Control Over Noisy Channels

Sekhar Tatikonda, *Member, IEEE,* and Sanjoy Mitter, *Fellow, IEEE,*

***Abstract*— Communication is an important component of distributed and networked controls systems. In our companion paper we presented a framework for studying control problems with a digital noiseless communication channel connecting the sensor to the controller [TM1]. Here we generalize that framework by applying traditional information theoretic tools of source coding and channel coding to the problem. We present a general necessary condition for observability and stabilizability for a large class of communication channels. Then we study sufficiency conditions for Internet-like channels that suffer erasures.**

***Index Terms*— Linear control, Communication, Distributed systems, Networked control**

I. Introduction

Communication is an important component of distributed and networked controls systems. A complete understanding of the interaction between control and communication will need to use tools from both control theory and information theory. In our companion paper we presented a framework for studying control problems with a digital noiseless communication channel connecting the sensor to the controller [TM1]. Here we generalize that framework by examining noisy communication channels. We apply the traditional information theoretic tools of source coding and channel coding to the controls problem. See [TM1] for a review of the relevant previous literature.

Here we study linear, discrete time, control problems with a noisy communication channel connecting the sensor to the controller. We view the initial condition and the process disturbances as the source. The job of the encoder and decoder is to transmit information about this source across the noisy channel in a causal, recursive manner. We apply the tools of information theory [CT] to determine the minimum channel capacity needed to almost surely asymptotically observe and stabilize the system. Specifically we provide a general necessary condition on the channel capacity needed to achieve the control objectives. We then show that this capacity condition is sufficient for erasure channels.

In section two we present our problem formulation and introduce the general channel model. In section three we present our necessary conditions. In section four we present sufficient conditions for observability and stabilizability over Internet-like channels that suffer from erasures.

Manuscript received March 12 2002; first revision April 15, 2003; second revision October 30, 2003. Support was provided by the Army Research Office under the MURI Grant: Data Fusion in Large Arrays of Microsensors DAAD19-00-1-0466 and the Department of Defense MURI Grant: Complex Adaptive Networks for Cooperative Control Subaward #03-132.

S. Tatikonda is with the Department of Electrical Engineering, Yale University, New Haven, CT, 06520, USA (e-mail:sekhar.tatikonda@yale.eduM. Shell is with the Georgia Institute of Technology.

II. Problem Formulation

Consider the following linear time-invariant system:

$$X_0 \in \Lambda_0, \quad X_{t+1} = AX_t + BU_t, \quad Y_t = CX_t, \quad \forall t \geq 0 \quad (1)$$

where $\{X_t\}$ is a $\mathbb{R}^d$-valued state process, $\{U_t\}$ is a $\mathbb{R}^m$-valued control process, and $\{Y_t\}$ is a $\mathbb{R}^l$-valued observation process. We have $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times m}$, and $C \in \mathbb{R}^{l\times d}$. The initial position, X_0, is distributed according to the probability density $p(X_0)$ with support on the open set $\Lambda_0 \subseteq \mathbb{R}^d$ and finite differential entropy $h(X_0)$. See Figure 1.

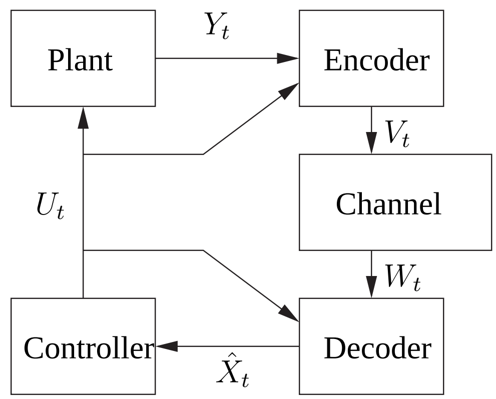

Fig. 1. System

Channel: The channel input and output alphabets are denoted by $\mathcal{V}$ and $\mathcal{W}$ respectively. Let $V^t \doteq (V_0, ..., V_t)$. The channel is modelled as a sequence of stochastic kernels $\{P(W_t \mid v^t, w^{t-1})\}$. Specifically for each realization of $(V^t, W^{t-1}) = (v^t, w^{t-1})$ the conditional probability of W_t given (v^t, w^{t-1}) is denoted by $P(W_t \mid v^t, w^{t-1})$. At time t the encoder produces a channel input symbol $V_t = v_t$ and the channel outputs the channel output symbol W_t according to the probability $P(W_t \mid v^t, w^{t-1})$. Some typical examples of channels include:

- **Noiseless digital channel with rate R**
  The channel input and output alphabets are the same: $\mathcal{V} = \mathcal{W}$. The alphabet size is $|\mathcal{V}| = 2^R$ where R is called the *rate* of the channel. The channel is noiseless and memoryless:

  $$p(W_t \mid v^t, w^{t-1}) = \begin{cases} 1 & \text{if } W_t = v_t \\ 0 & \text{if } W_t \neq v_t \end{cases}$$

  This is the channel we examined in our companion paper [TM1].
- **Delayed noiseless digital channel with delay Δ**
  This is a noiseless digital channel with delay Δ (Δ is a

nonnegative integer):

$$p(W_t \mid v^t, w^{t-1}) = \begin{cases} 1 & \text{if } W_t = v_{t-\Delta} \\ 0 & \text{if } W_t \neq v_{t-\Delta} \end{cases}$$

- **Erasure channel with erasure probability** α
  The channel input alphabet has size $|\mathcal{V}| = 2^R$. The channel output alphabet is $\mathcal{W} = \mathcal{V} \cup \{$ erasure symbol $\}$. The channel is memoryless with erasure probability $\alpha \in [0,1]$:

$$p(W_t \mid v^t, w^{t-1}) = \begin{cases} 1-\alpha & \text{if } W_t = v_t \\ \alpha & \text{if } W_t = \text{ erasure symbol} \\ 0 & \text{else} \end{cases}$$

  Thus with probability α the *packet* of R bits is erased. This channel is often used as a simplified model of packet loss on Internet-like channels.
- **Memoryless Gaussian channel with power** ρ
  The channel input and output alphabets are the real line: $\mathcal{V} = \mathcal{W} = \mathbb{R}$. The channel is memoryless with power ρ:

$$W_t = V_t + N_t$$

  where N_t is a Gaussian random variable with mean zero and variance 1. The input symbol V_t satisfies the power constraint: $E(V_t^2) \leq \rho$. This channel is often used as a simplified model of a wireless channel.

A. Information Pattern

The control problems we look at involve the design of an encoder, decoder, and controller. Just as in [TM1] we specify the information pattern [Wit] of each component. The difference here is the addition of a more general channel.

a) Encoder:: The encoder at time t is a map

$$\mathcal{E}_t : \mathbb{R}^{l(t+1)} \times \mathcal{V}^t \times \mathbb{R}^{mt} \to \mathcal{V} \quad \text{taking} \quad (Y^t, V^{t-1}, U^{t-1}) \mapsto V_t.$$

b) Decoder:: The decoder at time t is a map

$$\mathcal{D}_t : \mathcal{W}^{t+1} \times \mathbb{R}^{mt} \to \mathbb{R}^d \quad \text{taking} \quad (W^t, U^{t-1}) \mapsto \hat{X}_t.$$

The output of the decoder is an estimate of the state of the plant.

c) Controller:: The controller at time t is a map

$$\mathcal{C}_t : \mathbb{R}^d \to \mathbb{R}^m \quad \text{taking} \quad \hat{X}_t \mapsto U_t.$$

Note that we are assuming the controller takes as input only the decoder's state estimate. Hence we are assuming a separation structure between the decoder and the controller.

III. Necessary Conditions

Here we examine observability and stabilizability over general communication channels. See the appendix for background material on information theory.

Definition 3.1: Let the *error* be $E_t = X_t - \hat{X}_t$ where $\hat{X}_t$ is the state estimate. System (1) is *almost surely asymptotically observable* if there exists a control sequence $\{U_t\}$ and an encoder and decoder such that the state estimation error $\|E_t\|_2 \to 0$ almost surely. System (1) is *almost surely asymptotically stabilizable* if there exists an encoder, decoder, and controller such that $\|X_t\|_2 \to 0$ almost surely.

In [TM1] we provided necessary rate conditions for the noiseless digital channel under the stronger conditions of asymptotic observability and asymptotic stabilizability (as opposed to the almost sure version of the definition given above.) We repeat propositions 3.1 and 3.2 of [TM1] here:

Proposition 3.1: A necessary condition on the rate for asymptotic observability is $R \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$. A necessary condition on the rate for asymptotic stabilizability is $R \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$.

Our goal is to determine properties of the channel that ensure almost sure asymptotic observability and stabilizability for general channels. To that end we need a measure of channel quality. Shannon's channel capacity turns out to be the correct measure.

Channel Capacity: Given a channel $\{P(W_t \mid v^t, w^{t-1}\}$, the *Shannon capacity* over a time horizon of length T is defined as the supremum of the mutual information over all channel input distributions $P(V^{T-1})$. Specifically

$$C_T^{\text{cap}} = \sup_{P(V^{T-1})} I(V^{T-1}; W^{T-1})$$

where $I(\cdot;\cdot)$ is the mutual information [CT]. (See the appendix for a review of mutual information.) Here time starts at zero hence $V^{T-1} = (V_0, ..., V_{T-1})$. We list the channel capacity for the channels described above:

- **Noiseless digital channel with rate** R**:** $C_T^{\text{cap}} = TR$.
- **Delayed noiseless digital channel with delay** Δ**:** $C_T^{\text{cap}} = (T-\Delta)R$.
- **Erasure channel with erasure probability** α**:** $C_T^{\text{cap}} = (1-\alpha)TR$.
- **Memoryless Gaussian channel with power** ρ**:** $C_T^{\text{cap}} = \frac{T}{2}\log(1+\rho)$. Here the supremization in the definition of Shannon capacity is over all $P(V^{T-1})$ such that $E(V_i^2) \leq \rho, \quad \forall i$.

Rate-Distortion: As we have seen in [TM1] we need to be able to transmit information about the initial condition to the decoder and controller. One way to measure how much information is needed to reconstruct the initial condition to some distortion fidelity is given by the rate distortion function.

Let the source X have distribution $P(X)$. Let $d(x,\hat{x})$ be a distortion measure. Here a distortion measure is any nonnegative function that measures the relative fidelity in reconstructing x by $\hat{x}$. Given a source $P(X)$ the *rate distortion* function is defined as the infimum of the mutual information over all channels, $P(\hat{X}|x)$, that satisfy the distortion condition [CT]:

$$R(D) = \inf_{P(\hat{X} \mid x)} \left\{ I(X;\hat{X}) \quad \text{such that} \quad E\left(d(X,\hat{X})\right) \leq D \right\}.$$

Note that the expectation is taken with respect to the joint measure $P(x,\hat{x}) = P(\hat{x}|x)P(x)$.

We will find the following parameterized family of distortion measures useful in determining conditions for almost sure observability and stabilizability:

$$d^\epsilon(x,\hat{x}) \doteq \begin{cases} 0 & \text{if } \|x-\hat{x}\|_2 \leq \epsilon \\ 1 & \text{if } \|x-\hat{x}\|_2 > \epsilon \end{cases} \qquad \text{where } \epsilon > 0.$$

This choice of distortion measure will allow us to compute the probability that X and $\hat{X}$ are farther than ϵ apart. Specifically $E\left(d^{\epsilon}(X,\hat{X})\right) = \Pr(\|X-\hat{X}\|_2 > \epsilon)$.

Data-Processing Inequality: The traditional information theoretic setup involves a source X that we wish to transmit over a channel $P(W|v)$ and produce a reconstruction $\hat{X}$ satisfying some fidelity criterion. We have discussed both the rate distortion function and the Shannon capacity.

A necessary condition for reconstructing X upto some distortion D using the channel once is

$$R(D) \leq C_1^{\text{cap}} \tag{2}$$

To prove (2) we will need the following data-processing inequality whose proof can be found in [CT].

Lemma 3.1: Let $X \to V \to W \to \hat{X}$ be a Markov chain then $I(X;\hat{X}) \leq I(V;W)$.

We can generalize our encoder and decoder by modelling them as stochastic kernels. Deterministic encoders and decoders can be modelled as stochastic kernels that are Dirac measures. Then for any encoder $P(V|x)$ and decoder $P(\hat{X}|w)$ such that the resulting joint distribution $P(X,V,W,\hat{X})$ satisfies the distortion bound $E(d(X,\hat{X})) \leq D$ we have:

$$\begin{aligned} R(D) &= \inf_{\substack{P(\hat{X}|x) \\ E_P(d(X,\hat{X})) \leq D}} I(X;\hat{X}) \\ &\leq I(X;\hat{X}) \\ &\leq I(V;W) \\ &\leq \sup_{P(V)} I(V;W) \\ &= C_1^{\text{cap}}. \end{aligned}$$

Thus we have shown (2). More generally we will want to reconstruct the source X by using the channel T times instead of just once. In this case $X \to (V_0,...,V_{T-1}) \to (W_0,...,W_{T-1}) \to \hat{X}$ forms a Markov chain. Thus $I(X;\hat{X}) \leq I(V^{T-1},W^{T-1})$ and a necessary condition for reconstruction is $R(D) \leq C_T^{\text{cap}}$.

The following technical lemma gives a lower bound on the rate distortion function for reconstructing $X_t = A^t X_0$ at time t under the distortion measure $d^{\epsilon}(x,\hat{x})$.

Lemma 3.2: Assume X_0 has density $p(X_0)$ with finite differential entropy $h(X_0)$. Let $R_t^{\epsilon}(D)$ represent the rate distortion function for the source $X_t = A^t X_0$ under the distortion measure $d^{\epsilon}(x,\hat{x})$. Then

$$\begin{aligned} R_t^{\epsilon}(D) &\geq t(1-D)\sum_{\lambda(A)} \log|\lambda(A)| \\ &\quad + \left((1-D)h(X_0) - \log(K_d \epsilon^d) - \frac{1}{2}\right) \end{aligned}$$

where K_d is the constant in the formula for the volume of a d-dimensional sphere.

Proof: Let $\delta_t = d^{\epsilon}(X_t,\hat{X}_t)$. Let $P(X_t,\hat{X}_t)$ be any joint distribution such that the distortion constraint is met: $E\left(d^{\epsilon}(X_t,\hat{X}_t)\right) = \Pr(\|X_t-\hat{X}_t\|_2 > \epsilon) \leq D$. Hence the $\Pr(\delta_t = 1) \leq D$. Then

$$\begin{aligned} & I(X_t;\hat{X}_t) \\ =\ & I(X_t\ ;\ \delta_t,\hat{X}_t) - I(X_t;\ \delta_t \mid \hat{X}_t) \\ =\ & I(X_t\ ;\ \delta_t,\hat{X}_t) - \left(H(\delta_t \mid \hat{X}_t) - H(\delta_t \mid X_t,\hat{X}_t)\right) \\ \stackrel{(a)}{=}\ & I(X_t\ ;\ \delta_t,\hat{X}_t) - H(\delta_t \mid \hat{X}_t) \\ =\ & h(X_t) - h(X_t \mid \delta_t,\hat{X}_t) - H(\delta_t \mid \hat{X}_t) \\ =\ & h(X_t) - h(X_t \mid \delta_t = 0,\hat{X}_t)\Pr(\delta_t = 0) \\ & -h(X_t \mid \delta_t = 1,\hat{X}_t)\Pr(\delta_t = 1) - H(\delta_t \mid \hat{X}_t) \\ \stackrel{(b)}{\geq}\ & h(X_t) - h(X_t - \hat{X}_t \mid \delta_t = 0,\hat{X}_t) \\ & -h(X_t \mid \delta_t = 1,\hat{X}_t)D - \frac{1}{2} \\ \stackrel{(c)}{\geq}\ & h(X_t) - h(X_t - \hat{X}_t \mid \delta_t = 0) - h(X_t)D - \frac{1}{2} \\ \stackrel{(d)}{\geq}\ & h(X_t) - \log(K_d\epsilon^d) - h(X_t)D - \frac{1}{2} \\ =\ & h(A^t X_0)(1-D) - \log(K_d\epsilon^d) - \frac{1}{2} \\ \stackrel{(e)}{=}\ & (1-D)\left(t\log|A| + h(X_0)\right) - \log(K_d\epsilon^d) - \frac{1}{2} \end{aligned}$$

where (a) follows because δ_t is a deterministic function of X_t and $\hat{X}_t$. Point (b) follows because δ_t is a binary value random variable whose discrete entropy cannot be larger than $\frac{1}{2}$. Point (c) follows because conditioning reduces entropy. Point (d) follows because $\|X_t - \hat{X}_t\|_2 \leq \epsilon$ and the uniform distribution maximizes the continuous entropy over all random variables with bounded support. Point (e) follows because $h(A^t X_0) = t\log|A| + h(X_0)$. (See the appendix.)

The lower bound is independent of $P(\hat{X}_t|x_t)$. Hence

$$\begin{aligned} R_t^{\epsilon}(D) &= \inf_{P(\hat{X}_t|x_t)} I(X_t;\hat{X}_t) \\ &\geq t(1-D)\sum_{\lambda(A)} \log|\lambda(A)| \\ &\quad + \left((1-D)h(X_0) - \log(K_d\epsilon^d) - \frac{1}{2}\right). \end{aligned}$$

□

For any given channel define $C^{\text{cap}} \doteq \liminf_{T\to\infty} \frac{1}{T} C_T^{\text{cap}}$. We now present our necessary conditions for almost sure observability and stabilizability for general channels.

Proposition 3.2: For system (1) a necessary condition on the channel capacity for almost sure asymptotic observability is $C^{\text{cap}} \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$.

Proof: Assume that there exists an encoder and decoder such that system (1) is almost surely asymptotically observable. As in proposition 3.1 of [TM1] we see that, possibly after a coordinate transformation, the matrix A can be written in the form

$$\begin{bmatrix} A_s & \\ & A_u \end{bmatrix}$$

where the A_s block corresponds to the stable subspace (that subspace corresponding to the eigenvalues of A that are strictly inside the unit circle) and the A_u block corresponds to the marginally stable and unstable subspace (that subspace corresponding to the eigenvalues of A that are either on the unit circle or outside the unit circle.)

Let Π_s represent the projection onto the stable subspace. Fix an arbitrary control sequence $\{U_t\}$. Then $X_t = A^t X_0 + \alpha_t$ where $\alpha_t = \sum_{j=0}^{t-1} A^{t-1-j} B U_j$. For any control sequence we have $\lim_{t\to\infty} \Pi_s(X_t - \alpha_t) = 0$. Thus knowledge of the control signals alone is enough to estimate the projection of the state onto the stable subspace. Thus, without loss of generality, we can restrict our attention to A matrices that contain only unstable eigenvalues.

By almost sure asymptotic observability we know that for any $\epsilon > 0$ there exists a $T(\epsilon)$ such that the error $E_t = X_t - \hat{X}_t$ satisfies

$$\Pr\left(\sup_{t\geq T(\epsilon)} \|E_t\|_2 > \epsilon\right) \leq \epsilon.$$

Thus for $t \geq T(\epsilon)$ we have

$$\begin{aligned} & E\left(d^\epsilon(X_t, \hat{X}_t)\right) \\ = \;& 0 \times \Pr(\|X_t - \hat{X}_t\|_2 \leq \epsilon) + 1 \times \Pr(\|X_t - \hat{X}_t\|_2 > \epsilon) \\ \leq \;& \epsilon. \end{aligned}$$

Then by the data processing inequality and lemma 3.2 the channel capacity and rate distortion function must satisfy for all $t \geq T(\epsilon)$:

$$\begin{aligned} \frac{1}{t} C_t^{\text{cap}} &\geq \frac{1}{t} R_t^\epsilon(\epsilon) \\ &\geq (1-\epsilon) \sum_{\lambda(A)} \log|\lambda(A)| \\ &\quad + \frac{1}{t}\left((1-\epsilon)h(X_0) - \log(K_d \epsilon^d) - \frac{1}{2}\right). \end{aligned}$$

Hence

$$\begin{aligned} C^{\text{cap}} &= \liminf_{t\to\infty} \frac{1}{t} C_t^{\text{cap}} \\ &\geq \liminf_{t\to\infty} \left[(1-\epsilon)\sum_{\lambda(A)} \log|\lambda(A)| \right. \\ &\quad \left. + \frac{1}{t}\left((1-\epsilon)h(X_0) - \log(K_d\epsilon^d) - \frac{1}{2}\right)\right] \\ &= (1-\epsilon)\sum_{\lambda(A)} \log|\lambda(A)|. \end{aligned}$$

Since ϵ can be chosen arbitrarily small we see $C^{\text{cap}} \geq \sum_{\lambda(A)} \log|\lambda(A)|$ and, if we reintroduce A matrices with some stable eigenvalues, we get $C^{\text{cap}} \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$. □

Proposition 3.3: For system (1) with (A, B) a stabilizable pair a necessary condition on the channel capacity for almost sure asymptotic stabilizability is $C^{\text{cap}} \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$.

Proof: Assume there exists an encoder, decoder, and controller such that the system (1) is almost surely asymptotically stabilizable. For a given control sequence $U_0, U_1, ..., U_{t-1}$ we have

$$X_t = A^t X_0 - \alpha_t(U_0, ..., U_{t-1})$$

where

$$\alpha_t(U_0, ..., U_{t-1}) \doteq -\sum_{i=0}^{t-1} A^{t-1-i} B U_i.$$

Almost sure asymptotic stabilizability implies that for any ϵ there exists a $T(\epsilon)$ such that

$$\Pr\left(\sup_{t\geq T(\epsilon)} \|X_t\|_2 > \epsilon\right) \leq \epsilon.$$

We can view α_t as a reconstruction of $A^t X_0$ with distortion $E\left(d^\epsilon(A^t X_0, \alpha_t)\right) \leq \epsilon$. By proposition 3.2 a necessary condition to achieve this distortion is $C^{\text{cap}} \geq \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$. □

In the previous proposition we interpreted the following function of the control signals, $\alpha_t(U_0, ..., U_{t-1}) = -\sum_{i=0}^{t-1} A^{t-1-i} B U_i$, as a reconstruction of $A^t X_0$. We can view a particular sequence of control signals as a "codeword" in a reconstruction codebook [CT].

In proving the necessary conditions above we did not need to explicitly describe the encoder, decoder, and controller nor did we use the assumption of separation between the observer and the controller. Hence the conditions hold independently of the choice of these components. In the next section we will provide explicit constructions of the encoder, decoder, and controller that can achieve almost sure asymptotic observability and stabilizability for the erasure channel.

IV. Achievability Results

In this section we first quickly review our achievability results from [TM1] and then treat control over an erasure channel.

Recall that the encoder at time t is a map $\mathcal{E}_t$ that takes $(Y^t, V^{t-1}, U^{t-1}) \mapsto V_t$. In this case the encoder knows the past states, past channel input symbols, and past controls. In our companion paper we distinguished between two different encoder classes: one where the encoder observes the control signals, called *encoder class 1*, and one where it does not, called *encoder class 2* [TM1]. In this paper we restrict our attention to the situation where the encoder observes the control signals being applied to the plant.

Often times the rate condition for the noiseless digital channel will not be an integer. We can achieve an average rate by employing a time-sharing scheme as discussed in [TM1]. Hence the statement "a rate R can be achieved" should be interpreted to mean a fixed rate in the case R is an integer and an average rate in the case R is a not an integer. We repeat propositions 5.3 and 5.4 of [TM1] here:

Proposition 4.1: For system (1) with (A, C) an observable pair a sufficient condition for asymptotic observability over a noiseless digital channel is $R > \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$. For system (1) with (A, C) an observable pair and (A, B) a stabilizable pair a sufficient condition on the rate for asymptotic stabilizability over a noiseless digital channel is $R > \sum_{\lambda(A)} \max\{0, \log|\lambda(A)|\}$.

We will need the following technical lemma 5.1 of [TM1]:

Lemma 4.1: Let A be a stable matrix. Let B_t be a set of matrices such that $\|B_t\| \leq L < \infty$ and $\lim_{t\to\infty} \|B_t\| = 0$. Let $S_t = \sum_{i=0}^{t-1} A^{t-1-i} B_i$. Then $\lim_{t\to\infty} \|S_t\| = 0$.

From section 3 we know that the Shannon capacity of an erasure channel with erasure probability α over T channel uses is $C_T^{\text{cap}} = (1-\alpha)TR$. At each time step this channel will with probability $1-\alpha$ deliver a "packet" of size R bits and with probability α drop that "packet."

From proposition 3.2 we know a necessary condition for almost sure asymptotic observability is $C^{\text{cap}} = (1-\alpha)R \geq \sum_{\lambda(A)} \max\{0, |\log \lambda(A)|\}$. Hence we require a packet size of at least

$$R \geq \frac{1}{1-\alpha} \sum_{\lambda(A)} \max\{0, |\log \lambda(A)|\}.$$

Now we examine sufficiency. To that end we will extend the erasure channel model to include acknowledgements. Specifically the decoder will feed back to the encoder an acknowledgment whether the packet was erased or not. This acknowledgment feature is common in the TCP network protocol. The encoder then knows what information has been delivered to the decoder. Hence, in the language of [TM1], we say that the encoder and decoder are equi-memory. Because the erasure channel is memoryless, acknowledgement feedback cannot increase the channel capacity [CT]. Hence the necessity condition above continues to hold for erasure channels with acknowledgement feedback. We discuss how to relax this acknowledgment feature at the end of this section.

For simplicity we consider the system (1), $X_{t+1} = AX_t + BU_t$, with full state observation, $C = I$, at the encoder. The partially observed case can be treated in the manner described in [TM1].

Proposition 4.2: Given system (1), a bound on Λ_0, and an erasure channel with erasure probability α and feedback acknowledgements the packet size $R > \frac{1}{1-\alpha} \sum_{\lambda(A)} \max\{0, |\log \lambda(A)\}$ is sufficient to ensure almost sure asymptotic observability.

Proof: We first treat the scalar case: $X_{t+1} = aX_t + bU_t$. Let $E_t = X_t - \hat{X}_t$ and $E_0 = X_0 \in \Lambda_0 \subseteq [-L_0, L_0]$. At time t let L_t represent the box that the error lives in: $E_t \in [-L_t, L_t]$. We will construct a scheme such that $L_t \to 0$ almost surely and hence $E_t \to 0$ almost surely.

The decoder feeds back acknowledgments to the encoder. Hence the encoder can compute the decoder's uncertainty set $[-L_t, L_t]$. At time $t+1$ the encoder partitions the interval $[-|a|L_t,\ |a|L_t]$ into 2^R equal sized regions and sends the index of that region across the channel. If the erasure channel does not drop the packet then $L_{t+1} = \frac{|a|}{2^R} L_t$. If the packet is dropped then $L_{t+1} = |a|L_t$. This can be described by the stochastic difference equation: $L_{t+1} = |a|F_t L_t$ where the random variables F_t are IID with common distribution: $\Pr(F_t = 1) = \alpha$ and $\Pr(F_t = 2^{-R}) = 1-\alpha$.

Since $L_t = L_0 \prod_{j=0}^{t-1} |a|F_j$ we need to show that $\prod_{j=0}^{t-1} |a|F_j \to 0$ almost surely. By the strong law of large numbers we know $\frac{1}{t}\sum_{j=0}^{t-1} \log |a|F_j \to E(\log |a|F)$ almost surely.

If $E(\log |a|F) < 0$ then

$$\prod_{j=0}^{t-1} |a|F_j = 2^{t(\frac{1}{t}\sum_{j=0}^{t-1} \log |a|F_j)} \to 0 \text{ a.s.}$$

This result can be found in any standard text on large deviations. See for example [DZ]. Now

$$E(\log |a|F) = \alpha \log |a| + (1-\alpha) \log \frac{|a|}{2^R} = \log |a| - (1-\alpha)R.$$

Thus $E(\log |a|F)$ is negative if and only if $R > \frac{\log |a|}{1-\alpha}$.

In the vector case the stochastic difference equation takes the form $\underline{L}(t+1) = \overline{\Upsilon} F_{\underline{R}}(t) \underline{L}(t)$ where

$$P(F_{\underline{R}}(t)) = \begin{cases} \alpha & \text{if } F_{\underline{R}}(t) = \text{diag}(1, ..., 1) \\ 1-\alpha & \text{if } F_{\underline{R}}(t) = \text{diag}(2^{-R_1}, ..., 2^{-R_d}) \end{cases}$$

and $\overline{\Upsilon}$ is described in section 4 of [TM1]. Since $\underline{L}(0)$ is bounded we need only show that $\prod_{j=0}^{t-1} \overline{\Upsilon} F_{\underline{R}}(j)$ converges to zero almost surely. Since $\overline{\Upsilon}$ is upper triangular and $F_{\underline{R}}(t)$ is a random diagonal matrix we see from the argument above that each eigenvalue of $\prod_{j=0}^{t-1} \overline{\Upsilon} F_{\underline{R}}(j)$ converges to zero almost surely if and only if $R_i > \frac{\log |\lambda_i|}{1-\alpha}$ for each $i = 1, ..., d$. □

Proposition 4.3: Given an erasure channel with erasure probability α and feedback acknowledgments the packet size $R > \frac{1}{1-\alpha} \sum_{\lambda_A} \max\{0, |\log \lambda(A)\}$ is sufficient to ensure almost sure asymptotic stabilizability.

Proof: Let K be a stabilizing controller, i.e. $A + BK$ is stable. Apply the certainty equivalent controller $U_t = K\hat{X}_t$ where $\hat{X}_t$ is the decoder's state estimate. As before let $e_t = X_t - \hat{X}_t$. Then

$$X_t = (A+BK)^t X_0 - \sum_{j=0}^{t-1} (A+BK)^{t-1-j} BKe_j$$

Since $A + BK$ is stable the first addend in the above equation goes to zero almost surely. By proposition 4.3 we know that the state estimation error converges to zero almost surely: $\lim_{t\to\infty} \|e_t\| = 0\ a.s.$ Hence by lemma 4.1 the second addend goes to zero almost surely. □

Now we consider the case when there are process disturbances:

$$X_{t+1} = AX_t + BU_t + Z_t, \quad Y_t = X_t, \quad t \geq 0 \qquad (3)$$

where $\|Z_t\|_2 \leq D$.

Proposition 4.4: Given system (3), a bound on the Λ_0, and an erasure channel with erasure probability α and feedback acknowledgements the packet size $R > \frac{1}{1-\alpha} \sum_{\lambda_A} \max\{0, |\log \lambda(A)\}$ is sufficient to ensure that the state estimation error is bounded almost surely.

Proof: We first treat the scalar case: $X_{t+1} = aX_t + bU_t + Z_t$. Assume that $E_0 = X_0 \in [-L_0, L_0]$. At time t let L_t be such that $E_t \in [-L_t, L_t]$. We will construct a scheme such that the sequence L_t is bounded almost surely.

Just as in proposition 4.3 the encoder can compute the decoder's uncertainty set $[-L_t, L_t]$. If the erasure channel does not drop the packet at time t then $L_{t+1} = \frac{|a|}{2^R} L_t + D$. If the packet is dropped then $L_{t+1} = |a|L_t + D$. This is described by the stochastic difference equation:

$$L_{t+1} = |a|F_t L_t + D$$

where the random variables F_t are IID with common distribution: $\Pr(F_t = 1) = \alpha$ and $\Pr(F_t = 2^{-R}) = 1 - \alpha$. Now

$$L_t = L_0 \prod_{i=0}^{t-1} |a| F_i + \sum_{i=0}^{t-1} \left(\prod_{j=i+1}^{t-1} |a| F_j \right) D.$$

By proposition 4.3 we know that if $R > \frac{\log|a|}{1-\alpha}$ then the first addend converges to zero almost surely. We need to show that the second addend converges almost surely to a finite limit.

First note that $\sum_{i=0}^{t-1} \left(\prod_{j=0}^{i-1} |a| F_j \right)$ has the same distribution as $\sum_{i=0}^{t-1} \left(\prod_{j=i+1}^{t-1} |a| F_j \right)$. Choose δ so that $E(\log|a|F) + \delta < 0$. By the strong law of large numbers we have

$$\begin{aligned} \limsup_{T\to\infty} \left(\prod_{i=0}^{T-1} |a| F_i \right)^{\frac{1}{T}} &= \limsup_{T\to\infty} 2^{\frac{1}{T}\sum_{i=0}^{T-1} \log|a|F_i} \\ &\leq 2^{(E(\log|a|F)+\delta)} \\ &< 1 \quad \text{almost surely.} \end{aligned}$$

Hence, by applying Cauchy's root criterion, we see that the series $\lim_{t\to\infty} \sum_{i=0}^{t-1} \left(\prod_{j=0}^{i-1} |a| F_j \right)$ converges almost surely.

For the vector case, we know from proposition 5.2 of [TM1], that the stochastic difference equation takes the form

$$\underline{L}(t+1) = \overline{\Upsilon} F_{\underline{R}}(t) \underline{L}(t) + D \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

where

$$P(F_{\underline{R}}(t)) = \begin{cases} \alpha & \text{if } F_{\underline{R}}(t) = \mathrm{diag}(1, ..., 1) \\ 1 - \alpha & \text{if } F_{\underline{R}}(t) = \mathrm{diag}(2^{-R_1}, ..., 2^{-R_d}) \end{cases}$$

As in the scalar case we need to show that the product

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} \left(\prod_{j=0}^{i-1} \overline{\Upsilon} F_{\underline{R}}(j) \right)$$

converges almost surely. Since $\overline{\Upsilon}$ is upper triangular and $F_{\underline{R}}(t)$ is a random diagonal matrix we see from the argument above that this series converges if and only if $R_i > \frac{\log|\lambda_i|}{1-\alpha}$ for each $i = 1, ..., d$. See theorem 1.1 of [BP] and [Kes] for more details. □

In propositions 4.2-4.4 we assumed that there exists acknowledgement feedback from the decoder to the encoder. Relaxing this assumption is in general difficult. There are, though, a few scenarios where we do not need an explicit feedback acknowledgement. We discuss two here. Both require signaling the occurrence of an erasure via the control signal U_t. In this way the control takes on a "dual effect:" that of satisfying the control objective and of helping the encoder/decoder estimate the state. This signalling will ensure that the encoder continues to track the decoder's estimate of the state.

Scenario 1: Here we assume that the encoder knows the control policy K where $U_t = K\hat{X}_t$. We will prove using induction that the encoder can compute the decoder's estimate at each time step. At time zero the encoder knows the decoder's state estimate. Assume that at time $t-1$ the encoder knows the decoder's state estimate: $\hat{X}_{t-1}$. At time t the decoder's state estimate, based on $\hat{X}_{t-1}$ and the channel message, can take one of two values depending on whether there was an erasure or not. Hence the control U_t can take one of two values. The encoder, by observing U_t and using its knowledge of the control law K, can determine whether an erasure has occurred or not and hence can determine the decoder's estimate $\hat{X}_t$. Thus the encoder can compute the decoder's estimate at each time step.

Scenario 2: Here we assume that the controller adds signalling information, β_t, to the control signal: $U_t = K\hat{X}_t + \beta_t$. Then

$$X_t = (A + BK)^t X_0 - \sum_{j=0}^{t-1} (A + BK)^{t-i-j} B(Ke_j - \beta_j).$$

By lemma 4.1, if $\lim_{t\to\infty} \beta_t = 0$ then the sum $\lim_{t\to\infty} \sum_{j=0}^{t-1} (A + BK)^{t-i-j} B\beta_j = 0$ and hence does not effect the long term behavior of the state. We now show how to choose β_t. Fix an integer M and assume that the controller knows if an erasure has occurred or not. Let

$$\beta_t = \begin{cases} -2^{-Mt} \lfloor K\hat{X}_t \rfloor_{2^{-Mt}} & \text{if erasure} \\ -2^{-Mt} \lfloor K\hat{X}_t \rfloor_{2^{-Mt}} + 2^{-Mt}\text{ones(m)} & \text{if no erasure} \end{cases}$$

where $\lfloor K\hat{X}_t \rfloor_{2^{-Mt}}$ is a $\{0,1\}^m$-valued vector that contains the coefficient of 2^{-Mt} in the component-wise binary expansion of the vector $K\hat{X}_t$ and ones(m) is the $m-$dimensional vector of all ones. Note that $\beta_t \to 0$. The controller applies the control $U_t = K\hat{X}_t + \beta_t$ to the plant. In words U_t replaces the coefficient of 2^{-Mt} in the binary expansion of $K\hat{X}_t$ by a vector of all zeroes or all ones depending on whether there was an erasure or not. The encoder observes the control applied. Thus it can determine the coefficient of 2^{-Mt} in the binary expansion of U_t. Hence the encoder will know if an erasure has occurred or not. Thus the encoder can compute the decoder's estimate at each time step.

Neither scenario is completely satisfactory. The first case assumes the encoder knows the control policy. The second case is not robust if there is noise on the channel connecting the controller to the plant. But both cases show that the necessary conditions presented in propositions 3.2 and 3.3 are tight even for scenarios without explicit acknowledgement feedback.

V. Conclusion

In this paper we have been concerned with almost sure asymptotic observability and stabilizability. Sahai, in [Sa], [Sa2], treats the case of mean-square observability. In his *anytime capacity* framework he presents channel capacity results that ensure mean-square observability. In general the capacity conditions are different under the almost sure and the mean-square convergence criteria.

Depending on the control application, one may prefer an almost sure convergence criteria or a mean-square convergence criteria. In the former one is interested in finding a channel capacity so that almost all realizations of the system's trajectories are typical. And in fact with probability one all realizations will satisfy the control objective. Atypical realizations, also called large deviations excursions, can occur but with probability approaching zero. If in addition, one wants to penalize the atypical trajectories by the size of their large deviation excursion then the mean-square formulation is appropriate. The fact that one gets different results under the almost sure convergence criteria and the mean-square convergence criteria is a generic property of the multiplicative law of large numbers [DZ].

In this paper we examined linear systems with a communication channel connecting the plant to the controller. We generalized the necessity conditions first presented in [TM1] to general noisy channels. We then examined control over Internet-like channels that suffer erasures. Two important future research directions include generalizing the achievability results to more general classes of channels and analyzing erasure channels without acknowledgement feedback.

Appendix

The mutual information between two random variables X and Y with distribution $P(X,Y)$ is defined as

$$I(X;Y) \doteq \begin{cases} \int \log \frac{dP}{dQ} dP & \text{if } P << Q \\ +\infty & \text{else} \end{cases}$$

where $Q(x,y) = P(x) \times P(y)$ and $\frac{dP}{dQ}$ is the Radon-Nikodym derivative. The mutual information can be seen as a measure of the dependence between X and Y. Under the measure Q the random variables X and Y are independent. Note that the mutual information is a function of $P(X,Y) = P(Y|x)P(X)$. In the capacity computation one supremizes the mutual information over input distributions $P(X)$ and in the rate distortion computation one infimizes the mutual information over *forward channels* $P(Y|x)$ subject to the distortion criterion.

If X is a discrete random variable then its entropy is defined as:

$$H(X) \doteq -\sum_i P(X = x_i) \log P(X = x_i)$$

and its conditional entropy is defined as:

$$H(X|Y) \doteq -\int \left(\sum_i P(X = x_i|y) \log P(X = x_i|y) \right) p(dy).$$

If X is a random variable admitting a density, p_X, then its differential entropy is defined as:

$$h(X) \doteq -\int p_X(x) \log p_X(x) dx$$

and its conditional differential entropy is defined as:

$$h(X|Y) \doteq -\int \left(\int p_{X|Y}(x|y) \log p_{X|Y}(x|y) dx \right) p(dy).$$

The following useful properties can be found in [CT]:

(a) $I(X;Y) \geq 0$ and $I(X;Y) = \begin{cases} H(Y) - H(Y|X) & \text{if } Y \text{ is a discrete random variable} \\ h(Y) - h(Y|X) & \text{if } Y \text{ admits a density for each } x \end{cases}$
This implies conditioning reduces entropy.

(b) If X is a vector valued random variable admitting a density then $h(AX) = h(X) + \log |A|$ where $|A|$ is the absolute value of the determinant of A.

(c) If $Z = f(X)$ where Z is discrete then $H(Z \mid X) = 0$.

(d) $I(X;Z \mid Y) = I(X,Y;Z) - I(Y;Z)$

(e) $X \to Y \to Z$ forms a Markov chain if and only if $I(X;Z \mid Y) = 0$.

There are two extremal properties that will be important to us. If X is a discrete random variable taking on M values then $H(X) \leq \log M$. If X admits a density with bounded support Λ then $h(X) \leq \log(\text{volume}(\Lambda))$ [CT].

Acknowledgment

The authors would like to thank Nicola Elia and Anant Sahai for many helpful discussions and the reviewers for their detailed comments.

References

[BP] P. Bougerol and N. Picard, “Strict Stationarity of Generalized Autoregressive Processes.” The Annals of Probability, Volume 20, Number 4, pp. 1714-1730, October 1992.

[CT] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley, New York, 1991.

[DZ] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer-Verlag, 1998.

[Kes] H. Kesten, “Random Difference Equations and Renewal Theory for Products of Random Matrices.” Acta Math., Vol. 131, pp. 207-248, 1973.

[Sa] A. Sahai, “Evaluating Channels for Control: Capacity Reconsidered.” Proceedings of the American Control Conference, Vol. 4 , June 2000, pp. 2358 -2362.

[Sa2] A. Sahai, “Any-time Capacity and a Separation Theorem for Tracking Unstable Processes.” Proceedings of the IEEE International Symposium on Information Theory, June 2000.

[Tat] S. Tatikonda, “Control Under Communication Constraints.” MIT Ph.D. thesis, August 2000.

[TM1] S. Tatikonda and S. Mitter, “Control Under Communication Constraints.” Submitted to the IEEE Trans. on Automatic Control, February 2002.

[Wit] H. Witsenhausen, “Separation of Estimation and Control for Discrete Time Systems.” Proceedings of the IEEE, Volume 59, No. 11, November 1971.