Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

MoB06.6

State Estimation Utilizing Multiple Description Coding over Lossy Networks

Zhipu Jin, Vijay Gupta, Babak Hassibi, Richard M. Murray

Abstract— For state estimation in networked control systems, the impact of packet dropping over network links is an important problem. In this paper, we introduce multiple description (MD) source coding scheme to improve the statistical stability and performance of the estimation error covariance of Kalman filter with packet loss. We consider about two cases: when the packet loss over network links occurs in an i.i.d. fashion or in a bursty fashion. Compared with the traditional single description source coding, MD coding scheme can greatly improve the performance of Kalman filter over a large set of packet loss scenarios in both cases.

I. INTRODUCTION

A standard assumption in the classical control theory is that the data can be transmitted to controller or state estimator reliably and with infinite precision, or at most corrupted by an additive Gaussian white noise. Thus reliable communication channels with infinite bandwidth are needed. However, in the real world, any communication link has a limited channel capacity and data packets may be dropped. Increasing attentions are given to consider the effects of finite bit rate and stochastic packet losses, especially in the field of networked control systems (NCS) where the standard assumption is challenged most severely. Works like [1], [14], [15], [19], [20] have focused on answering the question: how much channel capacity do we need to achieve a certain control performance or estimation accuracy? In this paper, we are interested in another issue: how does the unreliability of the communication network affects NCS and what can we do to compensate for this unreliability? More specifically, how can we improve the state estimation with the presence of stochastic packet loss.

Most of the modern digital communication systems are implemented by packet-based communication protocols such as the transmission control protocol (TCP). For real-time networked control systems, this scheme faces couple of serious problems:

- TCP automatically retransmits lost packets which generates large delays.
- There may not exist reliable reverse channels from decoders to encoders to introduce feedbacks and cannot guarantee efficient retransmission.

Zhipu Jin is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. jzp@caltech.edu

Vijay Gupta is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA

Babak Hassibi and Richard M. Murray are with Faculty of the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA
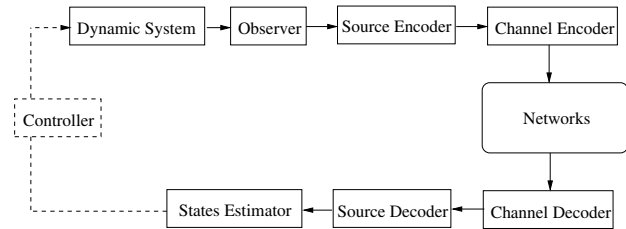
Fig. 1. Diagram of a networked control system

We make the following general assumptions for the networked control systems:

- Each data packet is protected from channel noise by perfect channel coding. The packet is either received and decoded successfully at the end of the links or totally lost.
- There are not computation delays, such as coding delay, shaping delay, packetisation delay, or receiver play out delay.
- We model the packet losses either according to an i.i.d. random process (the Bernoulli model in [21]) or according to a Markov chain (the Gilbert-Elliott channel model in [2], [5]) which can handel bursty channel losses.
- The network does not provide preferential treatment to packets. In other words, the network treats each single packet equally without inspecting the content.
- There is no feedback from decoders to encoders over the networks.
- The estimator only use new data packet to generate "real-time" estimation due to limited memory and computation ability.

In this paper, we focus on state estimation problem of a dynamical system over a packet dropping link. We choose the error covariance matrix of the estimation as our metric of the performance of the estimator. In their outstanding previous work, Sinopoli *et al.* [13] used a Modified Algebraic Riccati Equation (MARE) to solve the Kalman filtering problem with intermittent observations and discussed the statistical convergence properties of the estimation error covariance. Liu *et al.* [11] extended the result to the case with partial observation losses in sensor network. These works showed that the packet loss degrades the performance of Kalman filter. In this paper, we improve the performance of the estimator by using network source coding [3]. The specific scheme we consider is the multiple description (MD) source coding. MD coding has been studied in information theory

for over 30 years [4], [7] and has been successfully used in transmission real-time speech, audio/video over internet [6], [9], [10]. It has been showed that MD codes are very useful when the data can be used at various quality levels. However, this is the first time such coding schemes is applied to networked control system.

The remainder of this paper is organized as follows. In Section II, we introduce the MD source coding and briefly describe the theoretical limits. Also, the quantization noise of MD coding is modelled by gaussian white noise asymptotically. We formulate the state estimation problem in Section III and present results for the i.i.d. Bernoulli loss model. Examples and simulation results are given in Section IV. We then study MD coding over the Markov chain model in Section V and summarize the conclusions in Section VI.

II. MULTIPLE DESCRIPTION SOURCE CODING

Multiple description source coding [3] is used to generate a network source code that can achieve good rate-distortion performance over lossy links. The unique feature is that instead of using one single description to represent one possible output of the source, a MD code uses two or more descriptions. The distortion of the decoder output is in various quality levels which depend on how many descriptions are received. The order of descriptions is not important since MD coding is not hierarchical. The design of a MD code is a problem of optimizing the code over the efficiency and independence between descriptions.
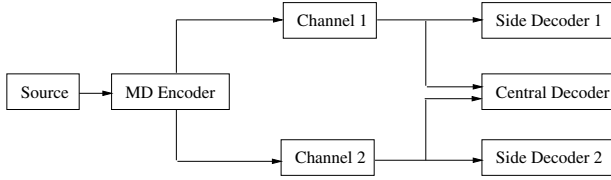

Fig. 2. Scenario for MD source coding with two channels and three receivers

Originally, MD coding refers to the case depicted in Fig. 2 and we call it 2-description MD coding problem. A sequence of source values $\{X_k\}_{k=1}^N$ are sent to three receivers over two noiseless channels. The encoder generates 2 descriptions for each source value and sends them through two different channels. One decoder receives the descriptions from both channels and we call it the central decoder. The reconstruction sequence at the central decoder is $\{\hat{X}_k^0\}_{k=1}^N$. The other two decoders receive descriptions only over their respective channels and we call them the side decoders. The reconstruction sequences are $\{\hat{X}_k^i\}_{k=1}^N$, $i = 1, 2$. The transmission rate over channel i is denoted by R_i, $i \in \{1, 2\}$. Three distortions are defined as:

$$D_i = \frac{1}{N} \sum_{k=1}^N E\left[\delta_i\left(X_k, \hat{X}_K^i\right)\right],$$

for $i = 0, 1, 2$, where the $\delta_i(\cdot, \cdot)$s are distortion measures. For now on, we let

$$\delta_i\left(X_k, \hat{X}_K^i\right) = \delta\left(X_k, \hat{X}_K^i\right) = \|X_k - \hat{X}_K^i\|_2.$$

If $R_1 = R_2$ and $D_1 \approx D_2$, we call the MD code is balanced. The MD coding problem can be generalized to $L(> 2)$ channels and we call it L-description MD coding [18]. The main theoretical problem in MD coding is to determine the achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$. As discussed in [4], the fundamental tradeoff in MD is making descriptions individually good and sufficiently different at the same time.

For the packet-based network, we use balanced MD codes and assume $R_1 = R_2 = R \gg 1$ and $D_1 = D_2 \ll 1$. Then we have the inequality [16]

$$D_0 \cdot D_1 \geq \frac{1}{4} 2^{-4R}.$$

This means the product of central and side distortions is approximately lower-bounded by $4^{-1}2^{-4R}$. If $D_1 \approx D_2 \approx 2^{-2(1-\alpha)R}$ where $\alpha \in [0, 1]$ is a pre-defined parameter for MD coding design, then the best distortion of the central distortion is

$$D_0 \approx \frac{1}{4} 2^{-2(1+\alpha)R}.$$

This shows a tradeoff between central and side distortions. The penalty in the exponential rate of decay of D_1 is exactly the increase in the rate of decay of D_0.

Since in networked control systems, the source data are real values of the observations, so the MD scalar quantizer (MDSQ) is a natural choice. The balanced 2-description MDSQ can be developed by the algorithm proposed in [17], and can be easily extended to 3-description or 4-description case. The first part of Table I shows some examples of distortions for different description loss cases when we keep the central distortions constant. It is clear that, in order to get the same accuracy, we need more bits per source sample (bpss). The second part shows that the distortions become bigger when the number of descriptions per sample increases and we keep bpss constant at the same time. In the table, "lost k" means k of the descriptions has been lost, and "N/A" means not available.

For other MD codes whose central decoder have same distortions, the distortions of decoders are also listed in second part of table I. It shows that we have to use more bits for MD codes if we want to get same central distortion. The MD coding actually provides various quality decoding levels corresponding to how many descriptions the decoder receives.

The similarity between the original 2-channel MD case (shown in Fig.2) and the 2-description MD with packet-based networks in NCS is obvious. We put 2 descriptions of each source sample into 2 different packets and sent them out in sequence. At the end of the link, the distortion of the MD decoder only depends on how many descriptions successfully pass though the network. Overtime delay equals to packet dropping since old data is not used for real-time state estimation.

The MD coding increases the computation complexity since the size of look-up tables increases a lot at the decoder side. For example, for the traditional uniform quantizer with N levels, the look-up table for a L-description MD code has

TABLE I

MSE FOR DIFFERENT MD CODING

Coding type	No loss	Lost 1	Lost 2	Lost 3	Total bpss
single description	8.33×10^{-6}	N/A	N/A	N/A	10
2-description	8.33×10^{-6}	1.56	N/A	N/A	12
3-description	8.33×10^{-6}	4.41×10^{-3}	1.53	N/A	15
4-description	8.33×10^{-6}	7.46×10^{-3}	1.34×10^{-2}	2.61	20

Coding type	No loss	Lost 1	Lost 2	Lost 3	Total bpss
single description	4.97×10^{-7}	N/A	N/A	N/A	12
2-description	8.33×10^{-6}	1.56	N/A	N/A	12
3-description	9.87×10^{-5}	1.97×10^{-2}	2.15	N/A	12
4-description	9.32×10^{-4}	8.04×10^{-2}	0.113	2.18	12

$(2^L - 1) \cdot N$ elements. Obviously, we need to consider this issue when choosing L. In most cases, a 2-description MD code is good enough.

As discussed in [12], the quantization error of a uniform scalar quantizer with the assumptions of small cells, reproduction levels at the cell's midpoints, and large support region can be approximately modelled as additive uncorrelated white noise. According to Section II, the central decoder of MD coding actually is a uniform scalar quantizer with the midpoints as the outputs and the central distortion is $D_0 \approx \frac{\Delta^2}{12}$ where Δ denotes the width of a quantization cell.

As discussed in [16], the side decoders introduce a slight asymmetry between the two side distortions and cause a small increase in distortion. However, as the bit rate increases, this asymmetry asymptotically disappears. According to the relationship discussed in Section II, we have

$$D_1 \approx D_2 \approx C_1 \cdot \left(\frac{C_2}{12}\right)^{\frac{1+\alpha}{1-\alpha}} \cdot \left(\Delta^{\frac{1-\alpha}{1+\alpha}}\right)^2.$$

For a balanced 2-description MD code, when α is a constant, D_1 will be asymptotically negligible relative to $\left(\Delta^{\frac{1+\alpha}{1-\alpha}}\right)^2$. So as long as the rate $R_1 (= R_2)$ is big enough, the addition noise model is also good enough to represent the quantization noise for the side decoders. In this paper, we model the MD quantization noise as Gaussian white noise with zero mean and covariance is D_0 for central decoder and D_1 for side decoders.

III. STATISTICAL CONVERGENCE OF KALMAN FILTER USING MD CODES

A. Problem Formulation

Consider the following discrete time linear dynamic system:

$$\begin{aligned} x_{t+1} &= Ax_t + w_t \\ y_t &= Cx_t + v_t \end{aligned}$$

where $x_t \in \mathcal{R}^n$ is the state vector, $y_t \in \mathcal{R}^m$ is the output vector, w_t and v_t are Gaussian white noise vectors with zero mean and covariance matrices are $Q \geq 0$ and $R > 0$ respectively. It is well known that if $(A, Q^{\frac{1}{2}})$ is controllable, (A, C) is detectable, and no measurement is lost, the estimation error covariance of Kalman filter converges to a unique value from any initial condition. In this Section, we assume that packets are lost stochastically and packet dropping occurs

according an independent, identically distributed Bernoulli random process.

We use a 2-description balanced MD coding scheme. The measurement output y_t goes through a MD encoder and is coded into two descriptions (i_t, j_t). These two descriptions are put into two packets and transmitted separately. We use the variables $\gamma_{i,t}$ and $\gamma_{j,t}$ to indicate whether the description i_t and j_t are received correctly. If i_t is received correctly, then $\gamma_{i,t} = 1$, otherwise, $\gamma_{i,t} = 0$, and similarly for $\gamma_{j,t}$. We assume that $\gamma_{i,t}$ and $\gamma_{j,t}$ are i.i.d. Bernoulli random variables with the probability distribution $P(\gamma_{i,t} = 1) = P(\gamma_{j,t} = 1) = \lambda$.

Since i_t and j_t are independently lost or received, we can have three measurement rebuilding scenarios. First, we may receive both of the descriptions correctly and the measurement noise is the white noise v_t plus the central distortion. We use $R_0 = R + D_0$ to indicate its covariance. Second, we may receive only one description correctly and the measurement noise will be $R_1 = R + D_1$ where D_1 is the side distortion. Third, we may receive none of the descriptions correctly, then the measurement is corrupted with a infinitely large noise. So the measurement noise is changed into a random variable $\hat{v}_t$ after the decoder at the end of the link and for the covariance Cov_t we have:

$$Cov_t = \begin{cases} R_0 & : \text{ probability is } \lambda^2 \\ R_1 & : \text{ probability is } 2(1-\lambda)\lambda \\ \sigma^2 I & : \text{ probability is } (1-\lambda)^2 \end{cases}$$

where $\sigma \to \infty$.

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

$$\begin{aligned} P_{t+1} &= AP_tA' + Q \\ &- \gamma(i,t)\gamma(j,t)AP_tC'[CP_tC' + R_0]^{-1}CP_tA' \\ &- (1-\gamma(i,t))\gamma(j,t)AP_tC'[CP_tC' + R_1]^{-1}CP_tA' \\ &- \gamma(i,t)(1-\gamma(j,t))AP_tC'[CP_tC' + R_1]^{-1}CP_tA'. \end{aligned}$$

This is a stochastic recursion and the sequence of the error covariance matrix $P_{t=0}^{\infty}$ is a random process for a given initial value. Using the same approach discussed in [13], we define the Modified Algebraic Riccati Equation (MARE) for the

Kalman filter with MD coding shceme as follows

$$
\begin{aligned}
g_\lambda(X) \;=\;\; & AXA' + Q \\
& -\lambda^2 AXC'(CXC' + R_0)^{-1}CXA' \\
& -2(1-\lambda)\lambda AXC'(CXC' + R_1)^{-1}CXA'
\end{aligned}
$$

where λ is the probability that a single packet can be received correctly.

B. Convergence Conditions and Boundaries

This subsection lists theorems which are used to study the convergence properties of the MARE. For brevity, we omit the proofs which can be obtained by following the same approach in [13], [11]. Considering the new MARE, we have the following theorem which states the uniqueness of the solution.

Theorem 3.1: Let the operator

$$
\begin{aligned}
\phi(K_0, K_1, X) \;=\;\; & (1-\lambda)^2(AXA' + Q) \\
& +\lambda^2(F_0 X F_0' + V_0) \\
& +2(1-\lambda)\lambda(F_1 X F_1' + V_1)
\end{aligned}
$$

where $F_0 = A + K_0 C$, $F_1 = A + K_1 C$, $V_0 = Q + K_0 R_0 K_0'$, and $V_1 = Q + K_1 R_1 K_1'$. Suppose there exists K_0, K_1, and $P > 0$ such that $P > \phi(K_0, K_1, P)$, then we have

(a) for any initial condition $P_0 \geq 0$, the MARE converges, i.e. the iteration $P_{t+1} = g_\lambda(P_t)$ converges, and the limit is independent of the initial value:

$$
\lim_{t\to\infty} P_t = \lim_{t\to\infty} g_\lambda^t(P_0) = \bar{P};
$$

(b) $\bar{P}$ is the unique positive semi-definite solution of MARE function $\bar{P} = g_\lambda(\bar{P})$.

The following theorem relates the packet receiving probability and the convergence of the MARE.

Theorem 3.2: If $(A, Q^{\frac{1}{2}})$ is controllable, (A, C) is detectable, and A is unstable, then there exists a $\lambda_c \in [0, 1)$ such that

(a) $\lim_{t\to\infty} E[P_t] = +\infty$ for $0 \leq \lambda \leq \lambda_c$ and some initial condition $P_0 \geq 0$,

(b) $E[P_t] \leq M_{P_0}$ $\forall t$ for $\lambda_c < \lambda \leq 1$ and any initial condition $P_0 \geq 0$,

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

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

Theorem 3.3: Let

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

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

$$
\underline{\lambda} \leq \lambda_c \leq \bar{\lambda}.
$$

This theorem states the upper and lower bound of the critical value of the packet receiving probability. The lower bound is in a closed form and the next theorem states how to get it. For some special cases, these two bounds are identical and we will discuss them later. According to [13], for the traditional single description coding scheme, the lower bound is $1 - \frac{1}{\alpha^2}$. So using MD coding pushes the lower bound to a smaller value and guarantee the convergence over a larger area.

Theorem 3.4: Assume $(A, Q^{\frac{1}{2}})$ is controllable and (A, C) is detectable, then the following statements are equivalent:

(a) $\exists \bar{X}$ such that $\bar{X} > g_\lambda(\bar{X})$;

(b) $\exists(\bar{K}_0, \bar{K}_1, \bar{X}) > 0$ such that $\bar{X} > \phi(\bar{K}_0, \bar{K}_1, \bar{X})$;

(c) $\exists \bar{Z}_0, \bar{Z}_1$ and $0 < \bar{Y} \leq I$ such that $\Psi_\lambda(\bar{Y}, \bar{Z}_0, \bar{Z}_1) > 0$ where

$$
\Psi_\lambda(Y, Z_0, Z_1)
$$
$$
= \begin{bmatrix}
Y & \Delta(Y, Z_1) & \Omega(Y, Z_0) & \Pi(Y) \\
\Delta(Y, Z_1)' & Y & 0 & 0 \\
\Omega(Y, Z_0)' & 0 & Y & 0 \\
\Pi(Y)' & 0 & 0 & Y
\end{bmatrix},
$$

$\Delta(Y, Z_1) = \sqrt{2(1-\lambda)\lambda}(YA + Z_1 C)$, $\Omega(Y, Z_0) = \lambda(YA + Z_0 C)$, and $\Pi(Y) = (1-\lambda)YA$.

According to this theorem, we can get the following corollary to reformulated the computation of $\bar{\lambda}$ as an LMI feasible problem.

Corollary 3.5: The upper bound $\bar{\lambda}$ is given by the solution of the following optimization problem,

$$
\bar{\lambda} = \arg\min_\lambda \left(\Psi_\lambda(Y, Z_0, Z_1) > 0 \right)
$$

where $0 < Y \leq I$.

Theorem 3.6: Assume $(A, Q^{\frac{1}{2}})$ is controllable, (A, C) is detectable, and $\bar{\lambda} < \lambda$, then for any initial condition $E[P_0] \geq 0$,

$$
0 \leq \bar{S} \leq \lim_{t\to\infty} E[P_t] \leq \bar{V}
$$

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

This theorem shows the upper and lower bound of the error covariance matrix when MARE converges. The lower bound $\bar{S}$ can be computed by standard Lyapunov Equation Solvers and the upper bound $\bar{V}$ 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.

There are some special cases in which the upper and lower bound of the critical value λ_c are identical.

(a) C is invertible. In this case, we choose $K_0 = K_1 = -AC^{-1}$ to make $F_0 = F_1 = 0$. Then the LMI in theorem 3.4 is equivalent to

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

Since $X \geq 0$ exists if and only if $(1-\lambda)A$ is stable, i.e. the magnitudes of eigenvalues of $(1-\lambda)A$ are all smaller than 1, we get $\bar{\lambda} = \underline{\lambda} = \frac{1}{1-\alpha}$.

(b) The matrix A has a single unstable eigenvalue. As long as (A, C) is detectable, we can always use decomposition to transform the system such that the Kalman filter only needs to estimate a single system state. Then it follows that the lower bound and upper bound are identical.

IV. EXAMPLES AND SIMULATION RESULTS

In this section examples and simulation results are given to show how MD coding affects the performance of the Kalman filtering. As discussed before, when C is invertible, the upper and lower bound on the critical value λ_c coincide. We choose a discrete time LTI system with $A = -1.25$ and $C = 1$. The noise w_t and v_t have zero means and variance $R = 2.5$ and $Q = 1$ respectively. A 2-description MD code is designed according to [17] such that the central MSE $D_0 \approx 8.33 \times 10^{-6}$ and $D_1 \approx 1.56$. According to the results in Section II, we can get $R_0 \approx 2.5$ and $R_1 \approx 4.06$.
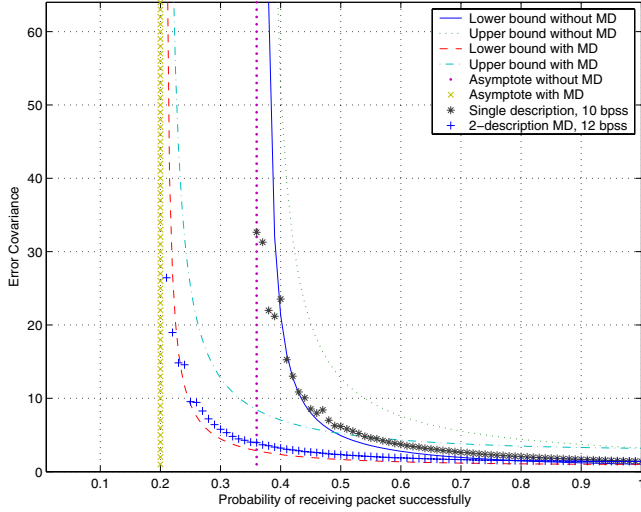

Fig. 3. Simulation results of error covariances with theoretical upper and lower bounds

Fig. 3 shows the expected estimation error covariance with different coding schemes. Using MD coding, the theoretical asymptote λ_c is pushed from 0.36 to 0.2. The convergence properties of error covariance at high packet loss rate region is decreased dramatically.

Some simulations have been done in MATLAB by implementing the MD encoder and decoder. We repeated each scenario 2000 times and used the average values as the approximations. In Fig. 3, the simulation results are consistent to the theoretical limits very well. Some of the simulation points are below the lower bound near the critical λ_c value because we only run simulation over limited time steps and the covariances take longer time to converge.

Fig. 4 shows some additional simulation results. For each certain packet dropping rate, the center of the error bar is the mean value and 95% of the simulation results are inside the error bar. It's clear that if we use a 3-description code, the critical value λ_c will be pushed even further. So the benefits of using MD are clear and the cost we need to pay is more bpss. Fig 5 shows the details about the error covariance when packet dropping rates are small. The MD scheme gives much better performance and robustness than the single-description scheme. Please Note that the 2-description MD achieves almost as good performance as sending the single description code twice and saves up to 40% bandwidth at the same time.
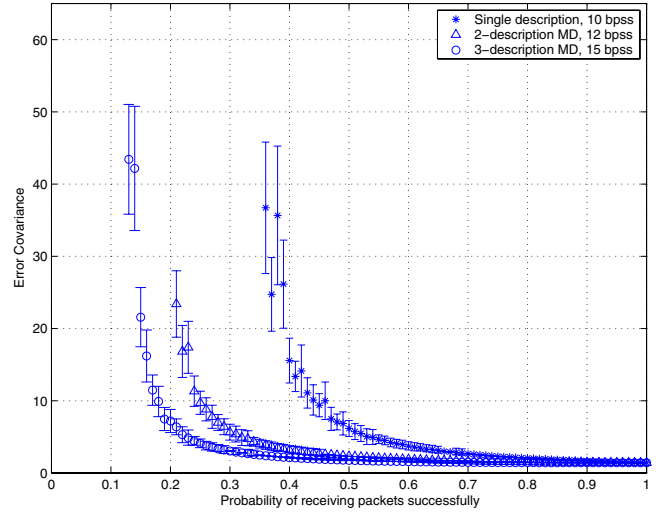

Fig. 4. Mean values of error covariance with same central distortion
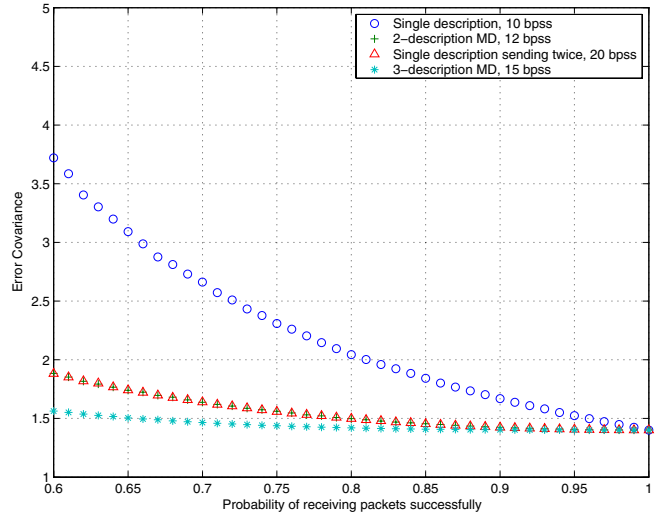

Fig. 5. Mean values of error covariance with low dropping rate

V. USING MD CODING OVER THE GILBERT-ELLIOT MODEL

So far, we have dealt with the situation when the packet loss occurs according to an Bernoulli loss model. Moreover, another popular model for packet drops in many channels (such as the wireless channel) is the one in which the losses occur in bursts. This bursty error behavior can be captured by a discrete-time Markov chain model. The simplest one of such models is the Gilbert-Elliot channel model. This model considers the channel transferring between two possible states - 'Good' and 'Bad'. In the good state, the packet is received successfully while the bad state corresponds to packets being dropped. The channel transfers between these two states according to a Markov chain with transition probability matrix $\mathcal{Q}$. Clearly, the model can easily be made more complicated by considering more than 2 states with different probabilities of packet drop. However, for reasons of simplicity and without loss of generality, we only consider

the 2-state model.

The analysis of the Markov channel case proceeds along similar lines as before. Suppose the channel can be modeled as a 2-state Markov chain with transition probability matrix $\mathcal{Q}$ given by

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

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. For the case of using a 2-MD code, we obtain a 4-state Markov chain where the states correspond to both packets lost, only the 1st description packet lost, only the 2nd description packet lost and no packet lost. The transition probability matrix of this chain is given by

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

Note that the state in which both packets are lost is equivalent to no observation coming through, while all the rest of the states correspond to the system being observed. We need results analogous to the Bernoulli case when packets are being dropped according to a Markov chain. We use the following results proven in [8].

Proposition 5.1: Consider the system

$$x_{t+1} = Ax_t + w_t,$$

being observed through n sensors with the i-th sensor of the form

$$y_t^i = C^i x_t + v_t^i.$$

Suppose only one sensor can be active at any time instant and the choice of the sensor is done according to a Markov chain with transition probability matrix $\mathcal{Q} = [q_{ij}]$. Denote the Ricatti update in error covariance when the i-th sensor is used by $f_i\left(.\right)$ and denote

$$f_i^t\left(.\right) = \underbrace{f_i\left(f_i\left(\cdots\left(.\right)\cdots\right)\right)}_{t \text{ times}}.$$

Then the expected error covariance at time step t, denoted by $E\left[P_t\right]$ is bounded as follows.

- Upper bound: Denote $q_i = \max_j q_{ji}$ and π_i is the initial probability of states i. Then an upper bound for $E\left[P_t\right]$ is X_t where

$$X_{t+1} = \begin{cases} \sum_i q_i f_i\left(X_t\right) & t \geq 1 \\ \sum_i \pi_i f_i\left(P_0\right) & t = 0. \end{cases}$$

Thus a sufficient condition for convergence of the error covariance is that X_t converges as t progresses.
- Lower bound: Denote the probability of being in Markov state j at time step t by π_t^j. Then a lower bound for $E\left[P_t\right]$ is Y_t where

$$\begin{aligned} Y_t &= q_{jj}^{t-1}\pi_0^j f_0^t\left(P_0\right) \\ &+ \sum_{i=1}^t q_{jj}^{i-1}\left(\pi_{t+1-i}^j - q_{jj}\pi_{t-i}^j\right)f_j^i\left(R\right), \end{aligned}$$

where R is the covariance matrix of the process noise w_t. Note that one such lower bound exists for each j. Thus a necessary condition for divergence of the error covariance is that

$$q_{jj}(\max|\sigma_i|)^2 > 1,$$

where $\max|\sigma_i|$ is the maximum magnitude among the eigenvalues of A when $\left(A, C^j\right)$ is put in observer canonical form.

We can easily apply these results to the same example with the transition probability matrices described above. When the single description code is applied, the system transitions according to a Markov chain between a state in which the system is observed and one in which it is not. With a 2-MD code, the four states corresponding to the transition probability matrix shown before. We can easily see the improvement in the performance by using MD codes. In figure 6 we plot the upper and lower bounds for the error variance as a function of q_{10} for the parameter $q_{11} = 0.95$. Although only the bounds are plotted, the lowering of the lower bound is indicative of the performance getting better with MD codes. This fact can be verified by actually simulating the system. The results for parameters $q_{11} = 0.05$ and $q_{11} = 0.95$ are shown in figure 7.
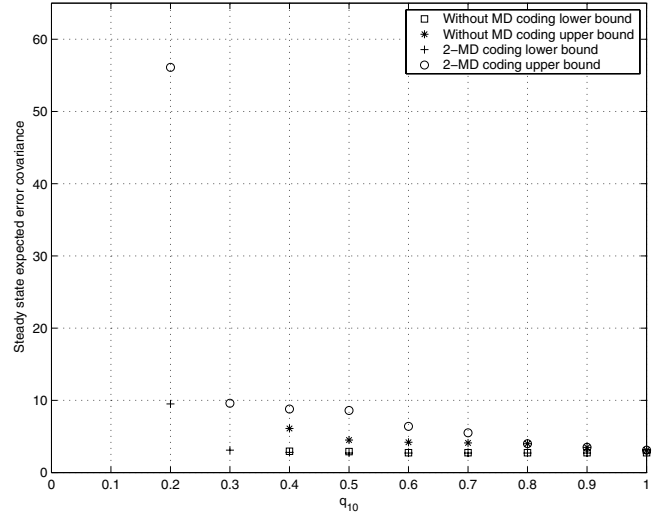

Fig. 6.   Upper and lower bounds for Markov chain case

It can be seen from the figures and the expressions given above that while the system diverges at $q_{10} = 0.36$ for the single description code case, and for the 2-MD code case, it diverges at $q_{10} = 0.2$. Thus the system stability margin is increased. It can actually be proven in this case (when the observation matrix C is invertible) that the necessary condition for divergence is sufficient as well.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we present a new scheme for the state estimation in networked control systems to compensate for packet dropping: using multiple description source coding to transfer state observations. In this scheme, we use 2
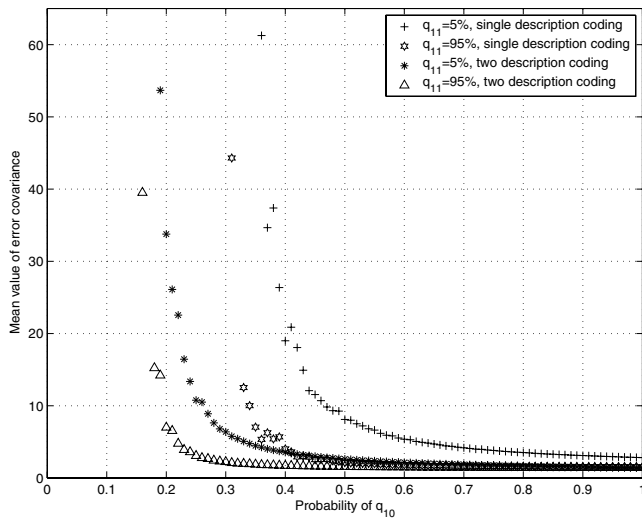
Fig. 7.   Simulation results for Markov chain case

descriptions to represent each observation sample instead of one description. The accuracy of the output of decoder only depends on how many descriptions has been successfully received. We consider about two models for packet transformation: the Bernoulli loss model and the Gilbert-Elliott model. In the high rate case, the estimation error covariance converges over a much larger receiving probability area using MD codes than using traditional single description source codes. Also, the scheme is advantaged over sending duplicate packets because it saves considerable communication bandwidth.

There are several issues we can work on in the future. First of all, we need a more complete theory to understand the MD coding for $L > 2$ case. Second, the validity of the quantization noise model of MD coding may need to be verified more carefully. Third, since MD coding will greatly increase the computation complexity of the decoder, a more efficient search algorithm for the source coding will be greatly helpful. Last, we would like to study the stability of the close-loop networked control systems with MD coding.

VII. Acknowledgements

References

[1] R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7):1279–1289, July 2000.

[2] E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech. J.*, 42:1977–1997, Sept. 1963.

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

[4] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, 28(6):851–857, 1982.

[5] E. N. Gilbert. Capacity of a burst-noise channel. *Bell Syst. Tech. J.*, 39:1253–1265, Sept. 1960.

[6] V. K. Goyal, J. Kovaceric, R. Arean, and M. Vetterli. Multiple description transform coding of images. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 674–678, Chicago, IL, Oct. 1998.

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

[8] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray. On a stochastic sensor switching algorithm with applications to sensor scheduling and sensor coverage. *Automatica (Accepted)*.

[9] A. Ingle and V. A. Vaishampayan. Dpcm system design for diversity systems with application to packetized speech. *IEEE Transaction on Speech and Audio Processing*, 3:48–57, Jan. 1995.

[10] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold. A robust codec for transmission of very low bit-rate video over channels with bursty errors. *IEEE Transaction on Circuits and systems for Video Technology*, 10:1403–1412, Dec. 2000.

[11] X. Liu and A. J. Goldsmith. Kalman filtering with partial observation losses. Paradise Island, Bahamas, Dec. 2004. IEEE Conference on Decision and Control.

[12] D. Marco and D. L. Neuhoff. The validity of the additive noise model for uniform scalar quantizers. *IEEE Transactions on Information Theory*, Submitted.

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

[14] S. Tatikonda and S. K. Mitter. Control over noisy channels. *IEEE Transactions on Automatic Control*, 49(7):1196–1201, July 2004.

[15] S. Tatikonda and S. K. Mitter. Control under communication constraints. *IEEE Transactions on Automatic Control*, 49(7):1056–1068, July 2004.

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

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

[18] R. Venkataramani, G. Kramer, and V. K. Goyal. Multiple description coding with many channels. *IEEE Transactions on Information Theory*, 49(9):2106–2114, 2003.

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

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

[21] M. Yajnik, S. Moon, J. Kurose, and D. Towsley. Measurement and modelling of the temporal dependence in packet loss. In *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 1, pages 345–352, March 1999.

[22] R. Zamir. Gaussian codes and shannon bounds for multiple descriptions. *IEEE transactions on Information Theory*, 45(7):2629–2635, 1999.