Automatica 45 (2009) 2134–2140

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Kalman filtering over a packet-delaying network: A probabilistic approach ☆

Ling Shi [a,*], Lihua Xie [b], Richard M. Murray [c]

[a] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[c] Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91106, United States

ARTICLE INFO

Article history:
Received 20 August 2008
Received in revised form
1 May 2009
Accepted 22 May 2009
Available online 27 June 2009

Keywords:
Kalman filters
Networked control systems
Packet-delaying networks
Estimation theory
Probabilistic performance

ABSTRACT

In this paper, we consider Kalman filtering over a packet-delaying network. Given the probability distribution of the delay, we can characterize the filter performance via a probabilistic approach. We assume that the estimator maintains a buffer of length D so that at each time k, the estimator is able to retrieve all available data packets up to time $k-D+1$. Both the cases of sensor with and without necessary computation capability for filter updates are considered. When the sensor has no computation capability, for a given D, we give lower and upper bounds on the probability for which the estimation error covariance is within a prescribed bound. When the sensor has computation capability, we show that the previously derived lower and upper bounds are equal to each other. An approach for determining the minimum buffer length for a required performance in probability is given and an evaluation on the number of expected filter updates is provided. Examples are provided to demonstrate the theory developed in the paper.

1. Introduction

The Kalman filter has played a central role in systems theory and has found wide applications in many fields such as control, signal processing, and communications. In the standard Kalman filter, it is assumed that sensor data are transmitted along perfect communication channels and are available to the estimator either instantaneously or with some fixed delays and no interaction between communication and control is considered. This abstraction has been adopted until recently when networks, especially wireless networks, are used in sensing and control systems for transmitting data from sensor to controller and/or from controller to actuator. While having many advantages such as low cost and flexibility, networks also induce many new issues due to their limited capabilities and uncertainties such as limited bandwidth, packet losses, and latency. On the other hand, in wireless sensor networks, sensor nodes also have limited computation capability in addition to their limitations in communications. These constraints undoubtedly affect system performance or even stability and cannot be neglected when designing estimation and control algorithms, which has inspired a lot of research in control with communication constraints.

Sinopoli, Schenato, Franceschetti, Poolla, Jordan and Sastry (2004) discussed how packet loss can affect state estimation. They showed that there exists a certain threshold of the packet loss rate above which the state estimation error diverges in the expected sense, i.e., the expected value of the error covariance matrix becomes unbounded as time goes to infinity. They also provided lower and upper bounds of the threshold value. Huang and Dey (2007) and Xie and Xie (2008) characterize packet losses as a Markov chain and give some sufficient and necessary stability conditions under the notion of peak covariance stability. The drawback of using mean covariance matrix as a stability measure is that it may conceal the fact that events with arbitrarily low probability may make the mean value diverge. For example, consider the simple unstable scalar system with $a = 2$ in Sinopoli et al. (2004). Let the arrival rate $\gamma = 0.74 < 1 - 1/a^2$. According to Sinopoli et al. (2004), the expected value of the estimation error covariance, $\mathbf{E}[P_k]$, is unbounded. This is easily verifiable by considering the event S where no packets are received in k time steps. Then $\mathbf{E}[P_k] \geq \mathbf{Pr}[S]\mathbf{E}[P_k|S] \geq (0.26^k)4^kP_0 = 1.04^kP_0$. By letting k go to infinity, we see that this event with almost zero probability makes the expected error diverge. Different from Huang and Dey (2007), Sinopoli et al. (2004), Xie and Xie (2008) and Shi, Epstein, Tiwari and Murray (2005) investigates the stability of the Kalman filter via a probabilistic approach. Gupta,

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Andrey V. Savkin under the direction of Editor Ian R. Peterson. The work by L. Shi is partially supported by DAG08/09.EG06. The work by L. Xie is supported by A*STAR SERC grant 052 101 0037 and NSFC grant 60828006. The work by R.M. Murray is supported in part by AFOSR grant FA9550-04-1-0169.

* Corresponding author. Tel.: +852 2358 7055; fax: +852 2358 1485.
E-mail addresses: eesling@ust.hk (L. Shi), elhxie@ntu.edu.sg (L. Xie), murray@cds.caltech.edu (R.M. Murray).

0005-1098/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2009.05.018

Spanos, Hassibi, and Murray (2005) considered LQG control over a packet-dropping network. They showed that preprocessing the measurement by the sensor before sending it to the controller is better than post-processing the measurement after receiving it by the controller. They also gave sufficient condition on the stability of the closed-loop system.

The problem of state estimation and control with delayed measurements is not new and has been studied even before the emergence of networked control (Ray, Liou, & Shen, 1993; Yaz & Ray, 1996). Nilsson (1998) analyzed delays that are either fixed or random according to a Markov chain. He solves the LQG optimal control problem for the different delay models. It has been well known that discrete-time systems with constant or known time-varying bounded measurement delays may be handled by state augmentation in conjunction with the standard Kalman filtering or by the reorganized innovation approach in Zhang, Xie, Zhang, and Soh (2004) and Zhang and Xie (2007). Although sensor data are usually time-stamped and thus transmission delays are known to the filter, the delays in networked systems are random in nature. For example, the ZigBee/IEEE 802.15.4 protocol is widely used in sensor network and wireless control applications (ZigBee, 2009). When multiple sensor nodes simultaneously access the channel, a random waiting time is generated by the CSMA/CA algorithm for each node before they try to access the channel again. Thus the experienced delay for data measurement is typically random.

Ray et al. (1993) present a modification of the conventional minimum variance state estimator to accommodate the effects of the random arrival of measurements whereas a suboptimal filter in the least mean square sense is given in Yaz and Ray (1996). In Matveev and Savkin (2003), a recursive minimum variance state estimator is presented for linear discrete-time partially observed systems where the observations are transmitted by communication channels with randomly independent delays. Using covariance information, recursive least squares linear estimators for signals with random delays are studied in Nakamori, Caballero-Aguila, Hermoso-Carazo, and Linares-Perez (2005). Costa, Fragoso, and Marques (2005) studied linear systems with random delays using a Markovian jump linear systems approach.

The goal of the present work is to study the performance of Kalman filter under random measurement delay. We assume that the probability distribution of the delay is given and aim to give a complete characterization of filter performance by a probabilistic approach. Due to the limited computation capability of the filtering center and also in consideration of the fact that a late arriving measurement related to the system state in the far past may not contribute much to the improvement of the accuracy of the current estimate, it is practically important to determine a proper buffer length for measurement data within which a measurement will be used to update the current state and beyond which the data will be discarded. The buffer provides a tradeoff between performance and computational load. In the paper, for a given buffer length, we shall give lower and upper bounds for the probability at which the filtering error covariance is within a prescribed bound, i.e., $\mathbf{Pr}[P_k \leq M]$ for some given M. The upper and lower bounds can be easily evaluated by the probability distribution of the delay and the system dynamics. An approach for determining the minimum buffer length for a required performance in probability is given and an evaluation on the number of expected filter updates is provided. Both the cases of sensor with and without necessary computation capability for filter updates are considered. Our results will have both theoretical and practical importance in networked sensing and control.

The rest of the paper is organized as follows. In Section 2, the mathematical models of the problem are given. In Section 3, we consider the case when measurement data is sent via the delaying network, and we provide lower and upper bounds for $\mathbf{Pr}[P_k \leq M]$. In Section 4, we consider the case when sensor estimate is sent via the delaying network, and we show that the previously derived lower and upper bounds equal to each other and hence give an exact form of $\mathbf{Pr}[P_k \leq M]$. Examples are provided in Section 5 to demonstrate the theory developed in the paper. Some concluding remarks are given in the end.

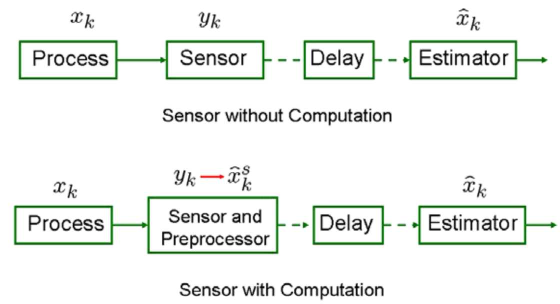

Fig. 1. System block diagram.

2. Problem setup

2.1. System model

We consider the problem of discrete-time state estimation over a packet-delaying network as seen from Fig. 1. The process dynamics and sensor measurement equation are given as follows:

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

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

In the above equations, k is the discrete-time index which is an integer, $x_k \in \mathbb{R}^n$ is the state vector, $y_k \in \mathbb{R}^m$ is the observation vector, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are zero-mean Gaussian random vectors with $\mathbb{E}[w_k w_j'] = \delta_{kj}Q, Q \geq 0$, $\mathbb{E}[v_k v_j'] = \delta_{kj}R, R > 0$, $\mathbb{E}[w_k v_j'] = 0\ \forall j, k$, where $\delta_{kj} = 0$ if $k \neq j$ and $\delta_{kj} = 1$ otherwise. The initial state x_0 is assumed to be zero-mean Gaussian with covariance $P_0 \geq 0$ and is uncorrelated with w_k and v_k for all k. We assume that the pair (A, C) is observable, $(A, \sqrt{Q})$ is controllable.

Depending on its computational capability, the sensor can either send y_k or preprocess y_k and send $\hat{x}_k^s$ to the remote estimator, where $\hat{x}_k^s$ is defined at the sensor as

$$\hat{x}_k^s \triangleq \mathbb{E}[x_k | y_1, \ldots, y_k].$$

The two cases correspond to the two scenarios in Fig. 1, i.e., sensor without/with computation capability.

2.2. Network delay model

After taking a measurement at time k, the sensor sends y_k (or $\hat{x}_k^s$) to a remote estimator for generating the state estimate. We assume that the measurement data packets from the sensor are to be sent across a packet-delaying network, with negligible quantization effects, to the estimator. Each y_k (or $\hat{x}_k^s$) is delayed by d_k times, where d_k is a random variable described by a probability mass function f, i.e.,

$$f(j) = \mathbf{Pr}[d_k = j], \quad j = 0, 1, \ldots. \tag{3}$$

For simplicity, we assume that d_{k_1} and d_{k_2} are independent if $k_1 \neq k_2$. We further assume that d_k carries no information about the state, e.g., d_k is independent of w_k, v_k and the initial state x_0. Similar to Shi, Epstein, and Murray (2008), we can use a Markov chain to model consecutive data packet delays, and the results extend straightforwardly to that case. Notice that the i.i.d packet drop with drop rate $1 - \gamma$ considered in the literature can be treated as a special case here, i.e.,

$$f(0) = \gamma, \quad f(\infty) = 1 - \gamma, \qquad f(j) = 0, \quad 1 \leq j < \infty.$$

Thus the theory developed in the paper includes the packet drop analysis as well.

2.3. Problems of interest

Assume the estimator discards any data y_k (or $\hat{x}_k^s$) that are delayed by D times or more. For example, if y_{k-D} is not received by the estimator before k, then even if y_{k-D} arrives at k or at a later time, it will be discarded by the estimator. This happens as long as the delay d_{k-D} associated with y_{k-D} satisfies $d_{k-D} \ge D$. Therefore "discarding" always precedes "receiving" the data.

Define $\mathbf{Y}_k$ as all the received data that are not discarded by the estimator up to k. Further define

$$\hat{x}_k \triangleq \mathbb{E}[x_k|\mathbf{Y}_k], \qquad e_k \triangleq x_k - \hat{x}_k, \qquad P_k \triangleq \mathbb{E}[e_k e_k'|\mathbf{Y}_k].$$

Notice that $\mathbf{Y}_k$ is a function of D and f, i.e., for different length of the buffer and for different delay distribution of the data packets, the received data that are not discarded would be different. Hence $\hat{x}_k, e_k$ and P_k as defined above are also functions of D and f. We sometimes write $\hat{x}_k(D,f), e_k(D,f), P_k(D,f)$ as $\hat{x}_k, e_k, P_k$ for simplicity if there is no confusion on what value D and f takes. Since P_k reflects how well the state estimate $\hat{x}_k$ is close to the true state x_k and P_k is a random quantity whose randomness is induced by the randomness of $d_i, i = 1, \ldots, k$, we focus on the statistic property of P_k in this paper.

Given the system and the network delay models in Eqs. (1)–(3), we are thus interested in the following problems.

(1) How should $\hat{x}_k(D,f)$ and $P_k(D,f)$ be computed?
(2) For a given data packet delay distribution f, $M \ge 0$, and $\epsilon \in [0,1]$ what is the minimum D such that

$$\mathbf{Pr}[P_k(D,f) \le M] \ge 1 - \epsilon.$$

The first problem is a classic estimation problem and its solution is actually known for a long time (e.g., Liptser and Shiryaev (1979)). The second problem, to the best of our knowledge, seems to be never studied in the literature. In the rest of the paper, we first present an alternative solution for the first problem and then solve the second problem for each of the two scenarios in Fig. 1.

The following terms that are frequently used in subsequent sections are first defined. It is assumed that (A, C, Q, R) are the same as they appear in Section 2; $X \in \mathbb{S}_+^n$ where $\mathbb{S}_+^n$ is the set of n by n positive semi-definite matrices; $h, g : \mathbb{S}_+^n \to \mathbb{S}_+^n$ are matrix functions defined below:

$$h(X) \triangleq AXA' + Q,$$
$$g(X) \triangleq h(X) - AXC'[CXC' + R]^{-1}CXA',$$
$$\tilde{g}(X) \triangleq X - XC'[CXC' + R]^{-1}CX,$$
$$h \circ g(X) \triangleq h(g(X)),$$
$$h^t(X) \triangleq \underbrace{h \circ \cdots \circ h}_{t \text{ times}}(X).$$

3. Sensor without computation capability

In this section, we consider the first scenario in Fig. 1, i.e., the sensor has no computation and sends y_k to the remote estimator. We assume that C is full rank, and without loss of generality, we assume that C^{-1} exists. The general C case will be considered in Appendix B.

3.1. Modified Kalman filtering

Let γ_t^k be the indicator function for y_t at time $k, t \le k$, i.e., $\gamma_t^k = 1$ if y_t arrives at k and $\gamma_t^k = 0$ otherwise. Further define $\gamma_{k-i} \triangleq \sum_{j=0}^{i} \gamma_{k-i}^{k-j}$, i.e., γ_{k-i} indicates whether y_{k-i} is received by the estimator at or before k.

Depending on whether y_k is received or not, i.e., $\gamma_k^k = 1$ or 0, $(\hat{x}_k, P_k)$ is known to be computed by a Modified Kalman Filter (**MKF**) (Sinopoli et al., 2004). We write $(\hat{x}_k, P_k)$ in compact form as follows.

$$(\hat{x}_k, P_k) = \mathbf{MKF}(\hat{x}_{k-1}, P_{k-1}, \gamma_k^k, y_k)$$

which represents the following set of equations:

$$\begin{cases} \hat{x}_k^- = A\hat{x}_{k-1}, \\ P_k^- = AP_{k-1}A' + Q, \\ K_k = P_k^- C'[CP_k^- C' + R]^{-1}, \\ \hat{x}_k = A\hat{x}_{k-1} + \gamma_k^k K_k(y_k - CA\hat{x}_{k-1}), \\ P_k = (I - \gamma_k^k K_k C)P_k^-. \end{cases}$$

Assume that $\gamma_k^k = 1$ for all k, then **MKF** reduces to the standard Kalman filter. In this case, P_k^- and P_k can be shown to satisfy

$$P_k^- = g(P_{k-1}^-), \qquad P_k = \tilde{g} \circ h(P_{k-1}).$$

Let P^* be the unique positive semi-definite solution[1] to $g(X) = X$, i.e., $P^* = g(P^*)$. Define $\overline{P}$ as $\overline{P} \triangleq \tilde{g}(P^*)$. Then we have

$$\tilde{g} \circ h(\overline{P}) = \tilde{g} \circ h \circ \tilde{g}(P^*) = \tilde{g} \circ g(P^*) = \tilde{g}(P^*) = \overline{P},$$

where we use the fact that $h \circ \tilde{g} = g$. In other words,

$$P^* = \lim_{k\to\infty} P_k^-, \qquad \overline{P} = \lim_{k\to\infty} P_k.$$

3.2. Optimal estimation with delayed measurements

As y_{k-i} may arrive at time k due to the delays introduced by the network, we can improve the estimation quality by recalculating $\hat{x}_{k-i}$ utilizing the new available measurement y_{k-i}. Once $\hat{x}_{k-i}$ is updated, we can update $\hat{x}_{k-i+1}$ in a similar fashion. The following proposition summarizes the estimation process.

Proposition 3.1. *Let $y_{k-i}, i \in [0, D-1]$ be the oldest measurement received by the estimator at time k. Then $\hat{x}_k$ is computed by $i+1$ **MKF** s as*

$$(\hat{x}_{k-i}, P_{k-i}) = \mathbf{MKF}(\hat{x}_{k-i-1}, P_{k-i-1}, 1, y_{k-i})$$
$$(\hat{x}_{k-i+1}, P_{k-i+1}) = \mathbf{MKF}(\hat{x}_{k-i}, P_{k-i}, \gamma_{k-i+1}, y_{k-i+1})$$
$$\vdots$$
$$(\hat{x}_{k-1}, P_{k-1}) = \mathbf{MKF}(\hat{x}_{k-2}, P_{k-2}, \gamma_{k-1}, y_{k-1})$$
$$(\hat{x}_k, P_k) = \mathbf{MKF}(\hat{x}_{k-1}, P_{k-1}, \gamma_k, y_k).$$

Proof. We know that the estimate $\hat{x}_k$ is generated from the estimate of $\hat{x}_{k-1}$ together with $\gamma_k y_k$ at time k through a **MKF**. Similarly, the estimate $\hat{x}_{k-1}$ is generated from the estimate of $\hat{x}_{k-2}$ together with $\gamma_{k-1}y_{k-1}$ at time k through a **MKF**, etc. This recursion for $i+1$ steps corresponds to the $i+1$ **MKF**s stated in the proposition. □

Remark 3.2. Similar results on estimation with delayed packets can be derived in other ways (e.g., Liptser and Shiryaev (1979) and Matveev and Savkin (2003)) but the result in Proposition 3.1 fits our intended uses later in the paper.

Remark 3.3. Although this section focuses on when C is invertible. Proposition 3.1 does not need to assume C to be invertible. Depending on whether (A, C) is detectable or not, the error covariance matrix may diverge. However, the algorithm always produces the optimal estimate at each time.

3.3. Lower and upper bounds of $\mathbf{Pr}[P_k(D,f) \le M]$

Since d_k is random and described by the probability mass function f, $\gamma_{k-i}^k (i = 0, \ldots, D-1)$ is also random. As a consequence, P_k computed as in Proposition 3.1 is a random variable. Let

[1] Since (A, C) is observable and $(A, \sqrt{Q})$ is controllable, from standard Kalman filtering analysis, P^* exists.

$$F(i)=\sum_{j=0}^{i} f(j)$$

be the cumulative distribution function of d_k. Define $\hat{\gamma}_i(D)$ as

$$\hat{\gamma}_i(D) \triangleq \begin{cases} F(i), & \text{if } 0 \le i < D, \\ F(D-1), & \text{if } i \ge D. \end{cases}$$

Recall that γ_{k-i} indicates whether y_{k-i} is received by the estimator at or before k, so it is easy to verify that

$$\mathbf{Pr}[\gamma_{k-i}=1]=\hat{\gamma}_i(D). \tag{4}$$

Define $\overline{M} \triangleq C^{-1}RC^{-1'}$. Then we have the following result that shows the relationship between P_k and $\overline{M}$.

Lemma 3.4. *For any $k \ge 1$, if $\gamma_k = 1$, then $P_k \le \overline{M}$.*

Proof. As $\gamma_k = 1$, we have $P_k = \tilde{g} \circ h(P_{k-1}) \le \overline{M}$, where the inequality is from Lemma A.1 in Appendix A. □

Remark 3.5. We can also interpret Lemma 3.4 as follows. One way to obtain an estimate $\tilde{x}_k$ of x_k when $\gamma_k = 1$ is simply by inverting the measurement, i.e., $\tilde{x}_k = C^{-1}y_k$. Therefore $\tilde{e}_k = C^{-1}v_k$ and $\tilde{P}_k = \mathbb{E}[\tilde{e}_k\tilde{e}_k'] = C^{-1}RC^{-1'} = \overline{M}$. Since Kalman filter is optimal among the set of all linear filters, we must have $P_k \le \tilde{P}_k = \overline{M}$.

Recall that $\overline{P}$ is defined in Section 3.1 as the steady state error covariance for the Kalman filter. For $M \ge \overline{M}$, let us define $k_1(M)$ and $k_2(M)$ as follows:

$$k_1(M) \triangleq \min\{t \ge 1 : h^t(\overline{M}) \nleq M\}, \tag{5}$$

$$k_2(M) \triangleq \min\{t \ge 1 : h^t(\overline{P}) \nleq M\}. \tag{6}$$

We sometimes write $k_i(M)$ as k_i, $i = 1, 2$ for simplicity for the rest of the paper. The following lemma shows the relationship between $\overline{P}$ and $\overline{M}$ as well as k_1 and k_2.

Lemma 3.6. (1) *$\overline{P} \le \overline{M}$; (2) If $k_1 < \infty$, then $k_1 \le k_2$.*

Proof. (1) $\overline{P} = \tilde{g}(P^*) \le \overline{M}$ where the inequality is from Lemma A.1 in Appendix A. (2) Without loss of generality, we assume that $k_2 < \infty$. If $k_1 > k_2$, then according to their definitions, we must have

$$M \ge h^{k_1-1}(\overline{M}) \ge h^{k_1-1}(\overline{P}) \ge h^{k_2}(\overline{P}),$$

which violates the definition of k_2. Notice that we use the property that h is nondecreasing as well as $h(\overline{P}) \ge \overline{P}$ from Lemmas A.1 and A.2 in Appendix A. □

Lemma 3.7. *Assume that $P_0 \ge \overline{P}$. Then $P_k \ge \overline{P}$ for all $k \ge 0$.*

Proof. Since **MKF** is used at each time k,

$$P_k = \hat{f}_k^k \circ \hat{f}_{k-1}^k \cdots \hat{f}_1^k(P_0) \ge \overline{P},$$

where $\hat{f}_{k-i}^k = h$ or $\hat{f}_{k-i}^k = \tilde{g} \circ h$ depending on the packet arrival sequence.[2] The inequality is from Lemma A.1 in Appendix A. □

Define N_k as the number of consecutive packets not received by k, i.e.,

$$N_k \triangleq \min\{t \ge 0 : \gamma_{k-t} = 1\}. \tag{7}$$

Thus N_k is the minimum of a sequence of independent Bernoulli random variables, therefore if we define

$$\theta(k_i, D) \triangleq \prod_{j=0}^{k_i-1}\left(1-\hat{\gamma}_j(D)\right), \tag{8}$$

then $\mathbf{Pr}[N_k \ge k_i]$ can be easily shown to be

$$\mathbf{Pr}[N_k \ge k_i] = \theta(k_i, D), \quad i = 1, 2. \tag{9}$$

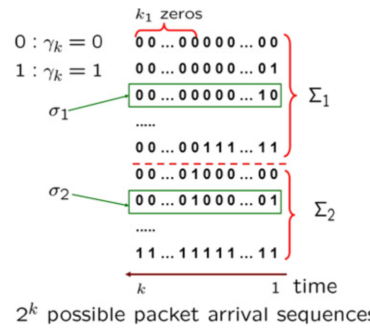

Fig. 2. $N_k \ge k_1$.

We are now ready to present the main result of this section.

Theorem 3.8. *Given the buffer length D and the delay distribution function f. Further assume that $\overline{P} \le P_0 \le \overline{M}$. Then for any $M \ge \overline{M}$, we have*

$$1-\theta(k_1, D) \le \mathbf{Pr}[P_k \le M] \le 1-\theta(k_2, D). \tag{10}$$

Proof. Let us first prove $1-\theta(k_1, D) \le \mathbf{Pr}[P_k \le M]$, or in other words, $1-\mathbf{Pr}[N_k \ge k_1] \le \mathbf{Pr}[P_k \le M]$. As $\gamma_k = 1$ or 0, there are in total 2^k possible realizations of γ_1 to γ_k as seen from Fig. 2. Let Σ_1 denote those packet arrival sequences of γ_1 to γ_k such that $N_k \ge k_1$. Similarly let Σ_2 denote those packet arrival sequences such that $N_k < k_1$. Let $P_k(\sigma_i)$ be the error covariance at time k when the underlying packet arrival sequence is σ_i, where $\sigma_i \in \Sigma_i$, $i = 1, 2$. Consider a particular $\sigma_2 \in \Sigma_2$. As $\gamma_{k-k_1+1} = 1$, from Lemma 3.4, $P_{k-k_1+1} \le \overline{M}$. Therefore we have

$$P_k(\sigma_2) \le h^{k_1-1}(P_{k-k_1+1}) \le h^{k_1-1}(\overline{M}) \le M,$$

where the first and second inequalities are from Lemma A.1 in Appendix A and the last inequality is from the definition of k_1. In other words, $\mathbf{Pr}[P_k \le M|\sigma_2] = 1$. Therefore we have

$$\begin{aligned} \mathbf{Pr}[P_k \le M] &= \sum_{\sigma \in \Sigma_1 \cup \Sigma_2} \mathbf{Pr}[P_k \le M|\sigma]\mathbf{Pr}(\sigma) \\ &\ge \sum_{\sigma_2 \in \Sigma_2} \mathbf{Pr}[P_k \le M|\sigma_2]\mathbf{Pr}(\sigma_2) \\ &= \sum_{\sigma_2 \in \Sigma_2} \mathbf{Pr}(\sigma_2) = \mathbf{Pr}(\Sigma_2) \\ &= 1-\mathbf{Pr}(\Sigma_1) = 1-\mathbf{Pr}[N_k \ge k_1]. \end{aligned}$$

Similarly, we can prove that $\mathbf{Pr}[P_k \le M] \le 1-\theta(k_2, D)$. □

3.4. Computing the minimum D

Assume that we require $\mathbf{Pr}[P_k(D, f) \le M] \ge 1-\epsilon$ for a given $\epsilon \in [0, 1]$. Then according to Eq. (10), a sufficient condition is $\theta(k_1, D) \le \epsilon$ and a necessary condition is $\theta(k_2, D) \le \epsilon$.

3.4.1. Sufficient minimum D

Notice that $\theta(k_1, D)$ is decreasing when $1 \le D \le k_1 - 1$ and remains constant when $D \ge k_1$. Hence if $\epsilon < \theta(k_1, k_1 - 1)$, no matter how large D is, there is *no guarantee* that $\mathbf{Pr}[P_k(D, f) \le M] \ge 1-\epsilon$. For any $\epsilon \in [0, 1]$, the corresponding sufficient minimum D, denoted D_s is thus given by:

$$D_s = \begin{cases} \infty, & \text{if } 0 \le \epsilon < \theta(k_1, k_1-1), \\ \min\{D : \theta(k_1, D) \le \epsilon\}, & \text{if } 1 \ge \epsilon \ge \theta(k_1, k_1-1). \end{cases}$$

[2] Notice that we use the superscript k in $\hat{f}_{k-i}^k$ to emphasize that it depends on the current time k. For example, if $d_{k-i} = i + 1$, i.e., $\gamma_{k-i} = 0$ and $\gamma_{k-i}^{k+1} = 1$, then $\hat{f}_{k-i}^k = h$ and $\hat{f}_{k-i}^{k+1} = \tilde{g} \circ h$.

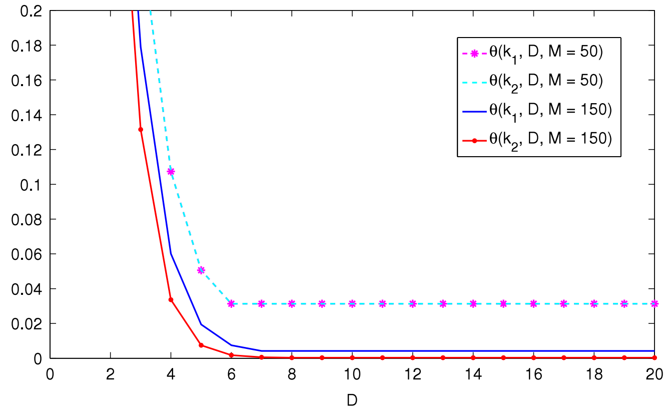

Fig. 3. $\theta(k_i, D)$ for different M.

3.4.2. Necessary minimum D

Similarly, $\theta(k_2, D)$ is decreasing when $1 \le D \le k_2 - 1$ and remains constant when $D \ge k_2$. Hence if $\epsilon < \theta(k_2, k_2 - 1)$, no matter how large D is, it is guaranteed that $\mathbf{Pr}[P_k(D, f) \le M] > 1 - \epsilon$. For any $\epsilon \in [0, 1]$, the corresponding necessary minimum D, denoted D_n is thus given by:

$$D_n = \begin{cases} \infty, & \text{if } 0 \le \epsilon < \theta(k_2, k_2 - 1), \\ \min\{D : \theta(k_2, D) \le \epsilon\}, & \text{if } 1 \ge \epsilon \ge \theta(k_2, k_1 - 1). \end{cases}$$

Remark 3.9. We can use binary search algorithm to find the exact value of D_s and D_n efficiently as $\theta(k_i, D)$ is a monotonically decreasing function of D.

Example 3.10. Consider Eqs. (1) and (2) with

$$A = 1.4, \quad C = 1, \quad Q = 0.2, \quad R = 0.5.$$

We model the packet delay as a poisson distribution with mean d, i.e., the probability density function $f(i)$ satisfies

$$f(i) = \frac{d^i e^{-d}}{i!}, \quad i = 0, 1, \ldots$$

where $d = \mathbb{E}[d_k]$ denotes the mean value of the packet delay.

When $M = 50$, it is calculated that $k_1(M) = k_2(M) = 7$, hence $\theta(k_1, D) = \theta(k_2, D)$ and $\theta(7, 6) = 0.0313$ which can be seen from the plot in Fig. 3. Thus we can find the minimum D that guarantees $\mathbf{Pr}[P_k \le 50] \ge 1 - \epsilon$ for any $\epsilon \ge 0.0313$. For any $\epsilon < 0.0313$, no matter how large D is, $\mathbf{Pr}[P_k \le 50] < 1 - \epsilon$.

When $M = 150$, it is calculated that $k_1(M) = 8$ and $k_2(M) = 9$, hence $\theta(k_1, D) > \theta(k_2, D)$. We also find that $\theta(8, 7) = 0.0042$ and $\theta(9, 8) = 0.0003$. Therefore if $\epsilon > 0.0042$, we can find minimum D that guarantees $\mathbf{Pr}[P_k \le 150] \ge 1 - \epsilon$; if $\epsilon < 0.0003$, no matter how large D is, $\mathbf{Pr}[P_k \le 150] > 1 - \epsilon$.

Remark 3.11. We find the minimum D that gives the desired filter performance, i.e., $\mathbf{Pr}[P_k \le M] \ge 1 - \epsilon$ for a given M and ϵ. Using the results developed in this section, it is straightforward to find the sufficient and necessary minimum D such that $\mathbb{E}[P_k(D)]$ is stable.

4. Sensor with computation capability

In this section, we consider the second scenario in Fig. 1, i.e., the sensor has necessary computation capability and sends $\hat{x}_k^s$ to the remote estimator. We assume that all the variables in this section, e.g., γ_t^k, γ_k, etc are the same as they are defined in Section 3 unless they are explicitly defined.

Consider the case when k is sufficiently large so that the Kalman filter enters steady state at the sensor side, i.e., $P_k^s = \overline{P}$. It is clear that the optimal estimation at the remote estimator is as follows. If $\gamma_k = 1$, then $\hat{x}_k = \hat{x}_k^s$ and $P_k = P_k^s = \overline{P}$. If $\gamma_k = 0$ and $\gamma_{k-1} = 1$, then $\hat{x}_k = A\hat{x}_{k-1}^s$ and $P_k = h(\overline{P})$. This is repeated until we examine γ_{k-D+1}. The full optimal estimation algorithm is presented in Fig. 4.

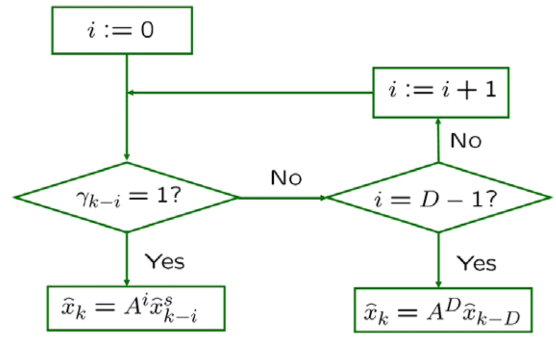

Fig. 4. Optimal estimation: Sensor with computation capability.

Theorem 4.1. *Given the buffer length D and the delay distribution function f. Further assume that k is sufficiently large such that $P_k^s = \overline{P}$. Then for any $M \ge \overline{P}$, we have*

$$\mathbf{Pr}[P_k \le M] = 1 - \theta(k_2, D). \tag{11}$$

Proof. The proof is similar to the proof of Theorem 3.8. □

Computing $\mathbf{Pr}[N_k \ge k_2]$ follows exactly the same way as in Section 3.4. Since we have a strict equality in Eq. (11), in order that $\mathbf{Pr}[P_k(D, f) \le M] \ge 1 - \epsilon$, a necessary and sufficient condition is $\mathbf{Pr}[N_k \ge k_2] \le \epsilon$, from which we can calculate the minimum D^* as

$$D^* = \min\{D : \theta(k_2, D) \le \epsilon, 1 \le D \le k_2 - 1\}. \tag{12}$$

Notice that since $\theta(k_2, D) \ge \theta(k_2, k_2 - 1) = \theta(k_2, k_2 - 1)$, D^* from the above equation exists if and only if $\epsilon \ge \theta(k_2, k_2 - 1)$.

5. Examples

5.1. Scalar system

Consider the same parameters as in Example 3.10. We run a Monte Carlo simulation of $\mathbf{Pr}[P_k \le M]$ for different D and d to demonstrate the result of Theorem 3.8. In Figs. 5–7, both the upper and lower bounds of $\mathbf{Pr}[P_k \le M]$ are plotted. From those figures, we can see that both smaller d and larger D lead to larger $\mathbf{Pr}[P_k \le M]$, which confirms the theory developed in this paper. We further notice that when $d = 3$, the filter's performances using $D = 10$ and $D = 5$ only differ slightly (though the former one is slightly better than the latter one), which confirms that using a large buffer may not improve the filter performance drastically.

We also run a Monte Carlo simulation of $\mathbf{Pr}[P_k \le M]$ to demonstrate the result of Theorem 4.1. Fig. 8 shows the result when $D = 10$ and $d = 5$. As we can see, the predicted value of $\mathbf{Pr}[P_k \le M]$ from Eq. (11) matches well with the actual value.

5.2. Vector system

Consider a vehicle moving in a two-dimensional space according to the standard constant acceleration model, which assumes that the vehicle has zero acceleration except for a small perturbation. The state of the vehicle consists of its x and y positions as well as velocities. Assume that a sensor measures the positions of the vehicle and sends the measurements to a remote estimator over a packet-delaying network. The system parameters are given according to Eqs. (1)–(2) as follows:

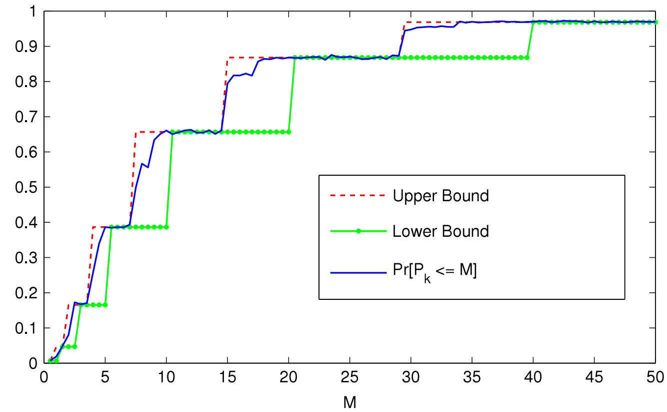

Fig. 5. $\mathbf{Pr}[P_k \le M]$ with $D = 10$, $d = 5$.

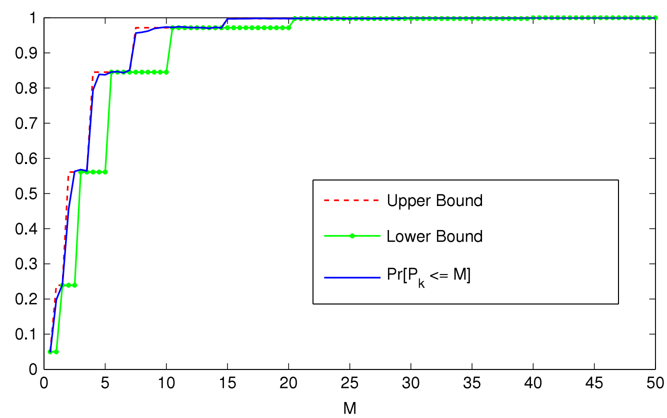

Fig. 6. $\mathbf{Pr}[P_k \le M]$ with $D = 10$, $d = 3$.

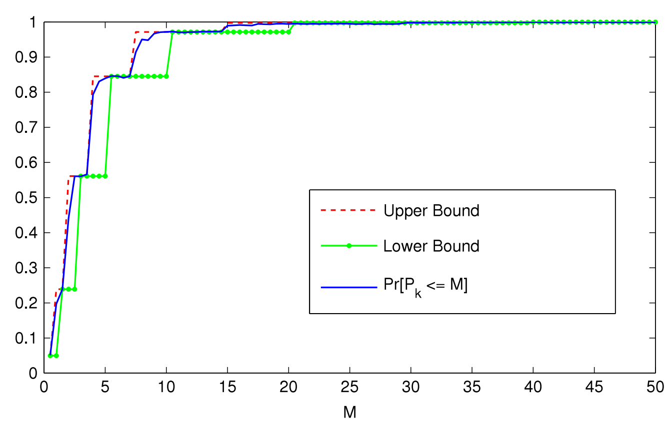

Fig. 7. $\mathbf{Pr}[P_k \le M]$ with $D = 5$, $d = 3$.

$$A = \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The process and measurement noise covariances are $Q = \mathbf{diag}(0.01, 0.01, 0.01, 0.01)$ and $R = \mathbf{diag}(0.001, 0.001)$. We assume the same delay profile as in the scalar system example with $D = 5$ and $d = 3$.

We run a Monte Carlo simulation for both cases when the sensor has or has not computation capability. As we can see from Fig. 9, the predicted values of $\mathbf{Pr}[P_k \le M]$ match well with the actual values. We also notice that when sensor has computation capability, the actual filter performance is better than when sensor

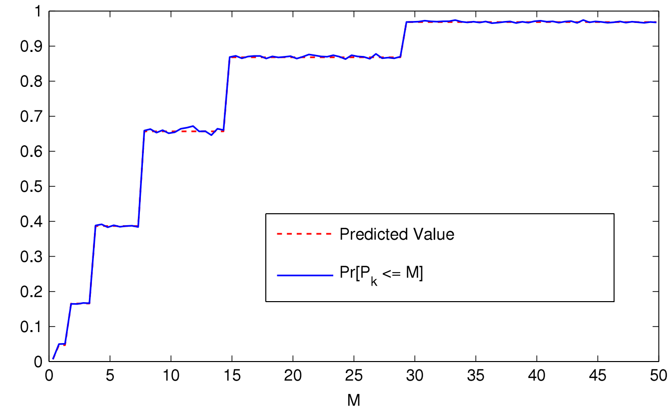

Fig. 8. $\mathbf{Pr}[P_k \le M]$ with $D = 10$, $d = 5$.

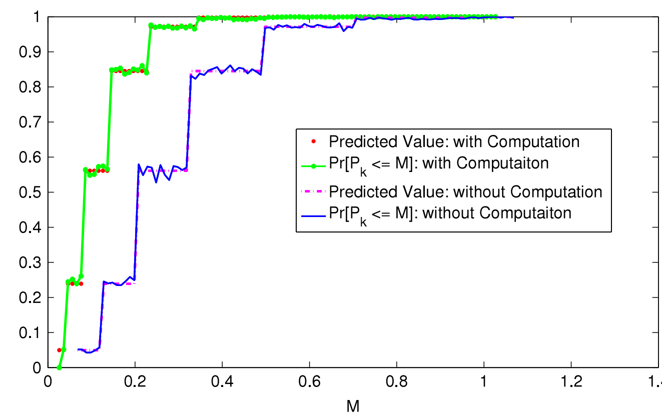

Fig. 9. $\mathbf{Pr}[P_k \le M]$ with $D = 5$, $d = 3$.

has no computation capability. In Fig. 9, the M in the x-axis means $M \times I_4$, where I_4 is the identity matrix of dimension 4.

6. Conclusion

In this paper, we have considered Kalman filtering over a packet-delaying network. Given the distribution of the network induced delay as well as the size of the buffer at the remote estimator, we have characterized the error covariance via a probabilistic approach, i.e., by finding $\mathbf{Pr}[P_k \le M]$. When measurement data is sent, we give lower and upper bounds on $\mathbf{Pr}[P_k \le M]$; when estimate data is sent, we provide an exact form on $\mathbf{Pr}[P_k \le M]$.

There are many interesting works that lie ahead which include: closing the loop using the filtering algorithms proposed in the paper and study closed-loop performance; experimentally evaluate the algorithms and theory developed in the paper; extend the results to multi-sensor scenarios. It is also interesting to study other buffering schemes, for example, consider the D available data packets instead of D most recent data packets, and give conditions on which scheme should be used under what circumstance.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments and suggestions which helped to improve the paper.

Appendix A. Supporting lemmas

Lemma A.1. *For any* $0 \le X \le Y$,

$$h(X) \le h(Y), \quad g(X) \le g(Y), \quad \tilde{g}(X) \le \tilde{g}(Y),$$
$$\tilde{g}(X) \le X, \quad g(X) \le h(X), \quad \tilde{g}(X) \le \overline{M}.$$

Proof. $h(X) \le h(Y)$ holds as $h(X)$ is affine in X. Proof for $g(X) \le g(Y)$ can be found in Lemma 1-c in Sinopoli et al. (2004). As $\tilde{g}$ is a special form of g by setting $A = I$ and $Q = 0$, we immediately obtain $\tilde{g}(X) \le \tilde{g}(Y)$. Next we have

$$\tilde{g}(X) = X - XC'[CXC' + R]^{-1}CX \le X,$$
$$g(X) = h(X) - AXC'[CXC' + R]^{-1}CXA' \le h(X).$$

For any $t > 0$, we have $\tilde{g}(t\overline{M}) = \frac{t}{t+1}\overline{M} \le \overline{M}$. For all $X \ge 0$, since $\overline{M} > 0$, it is clear that there exists $t_1 > 0$ such that $t_1\overline{M} > X$. Therefore $\tilde{g}(X) \le \tilde{g}(t_1\overline{M}) \le \overline{M}$. □

Lemma A.2. $\overline{P} \le h(\overline{P})$.

Proof. $h(\overline{P}) = h \circ \tilde{g}(P^*) = g(P^*) = P^* \ge \tilde{g}(P^*) = \overline{P}$, where the first and the last equalities are from the definition of $\overline{P}$, the third equality is from the definition of P^*. The remaining equality and inequality are from Lemma A.1. □

Appendix B. When C is not full rank

We consider the case when C is not full rank for the first scenario, i.e., sensor without computation capability. Since (A, C) is observable, there exists r $(2 \le r \le n)$ such that $[C; CA; \cdots; CA^{r-1}]'$ is full rank. In this section, we consider the special case when $r = 2$, and in particular, we assume that $\begin{bmatrix} C \\ CA \end{bmatrix}^{-1}$ exists. The idea readily extends to the general case.

Unlike the case when C^{-1} exists, and y_k is sent across the network, here we assume that the previous measurement y_{k-1} is sent along with y_k. This only requires that the sensor has a buffer that stores y_{k-1}. Then if $\gamma_k = 1$, both y_k and y_{k-1} are received. Thus the following linear estimator can be constructed in parallel to the Kalman filter

$$\tilde{x}_k = A\begin{bmatrix} CA \\ C \end{bmatrix}^{-1}\begin{bmatrix} y_k \\ y_{k-1} \end{bmatrix}.$$

The corresponding error covariance can be calculated as

$$\tilde{P}_k = \overline{M} \triangleq AM_1A' + Q - A\begin{bmatrix} CA \\ C \end{bmatrix}^{-1}\begin{bmatrix} CQ \\ 0 \end{bmatrix} - \begin{bmatrix} CQ \\ 0 \end{bmatrix}'\begin{bmatrix} CA \\ C \end{bmatrix}^{-1'} A',$$

where

$$M_1 = \begin{bmatrix} CA \\ C \end{bmatrix}^{-1}\begin{bmatrix} CQC' + R & 0 \\ 0 & R \end{bmatrix}\begin{bmatrix} CA \\ C \end{bmatrix}^{-1'}.$$

Since Kalman filter is optimal among the set of linear estimators, we have $P_k \le \tilde{P}_k$. Therefore once the packet for time k is received, i.e., $\gamma_k = 1$, we have $P_k \le \overline{M}$. Therefore we obtained the same results in Theorem 3.8.

References

Costa, O., Fragoso, M., & Marques, R. (2005). *Discrete-time Markov jump linear systems*. Springer.
Gupta, V., Spanos, D., Hassibi, B., & Murray, R. M. (2005). On lqg control across a stochastic packet-dropping link. In *American control conference*.
Huang, M., & Dey, S. (2007). Stability of kalman filtering with markovian packet losses. *Automatica*, *43*, 598–607.
Liptser, R. S., & Shiryaev, A. N. (1979). *Statistics of random processes*. Springer. 2005.
Matveev, A. S., & Savkin, A. V. (2003). The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Transactions on Automatic Control*, *48*, 670–676.
Nakamori, S., Caballero-Aguila, R., Hermoso-Carazo, A., & Linares-Perez, J. (2005). Recursive estimators of signals from measurements with stochastic delays using covariance information. *Applied Mathematics and Computation*, *162*, 65–79.
Nilsson, J. (1998). Real time control systems with delays. *Ph.D. dissertation.* Lund Institute of Technology, Lund, Sweden.
Ray, A., Liou, L. W., & Shen, J. (1993). State estimation using randomly delayed measurements. *Journal of Dynamical Systems, Measurement and Control*, *115*, 19–26.
Shi, L., Epstein, M., & Murray, R. M. (2008). Kalman filtering over a packet dropping network: A probabilistic approach. In *Tenth international conference on control, automation, robotics and vision*.
Shi, L., Epstein, M., Tiwari, A., & Murray, R. M. (2005). Estimation with information loss: Asymptotic analysis and error bounds. In *Proceedings of IEEE conf. on decision and control* (pp. 1215–1221).
Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M., & Sastry, S. (2004). Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, *49*(9), 1453–1464.
Xie, L., & Xie, L. (2008). Stability of a random riccati equation with markovian binary switching. *IEEE Transactions on Automatic Control*, *53*(7), 1759–1764.
Yaz, E., & Ray, A. (1996). Linear unbiased state estimation for random models with sensor delay. In *Proceedings of IEEE conf. on decision and control* (pp. 47–52).
Zhang, H., Xie, L., Zhang, D., & Soh, Y. (2004). A re-organized innovation approach to linear estimation. *IEEE Transactions on Automatic Control*, *49*(10), 1810–1814.
Zhang, H., & Xie, L. (2007). *Control and estimation of systems with input/output delays*. Springer.
ZigBee. (2009). www.zigbee.org.

Ling Shi received his B.S. degree in Electrical and Electronic Engineering from the Hong Kong University of Science and Technology in 2002 and Ph.D. degree in Control and Dynamical Systems from California Institute of Technology in 2008. He is currently an Assistant Professor at the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology. His research interests include networked control systems, wireless sensor networks and distributed control.

Lihua Xie received the B.E. and M.E. degrees in Electrical Engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in Electrical Engineering from the University of Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Professor and the Director, Centre for Intelligent Machines. He held teaching position in the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989.

Dr Xie's research interests include robust control and estimation, networked control systems, time delay systems, control of disk drive systems and sensor networks. In these areas, he has published many journal papers and co-authored two patents and the books *H-infinity Control and Filtering of Two-dimensional Systems* (with C. Du); *Optimal and Robust Estimation* (with F.L. Lewis and D. Popa) and *Control and Estimation of Systems with Input/Output Delays* (with H. Zhang). He is an Associate Editor of Automatica, IEEE Transactions on Control Systems Technology, the Transactions of the Institute of Measurement and Control, and an Associate Editor of Journal of Control Theory and Applications, and is also a member of the Editorial Board of IET Proceedings on Control Theory and Applications. He served as an Associate Editor of IEEE Transactions on Automatic Control from 2005 to 2007, IEEE Transactions on Circuits and Systems-II from 2006 to 2007, International Journal of Control, Automation and Systems from 2004 to 2006, and the Conference Editorial Board, IEEE Control Systems Society from 2000 to 2005. He was the General Chairman of the 9th International Conference on Control, Automation, Robotics and Vision. Dr Xie is a Fellow of IEEE.

Richard M. Murray received the B.S. degree in Electrical Engineering from California Institute of Technology in 1985 and the M.S. and Ph.D. degrees in Electrical Engineering and Computer Sciences from the University of California, Berkeley, in 1988 and 1991, respectively. He is currently the Thomas E. and Doris Everhart Professor of Control and Dynamical Systems and the Director for Information Science and Technology at the California Institute of Technology, Pasadena. Murray's research is in the application of feedback and control to mechanical, information, and biological systems.