Available online at www.sciencedirect.com

SCIENCE DIRECT®

automatica

Automatica 42 (2006) 1441–1452

www.elsevier.com/locate/automatica

State estimation over packet dropping networks using multiple description coding☆

Zhipu Jin*, Vijay Gupta, Richard M. Murray

Division of Engineering and Applied Science, California Institute of Technology, Pasadena, USA

Received 28 July 2005; received in revised form 19 January 2006; accepted 26 March 2006
Available online 9 June 2006

Abstract

For state estimation over a communication network, efficiency and reliability of the network are critical issues. The presence of packet dropping and communication delay can greatly impair our ability to measure and predict the state of a dynamic process. In this paper, multiple description (MD) codes, a type of network source codes, are used to compensate for this effect on Kalman filtering. We consider two packet dropping models: in one model, packet dropping occurs according to an independent and identically distributed (i.i.d.) Bernoulli random process and in the other model, packet dropping is bursty and occurs according to a Markov chain. We show that MD codes greatly improve the statistical stability and performance of Kalman filter over a large set of packet loss scenarios in both cases. Our conclusions are verified by simulation results.

Keywords: State estimation; Packet dropping; Kalman filter; Multiple description coding

1. Introduction

One of most significant challenges for control theory today is that the control objective is being enlarged from single physical systems to large-scale, complex systems and networks (Murray, 2002). Tremendous amounts of information need to be sensed, processed and transmitted among different subsystems. Examples include congestion control in the Internet, optimal operation of power grid, air traffic control networks, and many others. Communication networks play an important role in these examples. The theory of networked control systems (NCS) provides an approach to investigate the impact of communication constraints on feedback control systems by replacing the "ideal" feedback links with communication networks. Fig. 1 shows a simplified version of a NCS that omits the communication channel from the controller to the dynamic system. The link from observer to estimator is not modelled as a single, exclusive communication channel, but rather as a possible path through a large, complex communication network shared with many other users.

Efficient and reliable communication requires improvements in both source and channel coding. In addition, dynamical evolution of the system and a prior knowledge of the dynamics can give us extra benefits on top of just using current state-of-the-art communication theory and technology. This merger between control and communication has received considerable interest recently. Some works (Liberzon, 2003; Matveev & Savkin, 2005; Tatikonda & Mitter, 2004a, 2004b; Walsh, Ye, & Bushnell, 2002; Wong & Brockett, 1997) have focused on answering a fundamental question: how much information at least do we need to achieve stability? The main idea in these works is that the uncertainty of the dynamic system changes with respect to time. In order to stabilize the system, the minimum feedback information must be enough to compensate for the increase in the uncertainty. The sensitivity of the feedback system to quantization noise has also been studied (Brockett & Liberzon, 2000; Elia & Mitter, 2001). It has been noticed that feedback information can be useful even with different levels of resolution.

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Ioannis Paschalidis under the direction of Editor Ian Petersen.

* Corresponding author. Tel.: +1 626 395 3367; fax: +1 626 395 6170.
E-mail addresses: jzp@caltech.edu (Z. Jin), vijay@cds.caltech.edu (V. Gupta), murray@caltech.edu (R.M. Murray).

0005-1098/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2006.03.020

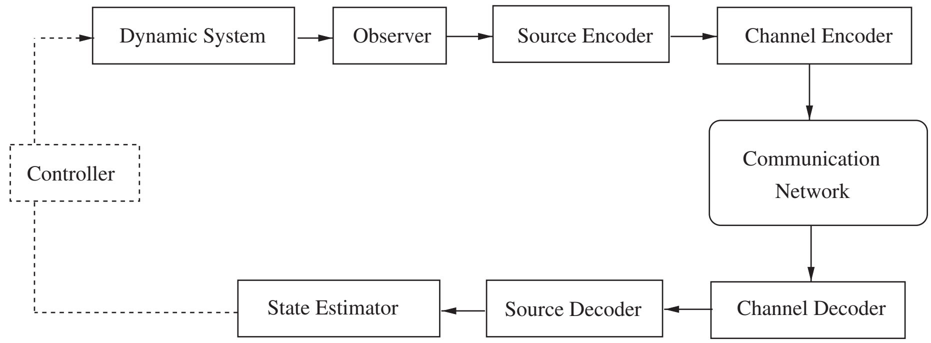

Fig. 1. Diagram of a networked control system.

Most modern communication networks are digital and are implemented using packet-based protocols. Thanks to the incredible developments in communication technology, the ratio of cost to bandwidth of communication links has dropped dramatically. However, the reliability of communication links in networks has become an important issue. Communication networks are not used exclusively for transmitting information between two single points. Packets have to be dropped whenever the network becomes congested. Stochastic packet dropping is very common in large-scale networks (Yajnik, Moon, Kurose, & Towsley, 1999). For widely used transmission control protocol (TCP), dropped packets are resent after certain delays. Using "recursive state estimator" (Matveev & Savkin, 2003) can generate minimum variance estimates in the presence of irregular communication delays. However, extra memory and computation costs are incurred. A multi-vehicle wireless testbed (MVWT) was built in Caltech (Jin et al., 2004) and user datagram protocol (UDP) was adopted to transmit data over a local wireless communication network. In this paper, according to our experiences on MVWT, we assume estimator and controller only use available new, "real-time" data packets to update estimation and control law. This assumption makes the model simple yet sophisticated enough such that we can focus on the effects of different coding schemes.

In this paper, we ignore the controller in Fig. 1 and focus on the problem of state estimation. Rather than worrying about limited bandwidth, we are concerned about the fact that packets can be dropped by the communication network. Our goal is to understand how the packet dropping affects state estimation and what we can do to compensate for this unreliability?

There are two popular models for packet dropping in large-scale networks. The Bernoulli model (Yajnik et al., 1999) describes packet losses according to an independent and identically distributed (i.i.d.) Bernoulli random process. Another model is the Gilbert–Elliott model (Elliott, 1963; Gilbert, 1960) which describes packet dropping as a Markov chain and is used to handle bursty packet dropping. Sinopoli et al. (2004) used the Bernoulli model to study the statistical convergence properties of the estimation error covariance in a Kalman filter by solving a modified algebraic Riccati equation (MARE). That work showed that packet dropping degrades the performance of Kalman filter. Liu and Goldsmith (2004) extended the results to the case with partial observation losses in sensor networks. In this paper, we show that multiple description (MD) source codes, a type of network source codes (Fleming, Zhao, & Effros, 2004), can be used to compensate for the unreliability of communication networks. MD codes have been studied in information theory for over 30 years (Gamal & Cover, 1982; Goyal & Kovacevic, 2001) and successfully used in transmission of real-time speech and audio/video over the internet (Goyal, Kovaceric, Arean, & Vetterli, 1998; Lee, Pickering, Frater, & Arnold, 2000). The efficiency of MD codes has been proved in situations where data can be used at various resolution levels. To the best of our knowledge, our work is the first to apply such a coding scheme to NCS.

This paper is organized as follows. In Section 2, state estimation problem in NCS is described and assumptions on communication networks are given. In Section 3, MD source codes are introduced and theoretical limits are discussed. We formulate the state estimation problem with MD coding in Section 4 and present results for the i.i.d. Bernoulli model. In addition, examples and simulation results are listed. Then the same estimation problem is studied for the Markov chain model in Section 5 and conclusions are summarized in Section 6.

2. State estimation problem and assumptions

We study the state estimation problem for the following discrete-time linear dynamic system:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \\ y_k &= Cx_k + v_k, \end{aligned} \tag{1}$$

where $x_k \in \mathscr{R}^n$ is the state vector, $y_k \in \mathscr{R}^m$ is the output vector, w_k and v_k are Gaussian white noise vectors with zero mean and covariance matrices are $Q \geqslant 0$, and $G > 0$, respectively. We assume that A is unstable and a standard discrete-time Kalman filter is used as the estimator. It is well known that if the pair $(A, Q^{1/2})$ is controllable, the pair (A, C) is detectable, and no measurements are lost, the estimation error covariance of Kalman filter converges to a unique value from any initial condition.

For the NCS in Fig. 1, the observation data is put into data packets and is sent through the communication network after going through source and channel encoders. We list some assumptions for the network which simplify our problem:

- We ignore channel coding and assume that the packet will be either received and decoded successfully at the end of the links or totally lost.
- For the estimator, only the new, "real-time" data is used for each update cycle. We only consider the transmission delay that is determined by the network bandwidth and length of the packet. If a packet arrives too late, it is discarded and treated as a dropped packet. So the packet transmission is in a "UPD-like" style which means that lost packets are not re-sent.
- The network does not provide preferential treatment to any packet. In other words, the network treats each single packet equally without inspecting the content. Thus, a multiple resolution code or a layered source code is not a good choice for us since they mark packets with different priorities according to the contents.
- There is no feedback from the decoder to the encoder.
- We assume that number of bits in each data packet is relatively large and the network is running at a high bit rate scenario.

We assume that packet dropping happens according to one of two following models:

- The i.i.d. Bernoulli model. A Bernoulli random variable γ_k indicates whether the packet k is received correctly. If it goes through the network successfully, then $\gamma_k=1$, otherwise, $\gamma_k=0$. For any value of k, γ_k is i.i.d with probability distribution $P(\gamma_k=1)=\lambda$ and $P(\gamma_k=0)=(1-\lambda)$. This is the simplest and often used model for packet dropping in large-scale networks.
- The Gilbert–Elliott model. This model considers the network as a discrete-time Markov chain with two possible states: "good" and "bad". In the "good" state, the packet is received correctly, and in the "bad" state, the packet is dropped. The network jumps between these two states according to a Markov chain with transition probability matrix $\mathcal{Q}$ as

$$\mathcal{Q}=\begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}, \tag{2}$$

  where 1 is the good state, 0 is the bad state, and q_{ij} is the probability from the previous state j to the next state i. The model can be easily extended to more possible states with different packet dropping probabilities. However, for simplicity and without loss of generality, we consider the two-state model in this paper. Unlike the first model, this one is able to capture the dependence between consecutive losses, i.e., bursty packet dropping.

3. Multiple description source coding

For NCS, the traditional source code is actually a quantizer $q:\mathscr{R}^m \to \mathscr{Z}$ with a state space partition set $\{V_i\}$ where $V_i \cap V_j = \emptyset$ for any $i, j \in [1, 2, \ldots, N]$ with $i \neq j$, and $\bigcup_{i=1}^{N} V_i = \mathscr{R}^m$. For each partition V_i, there exists a centroid $v_i \in V_i$ and the set of all the representatives of v_i is called a codebook. The encoder functions are $f_{\mathrm{e}}(x)=i$ if $x \in V_i$ and decoder functions are $f_{\mathrm{d}}(i)=v_i$. We call the integer i the description of the state x. The distortion function at the decoder is defined as $d(x, v_i)=\|x-v_i\|^2$. Generally, if N is bigger, each partition will be smaller and the average distortion at decoder side is smaller. However, the cost is sending more bits through the network. Rate distortion theory, a part of information theory, is used to study any possible partition set and the corresponding average distortion. In this section, we focus on uniform scalar quantizers and assume that the state of the dynamic system is uniformly distributed on a state space whose length is L. The optimal distortion rate function for single description source codes is

$$D(R) \leqslant \frac{L^2}{12} \cdot 2^{-2R}, \tag{3}$$

where $R=\log_2(N)$ is the bits per source sample (bpss) and N is the number of the quantization levels. For other state distributions, we have similar distortion rate functions that all decay at the speed 2^{-2R}.

MD source codes are designed to achieve good rate-distortion performance over lossy links. The unique feature of MD codes is that instead of using one single description to represent one source sample, MD codes use two or more descriptions. So at the end of the link, the decoder has much less chance of losing all descriptions. The distortion at the decoder depends on how many descriptions it receives and could be at various quality levels. Also we would like to keep the total bpss as small as possible. Thus, the design of a MD code is a problem of minimizing the size of the code over the redundancy between the descriptions. Moreover, MD codes need to be non-hierarchical so that the receiving order of descriptions is not important.

3.1. Theoretical limits of multiple description codes

In this subsection, we introduce some theoretical limits for MD codes that fit our discussion on NCS. We start with a two-description MD code. An encoder is fed by a sequence of source sample values $\{X_k\}$. The output of the encoder is $\{i_k, j_k\}$. The number of bits for descriptions are R_i and R_j. There are three cases according to which descriptions are received:

- At step k, the decoder receives none of the descriptions, we call this the "broken link" case. We will discuss this case in Section 4.
- At step k, the decoder receives both $\{i_k\}$ and $\{j_k\}$, we call this the "central decoder" case. In this case, the average distortion is D_{c}.
- At step k, the decoder only receives either $\{i_k\}$ or $\{j_k\}$, we call this the "side decoder" case. The average distortions are D_i and D_j, respectively.

The main theoretical problem of MD coding is to determine the achievable quintuple $(R_i, R_j, D_{\mathrm{c}}, D_i, D_j)$. The fundamental tradeoff in MD coding is making descriptions

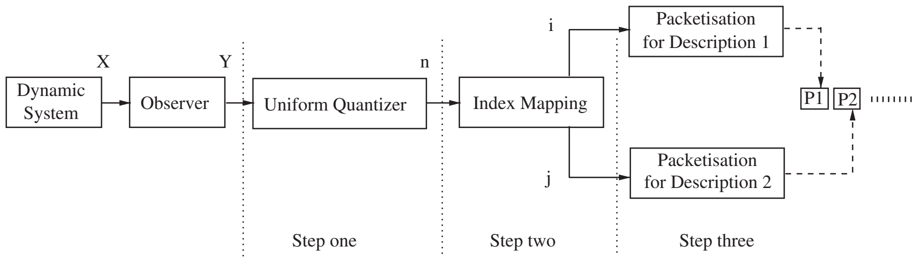

Fig. 2. Diagram of two-description MD source coding for NCS.

individually good and sufficiently different at the same time (Gamal & Cover, 1982). If $R_i = R_j$ and $D_i \approx D_j$, then we say the MD code is balanced. The achievable rate-distortion region of a two-description MD code for a memoryless unit variance Gaussian source with mean-squared error (MSE) distortion has been given (Ozarow, 1980):

$$\begin{cases} D_i \geqslant 2^{-2R_i}, \\ D_j \geqslant 2^{-2R_j}, \\ D_c \geqslant 2^{-2(R_i+R_j)} \cdot \gamma(D_i, D_j, R_i, R_j), \end{cases} \quad (4)$$

where

$$\gamma = \frac{1}{1 - \left(\sqrt{(1-D_i)(1-D_j)} - \sqrt{D_i D_j - 2^{-2(R_i+R_j)}}\right)^2}$$

for $D_i + D_j < 1 + 2^{-2(R_i+R_j)}$ and $\gamma = 1$ otherwise. For packet-based NCS, we use balanced MD codes and assume that $R_i = R_j = R \gg 1$ and $D_i = D_j = 2^{-2R(1-\alpha)} \ll 1$ with $0 < \alpha < 1$ (Vaishampayan & Batllo, 1998), then we get

$$\begin{aligned} \frac{1}{\gamma} &= 1 - \left((1-D_i) - \sqrt{D_i^2 - 2^{-4R}}\right)^2 \\ &\approx 1 - ((1-D_i) - D_i)^2 \approx 4D_i, \end{aligned}$$

and

$$D_c \cdot D_i \gtrapprox \tfrac{1}{4} 2^{-4R}. \quad (5)$$

Inequality (5) shows the tradeoff between central and side distortions. Compared with inequality (3), it is clear that the penalty in the exponential rate of decay of D_i is exactly the increase in the rate of decay of D_c.

3.2. Multiple description scalar quantization

In this section we present the actual design method for MD scalar quantizers (MDSQ). In NCS, each source sample is an observation of the dynamic system and MDSQ give two descriptions for each source sample. This approach was proposed and popularized by Vaishampayan (1993) and Fig. 2 shows the diagram of a two-description MDSQ:

- *Step one*: Select a uniform quantizer with an appropriate step size Δ and the number of step levels N. A source sample Y is quantized by rounding off to the nearest multiple of a step size Δ and the index output of the uniform quantizer is n which satisfies $0 < n \leqslant N$.
- *Step two*: The index n is assigned a pair of indexes (i, j) by using an index mapping matrix.
- *Step three*: The i and j are entropy coded and put into data packets, respectively.

Table 1
MSE for different MD codes

Description	No loss	Lost 1	Lost 2	bpss
One	8.33×10^{-6}	N/A	N/A	10
Two	8.33×10^{-6}	1.56	N/A	12
Three	8.33×10^{-6}	0.00441	1.53	15
One	4.97×10^{-7}	N/A	N/A	12
Two	8.33×10^{-6}	1.56	N/A	12
Three	9.87×10^{-5}	0.0197	2.15	12

The index mapping problem in step two is the main part of MDSQ. We state this problem as follows: there exists a $\sqrt{M} \times \sqrt{M}$ matrix where $M \geqslant N$. We need to arrange all these numbers from $[1, \ldots, N]$ into the cells of the matrix. Each cell can hold one number at most and each index n gets a pair of matrix index (i, j) according to it's location, i.e., matrix index i is the row number and j is the column number. By this index mapping matrix, step two transfers each single description n into two descriptions i and j. Since $N \leqslant M$, the total number of the possible index assignment methods is $\sum_{n=1}^{M} M!/(M-n)!$. By choosing a proper index mapping matrix (refer to (Vaishampayan, 1993 for details), we get

$$\begin{cases} D_c \approx C_0 2^{-2R(1+\alpha)}, \\ D_i \approx D_j \approx C_1 2^{-2R(1-\alpha)}, \end{cases} \quad (6)$$

where C_0 and C_1 are constants that depend on the distribution of the initial state and uniform quantizer in step one. The parameter $\alpha \in [0, 1]$ is a pre-defined parameter that indicates the tradeoff between the decay speeds of D_c and D_i. It is clear that the "central decoder" equals a decoder for the uniform quantizer in step one.

The index mapping method can be extended to three-description MD codes. Table 1 lists some MD codes. Each of them provides 1000 codewords for a variable between −5 and 5. The first column of the tables indicates how many descriptions we use to represent one sample value. The first part of Table 1 shows some examples of average distortions for different description loss cases when we keep the central distortion constant. It shows that more bpss are needed in order to get the same central decoder distortion. The second part shows that, if we keep bpss constant, the distortion increases when the number of descriptions per sample increases. In the table, "lost k" means k descriptions have been lost, and "N/A" means not available. The table shows that MD coding actually provides various quality levels corresponding to how many descriptions the decoder receives.

Another issue about MD coding is that the computation complexity of decoding increases since the size of the codebook increases at the decoder side as the number of descriptions increases. For example, for a traditional uniform quantizer with N levels, the codebook for the corresponding L-description MD code has $(2^L-1)N$ elements. We need to consider this issue when choosing the number of the descriptions per source sample.

3.3. Quantization noise of MD codes

As discussed in Marco and Neuhoff (2005), the quantization noise of a uniform scalar quantizer with the assumptions of small partition cells, reproduction values at cell's midpoints, and large support region can be approximately modelled as an additive uncorrelated white noise to the quantizer input. For balanced MD codes, the central decoder case actually is a uniform scalar quantizer with the midpoints as the outputs and the average distortion is $D_c \approx \Delta^2/12$ where Δ is the length of partition cells. For the side decoder case, index mapping introduces a slight asymmetry between the two side distortions and causes a small increase in distortion. However, for large bpss, this asymmetry asymptotically disappears. According to previous analysis, we have

$$D_i \approx D_j \approx C_1 \cdot \left(\frac{1}{12C_0}\right)^{(1-\alpha)/(1+\alpha)} \cdot (\Delta^{(1-\alpha)/(1+\alpha)})^2.$$

For a balanced two-description MD code, α is a constant and D_i will be asymptotically negligible relative to $(\Delta^{(1-\alpha)/(1+\alpha)})^2$. So as long as the bit rate $R_i (=R_j)$ is big enough, the additive noise model is still a good approximation to represent the quantization noise in the side decoder case. From now on, we model the MD quantization noise as Gaussian white noise with zero mean and covariance D_c for central decoder case and D_i for side decoder case.

4. Kalman filtering utilizing MD with i.i.d. packet dropping

4.1. Problem formulation

We consider the discrete-time linear dynamical system described by Eq. (1) and assume that packet dropping is independent and is described by an i.i.d. Bernoulli random process. We use two-description balanced MD codes. Each y_k^l in the measurement output $Y_k = [y_k^1, \ldots, y_k^m]$ is encoded by two descriptions $\{i_k^l, j_k^l\}$. We organize these descriptions into two description vectors as $\{I_k, J_k\}$ and put them into two different packets. Variables $\gamma_{I,k}$ and $\gamma_{J,k}$ are used to indicate whether the description vectors I_k and J_k are received correctly. If I_k is received correctly, then $\gamma_{I,k}=1$, otherwise, $\gamma_{I,k}=0$, and similarly for $\gamma_{J,k}$. We assume that $\gamma_{I,k}$ and $\gamma_{J,k}$ are i.i.d. Bernoulli random variables with probability distribution $P(\gamma_{I,k}=1) = P(\gamma_{J,k}=1)=\lambda$.

Since $\gamma_{I,k}$ and $\gamma_{J,k}$ are independent, we have three measurement rebuilding scenarios. First, we may receive both the descriptions correctly. In this case, the measurement noise is the white noise v_t plus the central distortion noise. We use $G_0 = G + D_c$ to indicate the covariance where G is the observation noise covariance defined in (1) and D_c is the central distortion covariance. Second, we may receive only one description correctly and the measurement noise is $G_1 = G + D_i$ where D_i is the side distortion covariance. Third, we may receive none of the descriptions correctly. In this case, we assume the measurement is corrupted by an infinitely large noise. This is corresponding to the "broken link" case in Section 3. The noise is changed into a random variable $\hat{v}_t$ after the decoder and the covariance C_k is:

$$C_k = \begin{cases} G_0 & \text{with probability } \lambda^2, \\ G_1 & \text{with probability } 2(1-\lambda)\lambda, \\ \sigma^2 I & \text{with probability } (1-\lambda)^2, \end{cases} \tag{7}$$

where $\sigma \to \infty$.

The Kalman filter recursion thus becomes stochastic and the error covariance evolves as

$$\begin{aligned} P_{k+1} = {} & AP_kA' + Q \\ & - \gamma_{I,k}\gamma_{J,k}AP_kC'[CP_kC' + G_0]^{-1}CP_kA' \\ & - (1-\gamma_{I,k})\gamma_{J,k}AP_kC'[CP_kC' + G_1]^{-1}CP_kA' \\ & - \gamma_{I,k}(1-\gamma_{J,k})AP_kC'[CP_kC' + G_1]^{-1}CP_kA'. \end{aligned}$$

Thus, the sequence of the error covariance matrix $P_{k=0}^{\infty}$ is a random process for any given initial value. Using the same approach as in Sinopoli et al. (2004), we define the MARE for Kalman filter using balanced two-description MD codes as:

$$\begin{aligned} g_\lambda(X) = {} & AXA' + Q \\ & - \lambda^2 AXC'(CXC' + G_0)^{-1}CXA' \\ & - 2(1-\lambda)\lambda AXC'(CXC' + G_1)^{-1}CXA' \end{aligned} \tag{8}$$

and the expected value of error covariance matrix $E[P_k]$ evolves according to this MARE.

4.2. Statistical convergence properties

This subsection lists theorems which describe the convergence properties of the MARE in Eq. (8). These theorems are based on the lemmas in Appendix A. The first theorem listed below states the uniqueness of the MARE solution.

Theorem 1. *Consider the operator*

$$\begin{aligned} \phi(K_0, K_1, X) = {} & (1-\lambda)^2(AXA' + Q) \\ & + \lambda^2(F_0XF_0' + V_0) \\ & + 2(1-\lambda)\lambda(F_1XF_1' + V_1) \end{aligned} \tag{9}$$

where $F_0 = A + K_0C$, $F_1 = A + K_1C$, $V_0 = Q + K_0G_0K_0'$, *and* $V_1 = Q + K_1G_1K_1'$. *Suppose there exist* K_0, K_1, *and* $P > 0$

such that $P > \phi(K_0, K_1, P)$. *Then, for any initial condition* $P_0 \geqslant 0$, *the iteration* $P_{k+1} = g_{\underline{\lambda}}(P_k)$ *converges to the unique positive semi-definite solution* $\bar{P}$ *of MARE* (8), *i.e.*,

$$\lim_{k\to\infty} P_k = \lim_{k\to\infty} g_\lambda^k(P_0) = \bar{P} \geqslant 0,$$

where $\bar{P} = g_\lambda(\bar{P})$.

Proof. First, we show that the MARE converges with initial value $Q_0 = 0$. Let $Q_k = g_\lambda(Q_{k-1}) = g_\lambda^k(Q_0)$, then $Q_1 \geqslant Q_0 = 0$ and

$$Q_1 = g_\lambda(Q_0) \leqslant g_\lambda(Q_1) = Q_2.$$

By induction, we know that the sequence $\{Q_k\}$ is non-decreasing. Also by Lemma 13 (most lemmas are listed in the Appendix), $\{Q_k\}$ is bounded and there exists an M_{Q_0} such that $Q_k \leqslant M_{Q_0}$ for any k. Therefore, the sequence converges and

$$\lim_{k\to\infty} Q_k = \bar{P} \geqslant 0,$$

where $\bar{P}$ is a fixed point of the iteration $\bar{P} = g_\lambda(\bar{P})$.

Next we show that the iteration $\bar{G}_k = g_\lambda^k(\bar{G}_0)$ initialized at $G_0 \geqslant \bar{P}$ also converges to $\bar{P}$. Since $G_1 = g_\lambda(G_0) \geqslant g_\lambda(\bar{P}) = \bar{P}$, $G_k \geqslant \bar{P}$ for any k. Also

$$\begin{aligned} 0 \leqslant G_{k+1} - \bar{P} &= g_\lambda(G_k) - g_\lambda(\bar{P}) \\ &= \phi(K_{G_k0}, K_{G_k1}, G_k) - \phi(K_{\bar{P}0}, K_{\bar{P}1}, \bar{P}) \\ &\leqslant \phi(K_{\bar{P}0}, K_{\bar{P}1}, G_k) - \phi(K_{\bar{P}0}, K_{\bar{P}1}, \bar{P}) \\ &= \hat{\mathscr{L}}(G_k - \bar{P}), \end{aligned}$$

where $\hat{\mathscr{L}}$ has a similar form as the operator $\mathscr{L}$ in Lemma 12. Note that

$$\bar{P} = g_\lambda(\bar{P}) > \hat{\mathscr{L}}(\bar{P}),$$

thus, $\hat{\mathscr{L}}$ meets all the conditions in Lemma 12. Using the same argument, we have, for any $Y \geqslant 0$,

$$\lim_{k\to\infty} \hat{\mathscr{L}}^k(Y) = 0.$$

So we get $0 \leqslant \lim_{k\to\infty}(G_k - \bar{P}) = 0$, i.e., the sequence G_k converges to $\bar{P}$.

As the last part, we show that, for any initial condition $P_0 \geqslant 0$, the iteration $P_k = g_\lambda^k(P_0)$ converges to $\bar{P}$. Let $G_0 = P_0 + \bar{P} \geqslant \bar{P}$, then $0 \leqslant Q_0 \leqslant P_0 \leqslant G_0$, by induction, we have $0 \leqslant Q_k \leqslant P_k \leqslant G_k$. Since $\{Q_k\}$ and $\{G_k\}$ converges to $\bar{P}$, $\{P_k\}$ also converges to $\bar{P}$ and the result follows. □

The following theorem states the conditions for MARE convergence.

Theorem 2. *If* $(A, Q^{1/2})$ *is controllable*, (A, C) *is detectable, and A is unstable, then there exists a* $\lambda_c \in [0, 1)$ *such that*

(a) *For* $0 \leqslant \lambda \leqslant \lambda_c$, *there exists some initial condition* $P_0 \geqslant 0$ *such that* $E[P_k]$ *diverges when* $k \to +\infty$, *i.e., there does not exist a matrix* M_{P_0} *such that* $E[P_k] \leqslant M_{P_0}$ *for any* $k > 0$;

(b) *For* $\lambda_c < \lambda \leqslant 1$, $E[P_k] \leqslant M_{P_0}$ *for any* $k > 0$ *and any initial condition* $P_0 \geqslant 0$;

where $M_{P_0} > 0$ *depends on the initial condition* P_0.

Proof. Please refer to Appendix A.

This theorem claims that there exists a critical value λ_c of the packet receiving probability. If λ is smaller than λ_c, MARE (8) does not converge and the expected value of error covariance matrix will diverge.

Theorem 3. *Let*

$$\begin{aligned} \underline{\lambda} &= \arg\inf_\lambda [\exists \hat{S} \geqslant 0 | \hat{S} = (1-\lambda)^2 A\hat{S}A' + Q] = 1 - \frac{1}{\alpha}, \\ \overline{\lambda} &= \arg\inf_\lambda [\exists \hat{X} \geqslant 0 | \hat{X} > g_\lambda(\hat{X})] \\ &= \arg\inf_\lambda [\exists (\hat{K}_0, \hat{K}_1, \hat{X} \geqslant 0) | \hat{X} > \phi(\hat{K}_0, \hat{K}_1, \hat{X})], \end{aligned}$$

where $\alpha = \max|\sigma_i|$ *and* σ_i *are the eigenvalues of A. Then*

$$\underline{\lambda} \leqslant \lambda_c \leqslant \overline{\lambda}. \tag{10}$$

Proof. Please refer to Appendix A.

This theorem states the upper and lower bounds for λ_c. The lower bound is in a closed form. According to the next theorem and corollary, we can reformulate the computation of $\overline{\lambda}$ as an LMI feasible problem.

Theorem 4. *Assume that* $(A, Q^{1/2})$ *is controllable and* (A, C) *is detectable, then the following statements are equivalent*:

(a) $\exists \bar{X} > 0$ *such that* $\bar{X} > g_\lambda(\bar{X})$;
(b) $\exists\ \bar{K}_0, \bar{K}_1$, *and* $\bar{X} > 0$ *such that* $\bar{X} > \phi(\bar{K}_0, \bar{K}_1, \bar{X})$;
(c) $\exists \bar{Z}_0, \bar{Z}_1$ *and* $0 < \bar{Y} \leqslant I$ *such that*

$$\Psi_\lambda(\bar{Y}, \bar{Z}_0, \bar{Z}_1) > 0,$$

where

$$\Psi_\lambda = \begin{bmatrix} Y & \varDelta(Y, Z_1) & \varOmega(Y, Z_0) & \varPi(Y) \\ \varDelta(Y, Z_1)' & Y & 0 & 0 \\ \varOmega(Y, Z_0)' & 0 & Y & 0 \\ \varPi(Y)' & 0 & 0 & Y \end{bmatrix},$$

$\varDelta(Y, Z_1) = \sqrt{2(1-\lambda)\lambda}(YA + Z_1 C)$, $\varOmega(Y, Z_0) = \lambda(YA + Z_0 C)$, *and* $\varPi(Y) = (1-\lambda)YA$.

When C is invertible, we choose $K_0 = K_1 = -AC^{-1}$ to make $F_0 = F_1 = 0$ and the LMI in Theorem 4 is equivalent to

$$X - (1-\lambda)^2 AXA' > 0.$$

Since the solution $X \geqslant 0$ exists if and only if $(1-\lambda)A$ is stable, i.e., all the magnitudes of eigenvalues of $(1-\lambda)A$ are smaller than 1, we obtain $\overline{\lambda} = \underline{\lambda} = (1/1-\alpha)$. According to Sinopoli et al. (2004), the lower bound of using single description codes is $1 - (1/\alpha^2)$ which is bigger than using MD codes. Also if either C is invertible or the quantization noise D_c and D_i are smaller than G (which is always true for the high bpss case), it is easy to show that the upper bound of using single description codes is also bigger than using MD codes. So using MD codes pushes λ_c to a smaller value and guarantee the convergence over a larger packet dropping scenario.

The following theorem gives the upper and lower bounds on the expected value of error covariance matrix when MARE converges. The lower bound $\overline{S}$ can be computed by standard Lyapunov equation solvers and the upper bound $\overline{V}$ can be either computed via iterating $V_{t+1}=g_\lambda(V_t)$ from any initial condition or transferred to a semi-definite programming (SDP) problem (Sinopoli et al., 2004).

Theorem 5. *Assume* $(A, Q^{1/2})$ *is controllable*, (A, C) *is detectable, and* $\overline{\lambda} < \lambda$, *then for any initial condition* $E[P_0] \geqslant 0$,

$$0 \leqslant \overline{S} \leqslant \lim_{k\to\infty} E[P_k] \leqslant \overline{V},$$

where $\overline{S}$ *and* $\overline{V}$ *are solutions of the equations* $\overline{S}=(1-\lambda)^2 A\overline{S}A' + Q$ *and* $\overline{V} = g_\lambda(\overline{V})$, *respectively.*

Proof. Let $S_{k+1} = \mathscr{M}(S_k) = (1-\lambda)^2 A S_k A' + Q$ and $V_{k+1} = g_\lambda(V_k)$ with initial conditions $S_0 = 0$ and $V_0 = E[P_0] \geqslant 0$. By induction and Theorem 3, we obtain

$$S_k \leqslant E[P_k] \leqslant V_k$$

for any t. According to Theorem 1, $\lim_{k\to\infty} V_k = \overline{V}$ where $\overline{V} = g_\lambda(\overline{V})$. Also since $(A, Q^{1/2})$ is controllable and all the magnitudes of the eigenvalues of $(1-\lambda)A$ are smaller than 1, the sequence of the Lyapunov iteration converges to the strictly positive definite solution of the Lyapunov function, i.e., $\lim_{k\to\infty} S_k = \overline{S} > 0$. Therefore, we can conclude that

$$0 < \overline{S} = \lim_{k\to\infty} S_k \leqslant \lim_{k\to\infty} E[P_k] \leqslant \lim_{k\to\infty} V_k = \overline{V}. \qquad \square$$

4.3. Simulation results

In this subsection simulation results are provided to verify the advantages of MD codes. We choose the discrete time LTI system with $A = -1.25$ and $C = 1$. The noises w_t and v_t have zero means and covariances $G = 2.5$ and $Q = 1$, respectively. A balanced two-description MD code is designed such that the central distortion $D_0 \approx 8.33 \times 10^{-6}$ and the side distortion $D_1 \approx 1.56$. The bpss of the MD code is 12 and the bpss of a single description code with the same distortion is 10 bits.

Fig. 3 shows the simulation results of the expected error covariance. The theoretical upper and lower bounds with or without MD codes are calculated according to Theorem 5 and the reference (Sinopoli et al., 2004), respectively. The simulations are run 1000 times and each simulation is run 2000 time steps. We use the average value $E[P_{2000}]$ as the expected error covariance. The asymptote λ_c has been pushed from 0.36 for the single description code to 0.2 for the MD code. Convergence properties of error covariance at high packet loss rate region are also improved dramatically. Note that when λ is close to the asymptote, some of the simulated error covariances values are below the lower bound. The reason is that we only take limited time steps for the simulation and residual effect of initial conditions remains.

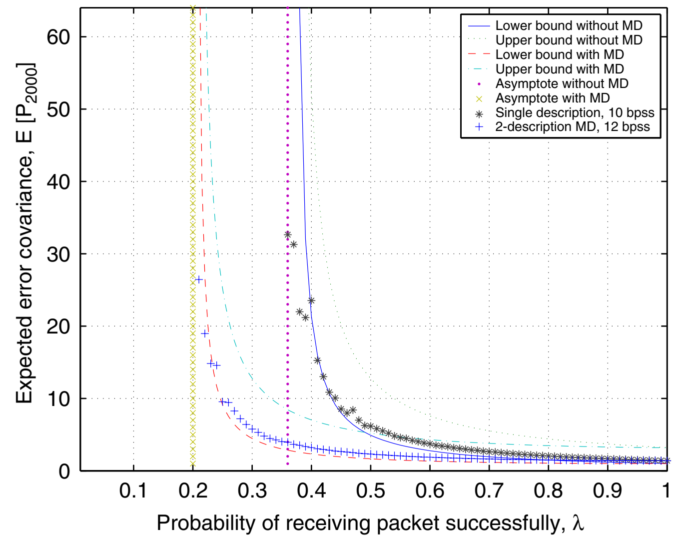

Fig. 3. Simulation results of expected error covariances with theoretical upper and lower bounds.

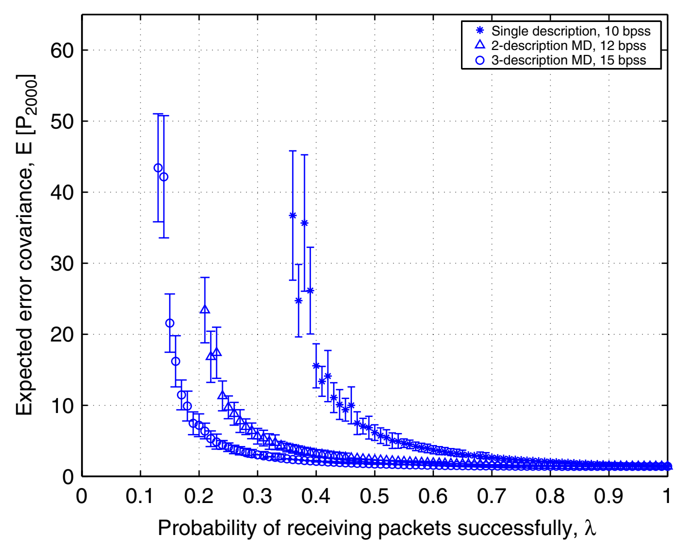

Fig. 4. Mean values of error covariance with same central distortions.

Fig. 4 shows some other simulation results. For each packet dropping rate, the centers of the error bars are the mean values and 95% of the simulation results are located inside the error bars. It shows that, if we use a balanced three-description MD code, the critical value λ_c is even smaller. So the benefits of using MD codes are clear and the cost we need to pay is more bits for each source sample. It can be shown that when C is invertible, using a L-description MD code we can get $\lambda_c = 1 - \alpha^{-2/L}$. Unfortunately, finding an optimal L-description MD code for arbitrary L is still an open problem in information theory. When we keep bpss constant, as shown in Fig. 5, we get bigger quantization noise as the number of descriptions increases.

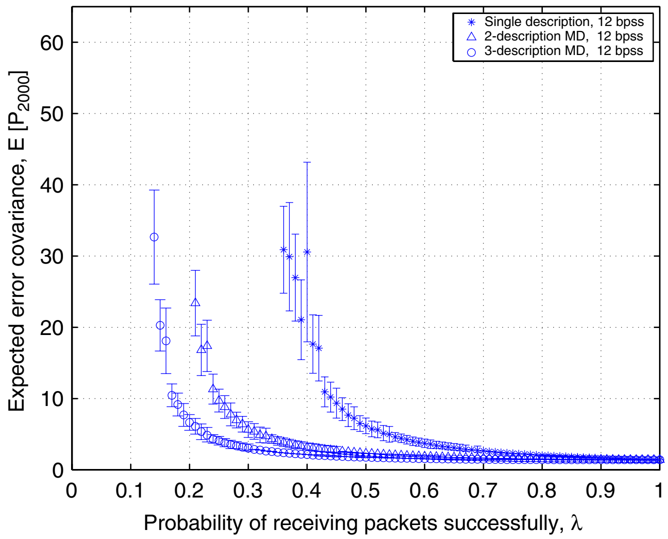

Fig. 5. Mean values of error covariance with same bpss.

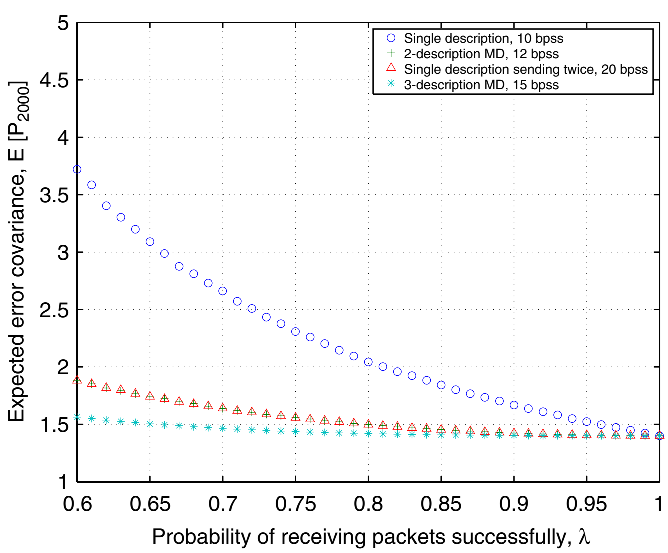

Fig. 6. Mean values of error covariance with low dropping rate.

Compared with the previous figure, there are no obvious differences due to the accuracy loss in this example.

In some cases, packet dropping rate of a practical communication network is fairly small. Fig. 6 shows the expected error covariance when packet dropping rate is low and MD codes give much better performance than the single description code. Note that the two-description MD code achieves as good performance as sending single description code twice but saves up to 40% bandwidth.

5. Kalman filtering utilizing MD with bursty packet dropping

5.1. Convergence conditions and boundaries

As mentioned in Section 2, a way to model the bursty packet dropping is using a two-state Markov chain with transition probability matrix $\mathcal{Q}$ given by Eq. (2). For the case of a balanced two-MD code, we are thus interested in a four-state Markov chain where the states correspond to both packets lost, only the 1st description packet lost, only the second description packet lost and no packet lost. The transition probability matrix of this chain is given as

$$\mathcal{Q}_{MD} = \begin{bmatrix} q_{00}^2 & q_{00}^2 & q_{01}q_{00} & q_{01}q_{00} \\ q_{01}q_{10} & q_{01}q_{10} & q_{11}q_{01} & q_{11}q_{01} \\ q_{10}q_{00} & q_{10}q_{00} & q_{01}q_{10} & q_{01}q_{10} \\ q_{10}q_{11} & q_{10}q_{11} & q_{11}^2 & q_{11}^2 \end{bmatrix}. \quad (11)$$

Note that the state in which both description packets are lost is equivalent to no observation coming through, while the other states correspond to the system being observed. If the Markov chain is stationary, the state probabilities tend to a stationary distribution as $k \to \infty$. However, we normally cannot directly study the problem over the steady-state distribution since this distribution might not be achievable (Gupta, Chung, Hassibi, & Murray, 2006). Mathematically, this problem is the same as the random sensor selection problem in sensor networks. Consider the system

$$x_{k+1} = Ax_k + w_k$$

being observed through n sensors with the ith sensor of the form

$$y_k^i = C^i x_k + v_k^i. \quad (12)$$

Suppose only one sensor can be active at any time instant and the choice of the sensor is done according to a Markov chain. We denote the Ricatti update in error covariance by $f_i(.)$ when the ith sensor is used and denote

$$f_i^k(.) = \underbrace{f_i(f_i(\cdots(.)\cdots))}_{k\text{times}}.$$

The expected error covariance at time step k is denoted by $E[P_k]$. Probability of the network in Markov state j at time k is denoted by π_k^j and q_{ij} is the probability of the network state is i at time $k+1$ given the network state is j at time k.

Lemma 6. *For any Ricatti update operator $f_i(P)$, we have*

(a) $f_i(P) \geqslant Q$;
(b) *If $X < Y$, then $f_i(X) \leqslant f_i(Y)$;*
(c) *$f_i(P)$ is concave w.r.t. P.*

When a single description code is applied, according to Eq. (12), packet dropping can be treated as the observation jumps between two sensors which have the same C^i matrices and different Gaussian noises with covariance G_0 and $\sigma^2 I$, respectively where $\sigma \to \infty$. The Kalman filter error covariance updates are

$$\begin{cases} f_0(P) &= APA' + Q, \\ f_1(P) &= APA' + Q - APC'(CPC' + G_0)^{-1}CPA'. \end{cases}$$

Similar to the i.i.d. Bernoulli model, we discuss the conditions and the upper/lower bounds for expected values of estimation error covariances converging.

Theorem 7. *When using a single-description code and with the Markov probability transition matrix given by Eq.* (2), *the lower bound for* $E[P_k]$ *is* Y_k, *where*

$$Y_k = q_{00}^k \pi_0^0 f_0^k\,(P_0) + \pi_k^1 f_1(Q) + \sum_{i=1}^{k-1} q_{00}^i (\pi_{k+1-i}^0 - q_{00}\cdot \pi_{k-i}^0) f_0^i(Q). \tag{13}$$

The upper bound is X_k, *where*

$$X_k = \sum_{j=0}^{1}\sum_{i=0}^{1} f_j(X_{k-1}^i) q_{ji}\pi_{k-1}^i \tag{14}$$

and $X_{k-1}^i = E[X_{k-1}|$ *state is i at time* $(k-2)]$.

Proof. Suppose k starts from 1, and for any k, we define event E_i as last packet was received at time $k-i$ where $i \in [0, \ldots, k]$. So the probability of E_i is

$$p_i = \begin{cases} q_{00}^k \pi_0^0, & i = k, \\ q_{00}^{i-1} q_{01} \pi_{k-i}^1, & 0 < i < k, \\ \pi_k^1, & i = 0 \end{cases}$$

and the error covariance P_k if E_i happens is

$$P_k|E_i = \begin{cases} f_0^k(P_0), & i = k, \\ f_0^i(f_1(P_{k-i})), & 0 < i < k, \\ f_1(P_k), & i = 0. \end{cases}$$

So

$$\begin{aligned} E[P_k] &= \sum_{i=0}^{k} p_i \cdot P_k|E_i \\ &= q_{00}^k \pi_0^0 f_0^k(P_0) + \sum_{i=1}^{k-1} q_{00}^i q_{01} \pi_{k-i}^1 f_0^i(f_1(P_{k-i})) \\ &\quad + \pi_k^1 f_1(P_k). \end{aligned}$$

According to Lemma 6, $f_1(P_{k-1}) \geqslant Q$, so

$$\begin{aligned} E[P_k] &\geqslant q_{00}^k \pi_0^0 f_0^k(P_0) + \pi_k^1 f_1(Q) \\ &\quad + \sum_{i=1}^{k-1} q_{00}^i q_{01} \pi_{k-i}^1 f_0^i(Q) \\ &= q_{00}^k \pi_0^0 f_0^k\,(P_0) + \pi_k^1 f_1(Q) \\ &\quad + \sum_{i=1}^{k-1} q_{00}^i \left(\pi_{k+1-i}^0 - q_{00}\cdot \pi_{k-i}^0\right) f_0^i\,(Q). \end{aligned}$$

For upper bound, let us denote S_k is the network state at time k. For single description code, $S_k \in [0, 1]$. Then $E[P_k] = \sum_{j=0}^{1} \pi_k^j \cdot E[P_k|S_k = j]$. Also

$$\begin{aligned} \pi_k^j \cdot E[P_k|S_k = j] &= \pi_k^j \sum_{i=0}^{1} E[P_k|S_k = j, S_{k-1} = i] \cdot p(S_{k-1} = i|S_k = j) \\ &= \sum_{i=0}^{1} E[f_j(P_{k-1})|S_{k-1} = i] q_{ji} \pi_{k-1}^i \\ &\leqslant \sum_{i=0}^{1} f_j([P_{k-1}|S_{k-1} = i]) q_{ji} \pi_{k-1}^i \end{aligned}$$

since $f_j(\cdot)$ is concave. □

Proposition 8. *A sufficient condition for divergence of expected error covariance is*

$$q_{00}\cdot \alpha^2 > 1, \tag{15}$$

where $\alpha = \max|\sigma_i|$ *and* σ_i *are the eigenvalues of* A.

Using a balanced two-MD code, the corresponding sensor selection problem has four sensors which have same C matrices and noise covariances are G_0, G_1, G_1, and $\sigma^2 I$, respectively. The Ricatti updates are

$$\begin{cases} f_0(P) = APA' + Q, \\ f_1(P) = APA' + Q - APC'(CPC' + G_1)^{-1}CPA', \\ f_2(P) = f_1(P), \\ f_3(P) = APA' + Q - APC'(CPC' + G_0)^{-1}CPA'. \end{cases}$$

Using the same approach, we get

Proposition 9. *When using two-description code and with the underlying Markov probability transition matrix given by* (2), *the lower bound for* $E[P_k]$ *is* Y_k, *where*

$$Y_k = q_{00}^{2k} \pi_0^0 f_0^k(P_0) + \sum_{j=1}^{3} \pi_k^j f_j(Q) + \sum_{i=1}^{k-1} q_{00}^{2i} (\pi_{k+1-i}^0 - q_{00}^2 \cdot \pi_{k-i}^0) f_0^i(Q). \tag{16}$$

The upper bound is X_k *where*

$$X_k = \sum_{j=0}^{3}\sum_{i=0}^{3} f_j(X_{k-1}^i) q_{ji}\pi_{k-1}^i \tag{17}$$

and $X_{k-1}^i = E[X_{k-1}|$ *state is i at time* $(k-2)]$.

A sufficient condition for divergence of expected error covariance is:

$$q_{00}\cdot \alpha > 1. \tag{18}$$

The equations for lower and upper bounds are pretty messy but can be calculated iteratively. Also, these bounds are dependent on value of q_{11} and initial distribution of packet dropping π_0^i.

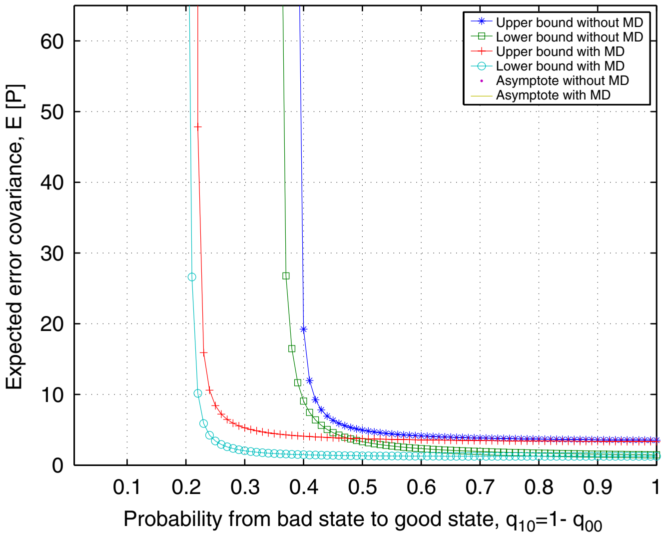

Fig. 7. Theoretical upper and lower bounds for burst packet dropping case with $q_{11} = 95\%$.

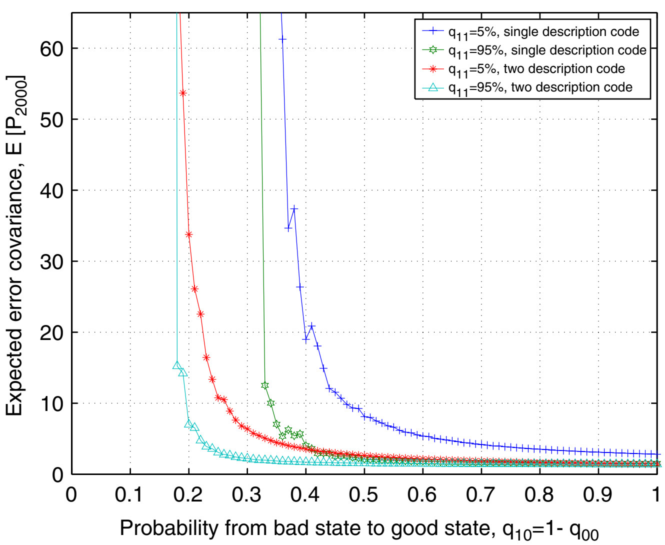

Fig. 8. Simulation results for burst packet dropping case.

5.2. Simulation results

We use the same LTI system as in Section 4. In Fig. 7 we plot the theoretical upper and lower bounds for the error variance as a function of q_{10} under the conditions as $q_{11} = 0.95$ and uniform distribution of π_0^i. The lowering of the bounds is indicative of the performance getting better with MD codes. The simulation results with parameters $q_{11} = 0.05$ and $q_{11} = 0.95$ with different coding schemes are shown in Fig. 8.

In the Markov chain model, $q_{01} = 1 - q_{11}$ is a measure of the frequency of the bursty dropping while $q_{00} = 1 - q_{10}$ indicates how big the burst dropping is. According to the simulation results, expected error covariance diverges more quickly with higher q_{01}. This makes sense since if the error bursts happen often, we expect the estimation error covariance to diverge. It is seen from the figures that the system diverges around $q_{10} = 0.36$ with the single description code case, and diverges around $q_{10} = 0.2$ with the two-MD code. (Limited simulation time steps make the results smaller than lower bounds near asymptotes.) Thus, the stability margin is enlarged if we use MD codes. Also, for same q_{10}, using MD codes greatly decreases the expected error covariance.

6. Conclusion

In this paper, we use the MD coding scheme to counteract the effect of packet dropping on state estimation problem. The accuracy of the decoder only depends on how many descriptions are successfully received. We considered two typical network packet dropping models: the i.i.d. Bernoulli model and the Gilbert-Elliott model. Using MD codes, the convergence region of the estimation error covariance is much larger than using traditional single description code and the steady expected values are also much smaller. Moreover, MD code is an optimal code which saves considerable bandwidth than sending duplicated packets.

The main goal of this paper is trying to understand the impact of communication constraints from another angle: in high bit rate scenario and with large, complex communication networks, what can we do to compensate for packet loss as well as to satisfy the real-time demands? In this paper, we have to compromise the accuracy of the source code to improve the convergency properties of MARE. This is a good demonstration to indicate the close relationship between communication theory and control theory when we face the challenges in "an information rich world" (Murray, 2002).

There are several promising research directions for the future. From the communication theory side, a more general theory and design method for MD coding for arbitrary number of description is needed. Also, since using MD codes greatly increases the computation complexity of the decoder, a more efficient search algorithm for decoding is needed. From the feedback control theory side, the stability and robustness of close-loop NCS with MD coding need to be studied.

Acknowledgments

The authors would like to thank Prof. M. Effros for the inspiring ideas on MD coding. They also want to thank B. Hassibi for valuable discussions and comments. This work is supported in part by AFOSR Grant FA9550-04-1-0169 and NSF-ITR Grant CCR-0326554.

Appendix A. Lemmas and some proofs of MARE convergence with i.i.d. packet dropping

Those theorems in Section 4 are based on these lemmas which can be easily proved by using similar approach in (Sinopoli et al., 2004). We list the lemmas here and omit the proofs.

Lemma 10. *Consider operator*

$$\phi(K_0, K_1, X) = (1-\lambda)^2(AXA' + Q) + \lambda^2(F_0XF_0' + V_0) + 2(1-\lambda)\lambda(F_1XF_1' + V_1),$$

where $F_0 = A + K_0C$, $F_1 = A + K_1C$, $V_0 = Q + K_0G_0K_0'$, *and* $V_1 = Q + K_1G_1K_1'$. *Assume* $X \in \{S \in \mathscr{R}^{n\times n} | S \geqslant 0\}$, $G_0 > 0$, $G_1 > 0$, $Q > 0$, *and* $(A, Q^{1/2})$ *is controllable. Then the following facts are true*:

(a) *With* $K_{x0} = -AXC'(CXC' + G_0)^{-1}$ *and* $K_{x1} = -AXC'(CXC' + G_1)^{-1}$, $g_\lambda(X) = \phi(K_{x0}, K_{x1}, X)$;
(b) $g_\lambda(X) = \min_{(K_0,K_1)} \phi(K_0, K_1, X) \leqslant \phi(K_0, K_1, X)$ *for any* (K_0, K_1);
(c) *If* $X \leqslant Y$, *then* $g_\lambda(X) \leqslant g_\lambda(Y)$;
(d) *If* $\lambda_1 \leqslant \lambda_2$, *then* $g_{\lambda_1}(X) \geqslant g_{\lambda_2}(X)$;
(e) *If* $\alpha \in [0, 1]$, *then* $g_\lambda(\alpha X + (1-\alpha)Y) \geqslant \alpha g_\lambda(X) + (1-\alpha)g_\lambda(Y)$;
(f) $g_\lambda(X) \geqslant (1-\lambda)^2AXA' + Q$;
(g) *If* $\bar{X} \geqslant g_\lambda(\bar{X})$, *then* $\bar{X} \geqslant 0$;
(h) *If* X *is a random variable, then* $(1-\lambda)^2AE[X]A' + Q \leqslant E[g_\lambda(X)] \leqslant g_\lambda(E[X])$.

Lemma 11. *Let* $X_{k+1} = h(X_k)$ *and* $Y_{k+1} = h(Y_k)$. *If* $h(X)$ *is a monotonically increasing function, then*:

$$X_1 \geqslant X_0 \Rightarrow X_{k+1} \geqslant X_k, \forall k \geqslant 0;$$
$$X_1 \leqslant X_0 \Rightarrow X_{k+1} \leqslant X_k, \forall k \geqslant 0;$$
$$X_0 \leqslant Y_0 \Rightarrow X_k \leqslant Y_k, \forall k \geqslant 0.$$

Lemma 12. *Define the linear operator* $\mathscr{L}(Y) = (1-\lambda)^2AYA' + \lambda^2F_0YF_0' + 2(1-\lambda)\lambda F_1YF_1'$ *and suppose there exists* $\bar{Y} > 0$ *such that* $\bar{Y} > \mathscr{L}(\bar{Y})$.

(a) *For all* $W \geqslant 0$, $\lim_{k\to\infty} \mathscr{L}^k(W) = 0$;
(b) *Let* $V \geqslant 0$ *and consider the linear system* $Y_{k+1} = \mathscr{L}(Y_k) + V$ *initial at* Y_0, *then the sequence* $\{Y_k\}$ *is bounded.*

Lemma 13. *Suppose there exists* $\bar{K}_0$, $\bar{K}_1$, *and* $\bar{P} > 0$ *such that*

$$\bar{P} > \phi(\bar{K}_0, \bar{K}_1, \bar{P}),$$

then for any initial value P_0, *the sequence* $P_k = g_\lambda^k(P_0)$ *is bounded, i.e., there exists* $M_{P_0} \geqslant 0$ *dependent of* P_0 *such that*

$$P_k \leqslant M_{P_0}, \forall k.$$

Proof of Theorem 2. Obviously there are two special cases:

- When $\lambda = 1$, the MARE reduces to the standard algebraic Riccati equation and it converges to a unique positive semi-definite solution.
- When $\lambda = 0$, all the packets are lost. Since A is unstable, the covariance matrix diverges for some initial values.

Next, we need to show that there exists a single point of transition between the two cases. Suppose for $0 < \lambda_1 \leqslant 1$, $E_{\lambda_1}[P_k]$ is bounded for any initial values. Then for any $\lambda_2 > \lambda_1$, we have

$$E_{\lambda_1}[P_k] = E[g_{\lambda_1}(P_k)] \geqslant E[g_{\lambda_2}(P_k)] = E_{\lambda_2}[P_k].$$

So $E_{\lambda_2}[P_k]$ is also bounded. Now we can choose

$$\lambda_c = \{\inf \lambda^* : \lambda > \lambda^* \Rightarrow E_\lambda[P_k] \text{is bounded for any initial value } P_0 \geqslant 0\}$$

and finish the proof. □

Proof of Theorem 3. For the lower bound of λ_c, we define the Lyapunov operator $\mathscr{M}(X) = \bar{A}X\bar{A}' + Q$ where $\bar{A} = (1-\lambda)A$. If $(A, Q^{1/2})$ is controllable, $(\bar{A}, Q^{1/2})$ is also controllable. Then the $\hat{S} = \mathscr{M}(\hat{S})$ has a unique strictly positive definite solution $\hat{S}$ if and only if $\max_i |\sigma_i(\bar{A})| < 1$, so we get $\underline{\lambda} = 1 - (1/\alpha)$. Consider the iteration $S_{t+1} = \mathscr{M}(S_t)$ for any $\lambda > \underline{\lambda}$, it converges. While for $\lambda \leqslant \underline{\lambda}$, it is unstable and S_k tends to infinity for any initial values.

For the mean value of the error covariance matrix $E[P_k]$ initialized at $E[P_0] \geqslant 0$, consider $0 = S_0 \leqslant E[P_0]$, it's easy to show that

$$\begin{aligned} S_k \leqslant E[P_k] \Rightarrow S_{k+1} &= \mathscr{M}(S_k) \\ &\leqslant (1-\lambda)^2AE[P_k]A' + Q \\ &\leqslant E[g_\lambda(P_k)] = E[P_{k+1}]. \end{aligned}$$

By induction, it is obvious that when $\lambda < \underline{\lambda}$, $\lim_{k\to\infty} E[P_k] \geqslant \lim_{k\to\infty} S_k = \infty$. This implies that for any initial condition $E[P_k]$ is unbounded for $\lambda < \underline{\lambda}$, therefore $\underline{\lambda} \leqslant \lambda_c$.

For the upper bound of λ_c, consider the sequence $V_{k+1} = g_\lambda(V_k)$ and $V_0 = E[P_0] \geqslant 0$, we have

$$\begin{aligned} E[P_k] \leqslant V_k \Rightarrow E[P_{k+1}] &= E[g_\lambda(P_k)] \\ &\leqslant g_\lambda(E[P_k]) \\ &\leqslant g_\lambda(V_k) = V_{k+1}. \end{aligned}$$

A simple induction shows that for any k, $V_k \geqslant E[P_k]$. So for $\lambda > \bar{\lambda}$, according to Lemma 10 part (g), there exists $\bar{X} > 0$. Therefore, all conditions of Lemma 13 are satisfied and we have

$$E[P_k] \leqslant V_k \leqslant M_{V_0}$$

for any k. This shows that $\lambda_c \leqslant \bar{\lambda}$. □

Proof of Theorem 4. Using Lemma 10, it is easy to show that if there exists a $\bar{X} > 0$ such that $\bar{X} > g_\lambda(\bar{X})$, $\bar{X} > g_\lambda(\bar{X}) = \phi(K_{\bar{X}0}, K_{\bar{X}1}, \bar{X})$. Also it is obvious that $\bar{X} > \phi(K_0, K_1, \bar{X}) \geqslant g_\lambda(\bar{X})$, so (a) is equivalent to (b). The only trick we need for the remaining proof is to use Schur complement decomposition to obtain the function Ψ_λ. Please note that

$$\begin{aligned} \phi(K_0, K_1, X) &= (1-\lambda)^2(AXA' + Q) + \lambda^2(F_0XF_0' + V_0) \\ &\quad + 2(1-\lambda)\lambda(F_1XF_1' + V_1) \\ &= (1-\lambda)^2AXA' + Q + \lambda^2F_0XF_0' \\ &\quad + 2(1-\lambda)\lambda F_1XF_1' + \lambda^2K_0G_0K_0' \\ &\quad + 2(1-\lambda)\lambda K_1G_1K_1'. \end{aligned}$$

The part (b) is equivalent to

$$\begin{bmatrix} X-(1-\lambda)^2AXA'+\lambda^2F_0XF_0' & \sqrt{2(1-\lambda)\lambda}F_1 \\ \sqrt{2(1-\lambda)\lambda}F_1' & X^{-1} \end{bmatrix} > 0.$$

Using Schur complement decomposition two more times to obtain

$$\begin{bmatrix} X & \sqrt{2(1-\lambda)\lambda}F_1 & \lambda F_0 & (1-\lambda)A \\ \sqrt{2(1-\lambda)\lambda}F_1' & X^{-1} & 0 & 0 \\ \lambda F_0' & 0 & X^{-1} & 0 \\ (1-\lambda)A' & 0 & 0 & X^{-1} \end{bmatrix} > 0.$$

Let $Y=X^{-1}$, $Z_1=X^{-1}K_1$, and $Z_1=X^{-1}K_1$, we get

$$\begin{bmatrix} Y & \sqrt{2(1-\lambda)\lambda}I_1 & \lambda I_0 & (1-\lambda)YA \\ \sqrt{2(1-\lambda)\lambda}I_1' & Y & 0 & 0 \\ \lambda I_0' & 0 & Y & 0 \\ (1-\lambda)A'Y & 0 & 0 & Y, \end{bmatrix} > 0,$$

where $I_1=YA+Z_1C$ and $I_0=YA+Z_0C$, and this is what we define as $\Psi_\lambda(Y,Z_0,Z_1)$. Since $\Psi_\lambda(\alpha Y,Z_0,Z_1)=\alpha\Psi_\lambda(Y,Z_0,Z_1)$, so Y can be restricted to $0<Y\leqslant I$. □

References

Brockett, R. W., & Liberzon, D. (2000). Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, *45*(7), 1279–1289.

Elia, N., & Mitter, S. K. (2001). Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, *46*(9), 1384–1400.

Elliott, E. O. (1963). Estimates of error rates for codes on burst-noise channels. *Bell System Technical Journal*, *42*, 1977–1997.

Fleming, M., Zhao, Q., & Effros, M. (2004). Network vector quantization. *IEEE Transactions on Information Theory*, *50*(8), 1584–1604.

Gamal, A. A. E., & Cover, T. M. (1982). Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, *28*(6), 851–857.

Gilbert, E. N. (1960). Capacity of a burst-noise channel. *Bell System Technical Journal*, *39*, 1253–1265.

Goyal, V. K., & Kovacevic, J. (2001). Generalized multiple description coding with correlating transforms. *IEEE Transactions on Information Theory*, *47*(6), 2199–2224.

Goyal, V.K., Kovaceric, J., Arean, R., & Vetterli, M., (1998). Multiple description transform coding of images. Proceedings of IEEE international conference on image processing. (Vol. 2, pp. 674–678), Chicago, IL.

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.

Jin, Z., Waydo, S., Wildanger, E. B., Lammers, M., Scholze, H., Foley, P., et al. (2004). Mvwt-ii: The second generation caltech multi-vehicle wireless testbed, Proceedings of the 2004 American control conference, (Vol. 1, pp. 5321–5326), Boston, MA.

Lee, W. S., Pickering, M. R., Frater, M. R., & Arnold, J. F. (2000). A robust codec for transmission of very low bit-rate video over channels with bursty errors. *IEEE Transactions on Circuits and Systems for Video Technololgy*, *10*, 1403–1412.

Liberzon, D. (2003). On stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, *48*(2), 304–307.

Liu, X., & Goldsmith, A. J. (2004). *Kalman filtering with partial observation losses. IEEE conference on decision and control*. Bahamas: Paradise Island.

Marco, D., & Neuhoff, D. L. (2005). The validity of the additive noise model for uniform scalar quantizers. *IEEE Transactions on Information Theory*, *51*(5), 1739–1755.

Matveev, A. S., & Savkin, A. V. (2003). The problem of state estimation via asynchronous communication channels with irregular transmissioni times. *IEEE Transactions on Automatic Control*, *48*(4), 670–675.

Matveev, A. S., & Savkin, A. V. (2005). Comments on "control over noisy channels" and relevant negative results. *IEEE Transactions on Automatic Control*, *50*(12), 2105–2110.

Murray, R.M., (Ed.) (2002). Control in an information rich world (panel report). ⟨http://www.cds.caltech.edu/murray/cdspanel/⟩.

Ozarow, L. (1980). On a source-coding problem with two channels and three receivers. *Bell System Technical Journal*, *59*(10), 1909–1921.

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.

Tatikonda, S., & Mitter, S. K. (2004a). Control over noisy channels. *IEEE Transactions on Automatic Control*, *49*(7), 1196–1201.

Tatikonda, S., & Mitter, S. K. (2004b). Control under communication constraints. *IEEE Transactions on Automatic Control*, *49*(7), 1056–1068.

Vaishampayan, V. A. (1993). Design of multiple description scalar quantizers. *IEEE Transactions on Information Theory*, *39*(3), 821–834.

Vaishampayan, V. A., & Batllo, J. C. (1998). Asymptotic analysis of multiple description quantizers. *IEEE Transactions on Information Theory*, *44*(1), 278–284.

Walsh, G. C., Ye, H., & Bushnell, L. G. (2002). Stability analysis of networked control systems. *IEEE Transactions on Control Systems Technology*, *10*(3), 438–446.

Wong, W. S., & Brockett, R. W. (1997). Systems with finite communication bandwidth constraints—part I: State estimation problems. *IEEE Transactions on Automatic Control*, *42*(9), 1294–1299.

Yajnik, M., Moon, S., Kurose, J., & Towsley, D., (1999). Measurement and modelling of the temporal dependence in packet loss, INFOCOM '99. Eighteenth annual joint conference of the IEEE computer and communications societies. Proceedings (Vol. 1, pp. 345–352), New York, NY, USA: IEEE.

Zhipu Jin received the B.S. degree and M.S. degree in Electrical Engineering from Tsinghua University, Beijing, China, in 1998 and 2001. He also received the M.S. degree in Electrical Engineering from California Institute of Technology in 2002. He is currently pursuing his Ph.D. degree at the California Institute of Technology, Pasadena. His research interests include packet-based networked control systems, cooperation/coordination between multiple vehicles with communication and computation constraints, and estimation over sensor networks.

Vijay Gupta received his B.Tech degree from Indian Institute of Technology, Delhi, India in 2001 and his M.S. degree from California Institute of Technology in 2002, both in Electrical Engineering. He is currently a Ph.D. candidate at the California Institute of Technology, Pasadena. His research interests include networked control systems, distributed estimation and detection, usage-based value of information and the interaction of communication and control.

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. Current projects include integration of control, communications, and computer science in multi-agent systems, information dynamics in networked feedback systems, analysis of insect flight control systems, and synthetic biology using genetically-encoded finite state machines.