimation in networked control systems subject to random delay and packet drop. *IEEE Transactions on Automatic Control*, *53*(5), 1311–1317.

Sennott, L. I. (1996). The convergence of value iteration in average cost Markov decision chains. *Operations Research Letters*, *19*(1), 11–16.

Shi, D., Chen, T., & Shi, L. (2014). Event-triggered maximum likelihood state estimation. *Automatica*, *50*(1), 247–254.

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

Shi, L., & Zhang, H. (2012). Scheduling two Gauss–Markov systems: An optimal solution for remote state estimation under bandwidth constraint. *IEEE Transactions on Signal Processing*, *60*(4), 2038–2042.

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

Spall, J. C. (2005). *Introduction to stochastic search and optimization: Estimation, simulation, and control, Vol. 65*. John Wiley & Sons.

Tamboli, C., & Manikopoulos, C. N. (1995). Determination of the optimum packet length and buffer sizes for the industrial building automation and control networks. In *Proceedings of the IEEE international symposium on industrial electronics, Vol. 2* (pp. 831–836). IEEE.

Watteyne, T., Palattella, M., & Grieco, L. (2015). Using IEEE 802.15. 4e time-slotted channel hopping (TSCH) in the internet of things (IoT): Problem statement. In *Tech. rep.*.

Wu, J., Jia, Q.-S., Johansson, K. H., & Shi, L. (2013). Event-based sensor data scheduling: Trade-off between communication rate and estimation quality. *IEEE Transactions on Automatic Control*, *58*(4), 1041–1046.

Wu, J., Li, Y., Quevedo, D. E., Lau, V., & Shi, L. (2015). Data-driven power control for state estimation: A Bayesian inference approach. *Automatica*, *54*, 332–339.

Wu, S., Ren, X., Dey, S., & Shi, L. (2017). Optimal scheduling of multiple sensors with packet length constraint. In *20th IFAC world congress* (pp. 14430–14435).

You, K., & Xie, L. (2011). Minimum data rate for mean square stabilizability of linear systems with Markovian packet losses. *IEEE Transactions on Automatic Control*, *56*(4), 772–785.

You, K., & Xie, L. (2013). Survey of recent progress in networked control systems. *Acta Automatica Sinica*, *39*(2), 101–117.

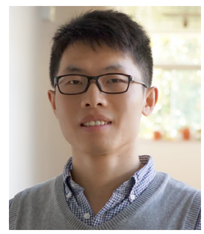
Zhang, L., Gao, H., & Kaynak, O. (2013). Network-induced constraints in networked control systems: A survey. *IEEE Transactions on Industrial Informatics*, *9*(1), 403–416.

Zhao, Y.-B., Kim, J., Shi, P., & Liu, G.-P. (2011). PDV-based packet length allocation for networked control systems. In *2nd international conference on intelligent control and information processin, Vol. 2* (pp. 791–796). IEEE.

Zhao, L., Zhang, W., Hu, J., Abate, A., & Tomlin, C. J. (2014). On the optimal solutions of the infinite-horizon linear sensor scheduling problem. *IEEE Transactions on Automatic Control*, *59*(10), 2825–2830.

Zhou, B., Cui, Y., & Tao, M. (2017). Optimal dynamic multicast scheduling for cache-enabled content-centric wireless networks. *IEEE Transactions on Communications*.

Shuang Wu received the B.Eng. (Hons) degree in the Department of Mechanical Engineering from Zhejiang University, Hangzhou, China, in 2015. He is currently working towards the Ph.D. degree in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology, Hong Kong, China. His current research interests include networked control systems, distributed state estimation and wireless sensor networks.



Xiaoqiang Ren received the B.E. degree in the Department of Control Science and Engineering from Zhejiang University, Hangzhou, China, in 2012 and the Ph.D. degree in the Department of Electronic and Computer Engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2016. He is currently a postdoctoral researcher at the Department of Automatic, KTH Royal Institute of Technology, Sweden. Prior to this, he was a postdoctoral researcher in the Hong Kong University of Science and Technology from September to November 2016 and Nanyang Technological University from December 2016 to February 2018. His research interests include security of cyber–physical systems, sequential detection, and networked estimation and control.

Subhrakanti Dey received the Bachelor in Technology and Master in Technology degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, in 1996.

He is currently a Professor with the Institute of Telecommunications Research, University of South Australia in Adelaide, Australia. He is also a Professor with the Dept. of Engineering Sciences in Uppsala University, Sweden, since 2013. Prior to this, he was a Professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, from 2000 until early 2013. From September 1995 to September 1997, and September 1998 to February 2000, he was a Postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park.

His current research interests include wireless communications and networks, signal processing for sensor networks, networked control systems, and molecular communication systems.

Professor Dey currently serves on the Editorial Board of IEEE Transactions on Signal Processing, IEEE Transactions on Control of Network Systems, and IEEE Transactions on Wireless Communications. He was also an Associate Editor for IEEE Transactions on Signal Processing during 2007–2010 and IEEE Transactions on Automatic Control during 2004–2007, and Associate Editor for Elsevier Systems and Control Letters during 2003–2013.

Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002 and the Ph.D. degree in Control and Dynamical Systems from California Institute of Technology, Pasadena, CA, USA, in 2008. He is currently an associate professor at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include cyber–physical systems security, networked control systems, sensor scheduling, and event-based state estimation. He is a senior member of IEEE. He served as an editorial board member for The European Control Conference 2013–2016. He has been serving as a subject editor for International Journal of Robust and Nonlinear Control from March 2015, an associate editor for IEEE Transactions on Control of Network Systems from July 2016, and an associate editor for IEEE Control Systems Letters from Feb 2017. He also served as an associate editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE Transactions on Control of Network Systems in 2015–2017. He serves as the General Chair of the 23rd International Symposium on Mathematical Theory of Networks and Systems (MTNS 2018).