On The Rate-Cost of Gaussian Linear Control Systems with Random Communication Delays

Jia Zhang, Purdue ECE
Email: zhan2030@purdue.edu

Chih-Chun Wang, Purdue ECE
Email: chihw@purdue.edu

***Abstract*—This work considers Gaussian linear control systems where the goal is to guarantee *system-state mean-square stability* (i.e., $E\left(\|\mathbf{x}(t)\|^2\right) \leq \mathsf{D}$) while minimizing the traffic rate R between the sensor(s) and the controller. Most existing results either assume zero delay or focus on the asymptotic setting that overlooks the impact of delay. Nonetheless, in practice the communication delay is randomly distributed due to varying channel/network conditions. When the sensor measurement finally arrives at the controller, the *age-of-information* is thus random. Heuristically, an "old" measurement provides less valuable information than a "young" measurement but the quantitative impact of random delay on the optimal rate-cost tradeoff $R^*(\mathsf{D})$ remains an open problem. This work provides the first lower bound $R_{\rm LB}(\mathsf{D})$ for the random delay setting and designs a simple scheme that leads to a numerically evaluated upper bound $R_{\rm UB}(\mathsf{D})$. Jointly $R_{\rm LB}(\mathsf{D})$ and $R_{\rm UB}(\mathsf{D})$ bracket the optimal tradeoff $R^*(\mathsf{D})$. The new $R_{\rm LB}(\mathsf{D})$ is asymptotically tight when either $\mathsf{D} \to \infty$ or $R \to \infty$, and sheds further insights on how the (random) age of information could impact the performance of a cyber-physical control system.**

I. Problem Formulation

Consider a discrete-time Gaussian linear control system:

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + \mathbf{u}(t) + \mathbf{w}(t) \quad (1)$$

$$\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{v}(t), \quad (2)$$

where $\mathbf{x}(t)$ is the N-dimension system state column vector at time t, A is the state evolution matrix, $\mathbf{u}(t)$ is the control action to be decided by the controller at time t, $\mathbf{y}(t)$ is the M-dimensional observation column vector of the sensor, and C is the observation matrix. The disturbance $\mathbf{w}(t)$ and the measurement noise $\mathbf{v}(t)$ are zero-mean i.i.d. Gaussian vectors with covariance matrices $\Sigma_{\mathbf{w}}$ and $\Sigma_{\mathbf{v}}$, respectively.

The system starts with $\mathbf{x}(-1) = \mathbf{0}$ and $\mathbf{u}(-1) = \mathbf{0}$. In each time slot t a *variable-length* bit string $\mathbf{s}(t) \in \{0,1\}^*$ is transmitted by the sensor over a noise-free digital channel. $\mathbf{s}(t)$ will eventually reach the controller with zero error. That is, any corruption will be corrected either by error-correcting codes or by Automatic Repeat reQuest. The delay effects of ECC and ARQ are modeled by a non-negative integer random variable $\delta(t)$. That is, $\mathbf{s}(\tau)$ sent at time τ will be received at time $t = \tau + \delta(\tau)$. We assume $\delta(t)$ is i.i.d., has bounded support $[0, \delta_{\max}]$, and is independent of $\mathbf{w}(t)$ and $\mathbf{v}(t)$. The distribution of $\delta(t)$ is described by its pmf $\{p_\delta : \delta \in [0, \delta_{\max}]\}$.

This work was supported in part by NSF grants ECCS-1407603, CCF-1422997, and CCF-1618475.

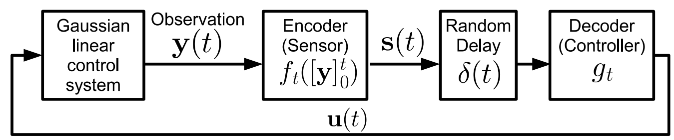


Figure 1: The system diagram

For each time t, the string $\mathbf{s}(t)$ transmitted by the sensor is a function of the current and past observations:

$$\mathbf{s}(t) = f_t(\{\mathbf{y}(\tau) : \tau \in [0,t]\}) \quad (3)$$

The length of the variable-length string $\mathbf{s}(t)$ is denoted by $|\mathbf{s}(t)|$, which is a random variable. We also define

$$\overline{\mathcal{T}}_{[t_1,t_2]} = \{\tau : \tau + \delta(\tau) \in [t_1, t_2]\} \quad (4)$$

as the set of indices τ for which $\mathbf{s}(\tau)$ arrives between $[t_1, t_2]$.

Assuming each $\mathbf{s}(\tau)$ is time stamped, the controller thus knows the set $\overline{\mathcal{T}}_{[0,t]}$ at time t. It then computes $\mathbf{u}(t)$ by

$$\mathbf{u}(t) = g_t(\overline{\mathcal{T}}_{[0,t]}, \{\mathbf{s}(\tau) : \tau \in \overline{\mathcal{T}}_{[0,t]}\}) \quad (5)$$

See Fig. 1 for the illustration of the system operation.

For any given encoder and controller pair $(\{f_t\}, \{g_t\})$, we define the *average bit length* of the scheme as

$$L(\{f_t\}, \{g_t\}) \triangleq \sup_{t \in [0,\infty)} E(|\mathbf{s}(t)|). \quad (6)$$

Given any constant $\mathsf{D} > 0$, the optimal rate-cost tradeoff is

$$R^*(\mathsf{D}) \triangleq \min_{f,g} L(\{f_t\}, \{g_t\}) \quad (7)$$

$$\text{subject to} \sup_{t \in [0,\infty)} E\left(\|\mathbf{x}(t)\|^2\right) \leq \mathsf{D}. \quad (8)$$

where (8) represents the mean square stability constraint.

While the characterization of $R^*(\mathsf{D})$ for random delay $\delta(t)$ remains an open problem, this work provides a pair of lower and upper bounds denoted by $R_{\rm LB}(\mathsf{D})$ and $R_{\rm UB}(\mathsf{D})$, respectively.

II. Motivation and Existing Results

The increasing importance of cyber physical systems (CPS) has been a driving force for new schemes in both control and information theory. For example, the modern electricity grid needs to accommodate the inherent system baseloads, diverse energy generating and consumption units, e.g., electric cars. Therefore, sensors, controller(s), generators, and smart consumption units are constantly communicating with each other with the goal of maximizing the overall system stability. One key question is how the underlying random communication delay could affect the rate-cost tradeoff $R^*(\mathsf{D})$. The answer would shed further insights on how the varying network conditions could impact the controllability of a CPS.

Rate-cost tradeoff $R^*(\mathsf{D})$ is a classic subject in control and information theory with dozens of important results over the years. One way of categorizing the existing results is by the types of system state disturbance and by the communication channel models. A summary for a very limited selection of existing works is provided in Table I. For example, when fixing the delay to zero $\delta(t) = 0$, our work assumes Gaussian disturbance with error-free digital communication channels, which very closely follows the settings of variable-length coding [3], [10], fixed-length coding [5], and entropy-based studies [13]. See the first row of Table I.

In the following, we highlight two major differences between this work and the existing results.

A. Contribution 1: A comprehensive treatment for finite D.

A recurring fundamental result in various settings [5], [6], [11], [12] is the *minimum asymptotic rate*. For any large but finite stability requirement D, we always have

$$R^*(\mathsf{D}) > \sum_{n=1}^{N} \max(\log_2(|\lambda_n(A)|), 0) \tag{9}$$

where $\{\lambda_n(A)\}$ are the eigenvalues of the evolution matrix A. This lower bound is loose for small D but is provably tight when $\mathsf{D} \to \infty$. Note that most existing works [5], [6], [9], [11], [12] first derived analytical lower bounds for any finite D and then let $\mathsf{D} \to \infty$. During the derivation, only the unstable directions of the system states are considered. The reason is that from the perspective of asymptotic stability, the directions with $|\lambda_n(A)| < 1$ are inherently stable so it does not contribute to the lower bound in (9). However, if we leave those stable directions "uncontrolled" they could still contribute to a finite but large mean-square norm. As a result, any high-performance scheme for *small* D must take into account both the stable and unstable directions. This is the approach taken in this work since we aim to estimate/bracket the optimal $R^*(\mathsf{D})$ for *all* D. This finite D emphasis is also studied in [3].

Also note that our rate-cost tradeoff definition considers the supremum in both (6) and (8). We call this setting *individual rate* and *individual stability* requirements, which is in contrast with the following requirements:

$$\text{[Time avg. rate:]}\ L \triangleq \limsup_{T} \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} E\left(|\mathbf{s}_k(t)|\right) \tag{10}$$

$$\text{[Time avg. stability:]}\ \limsup_{T} \frac{1}{T} \sum_{t=1}^{T} E\left(\|\mathbf{x}(t)\|^2\right) \leq \mathsf{D}. \tag{11}$$

Relatively speaking, the time averaged rate/stability requirement (e.g., the results in [3], [13]) is more relevant in the asymptotic regime while the individual rate/stability requirement is more stringent and more relevant in the finite regime, the main focus of this work.

B. Contribution 2: Random delay $\delta(t)$.

Most existing works assume zero delay, which may not fully capture the varying network dynamics like scheduling and congestion in modern CPSs.

One challenge of analyzing random delay $\delta(\tau)$ is that the packets arriving at the controller have different[1] *ages*. How to actively minimize the *age of information* has recently attracted much attention in both the information theory and networking communities [15]. In a broad sense this work studies how the randomly distributed age of information would alter the rate-cost tradeoff $R^*(\mathsf{D})$ for a Gaussian linear control system.

It is worth mentioning that although there is no dedicated communication channel from the controller back to the sensor, see Fig. 1, the controller, if desired, can feedback (some of) the delay information $\delta(t)$ back to the sensor by *embedding* it in the control action $\mathbf{u}(t)$. The sensor can then deduce it by observing $\mathbf{y}(t)$ and change the encoder/decoder $\{f_t, g_t\}$ functions to better adapt to the delay realization $\delta(t)$. This phenomenon of controlling the state through the "random information/delay disturbance" is absent when delay is fixed to zero or some fixed number. One main contribution of this work is to derive a new lower bound that holds even with this new delay-information feedback loop.

The most related works are [7], [8]. [8] studies the rate-cost tradeoff under a zero-delay erasure channel model. [7] studies the stability under a random delay model but does not impose any communication rate constraint (measurements are of infinite precision). The most defining feature of [7], [8] is that they analyze special schemes called *packetized predictive controllers* and *sequence-based controllers*, and thus essentially derive new achievable upper bounds $R_{\mathrm{UB}}(\mathsf{D})$. In contrast, this work focuses on deriving a fundamental lower bound $R_{\mathrm{LB}}(\mathsf{D})$ that holds for any possible designs.

Assuming zero disturbance, see row three of Table I, [2] studies the transmission rate when the sensor can actively modulate the timing of the transmission over a communication channel with unknown but bounded delay. Time- and event-triggering schemes are proposed and analyzed. The results can again be interpreted as deriving new upper bounds $R_{\mathrm{UB}}(\mathsf{D})$.

III. Preliminaries and Some Simple Derivations

We summarize the Kalman filter (KF) formulas that will be used extensively and provide some new but simple derivations.

A. The KF Formulas

Consider a zero-delay Gaussian linear control system (i.e., $P(\delta(t) = 0) = 1$) and its KF estimate $\hat{\mathbf{x}}(t)$. Define P_t as the covariance of the estimate error $\mathbf{x}(t) - \hat{\mathbf{x}}(t)$. Since the system starts from $\mathbf{x}(-1) = 0$, we have $\hat{\mathbf{x}}(-1) = \mathbf{0}$ and $P_{-1} = \mathbf{0}$. $\hat{\mathbf{x}}(t)$ is calculated iteratively for $t \in [0, \infty)$:

$$\Phi_t = A P_{t-1} A^{\mathrm{T}} + \Sigma_{\mathbf{w}} \tag{12}$$

$$\Gamma_t = \Phi_t C^{\mathrm{T}} (C \Phi_t C^{\mathrm{T}} + \Sigma_{\mathbf{v}})^{-1} \tag{13}$$

[1] Packet arrivals can also be out of order. I.e., an older string $\mathbf{s}(\tau_1)$ may arrive later than a newer string $\mathbf{s}(\tau_2)$ if $\tau_1 + \delta(\tau_1) > \tau_2 + \delta(\tau_2)$.

Table I: Summary of a non-comprehensive selection of existing works.

	Noiseless communication channel	Noisy communication channel
Stochastic or Gaussian disturbance	This work, [5], [10]. (i) [5], [10] consider only the unstable directions; (ii) [5] develops the asymptotic lower bound $R > \sum_{n=1}^{N} \max(\log(\vert\lambda_n\vert), 0)$ and a stabilizing coder-controller with time-sharing protocals; (iii) [10] studies the optimality of LTI controllers.	[3], [13]. (i) [3], [13] consider both stable and unstable directions; (ii) [13] introduces the novel sequential rate-distortion (SRD) framework. Full evaluation of the SRD results involves finding the minimizing distribution P; (iii) [3] develops new matrix-based finite-regime lower bounds.
Deterministic (i.e., unknown but bounded) disturbance	[6], [14]. (i) [6] considers only the unstable directions; (ii) [6] develops the asymptotic lower bound: $R > \sum_{n=1}^{N} \max(\log(\vert\lambda_n\vert), 0)$; (iii) [14] focuses exclusively on linear time-invariance controllers. A continuity assumption is needed, which may make it difficult to evaluate.	[9], [11]. (i) [9] considers only the unstable directions; (ii) introduces the novel concept of *anytime capacity*, which has led to many follow-up works; (iii) studies the asymptotic regime and may be difficult to evaluate.
Zero disturbance	[12] (i) considers only the unstable directions; (ii) develops the asymptotic lower bound $R > \sum_{n=1}^{N} \max(\log(\vert\lambda_n\vert), 0)$ [2] follows [12] but considers, additionally, bounded but unknown communication delay. Time-triggering and event-triggering schemes are proposed and analyzed.	[4], [11]. (i) [11] considers only the unstable directions; (ii) [11] develops the asymptotic lower bound: $R > \sum_{n=1}^{N} \max(\log(\vert\lambda_n\vert), 0)$; (iii) [4] studies the observability and controllability with feedback communication channels.

$$\hat{\mathbf{w}}_{t-1} = \Gamma_t \left(\mathbf{y}(t) - C(A\hat{\mathbf{x}}(t-1) + \mathbf{u}(t-1))\right) \quad (14)$$

$$\hat{\mathbf{x}}(t) = A\hat{\mathbf{x}}(t-1) + \mathbf{u}(t-1) + \hat{\mathbf{w}}_{t-1} \quad (15)$$

$$P_t = \Phi_t - \Gamma_t C \Phi_t \quad (16)$$

where $\hat{\mathbf{w}}_{t-1}$ is the *external update* when computing $\hat{\mathbf{x}}(t)$ from $\hat{\mathbf{x}}(t-1)$. Each $\hat{\mathbf{w}}_{t-1}$, $t \in [0, \infty)$, is independently distributed Gaussian with zero mean and covariance

$$\Sigma_{\hat{\mathbf{w}}_{t-1}} = \Phi_t C^{\mathrm{T}} (C\Phi_t C^{\mathrm{T}} + \Sigma_{\mathbf{v}})^{-1} C \Phi_t^{\mathrm{T}}. \quad (17)$$

We assume the convergence of P_t , Φ_t, Γ_t, and $\Sigma_{\hat{\mathbf{w}}_{t-1}}$ and denote their $t \to \infty$ limits by P, Φ, Γ, and $\Sigma_{\hat{\mathbf{w}}}$, respectively.

B. The distribution of the ages of information

Throughout this paper, we use the convention that $\sum_{\tau=a}^{b} f(\tau) = 0$ and $\prod_{\tau=a}^{b} f(\tau) = 1$ if $b < a$, regardless of the underlying function $f(\cdot)$, and the convention $\log\left(\frac{0}{0}\right) = 0$.

Recall that the delay pmf $\{p_\delta\}$ is defined from the sensor's perspective. Namely, the string $\mathbf{s}(\tau)$ will be received at time $t = \tau + \delta(\tau)$. We now consider *the delay experienced by the controller at time* t, denoted by $\theta(t)$. Namely, at time t, the *latest* string received by the controller is $\mathbf{s}(t - \theta(t))$. Given $\{p_\delta\}$, we compute the pmf of $\theta(t)$ as follows. Define $p_S \triangleq \sum_{\delta \in S} p_\delta$ as the probability that $\delta(t)$ belongs to a set S. Assume any arbitrary but fixed t value. The pmf of $\theta(t)$ is

$$\overline{p}_\theta = p_{[0,\theta]} \cdot \prod_{\tau=t-\theta+1}^{t} p_{(t-\tau, \delta_{\max}]}, \ \forall \theta \in [0, \delta_{\max}] \quad (18)$$

where the first term of (18) is the probability that string $\mathbf{s}(t-\theta)$ is received by time t, and the rest of (18) computes the probability that none of $\mathbf{s}(\tau)$, $\tau \in [t-\theta+1, t]$, has arrived by time t. Although the formula (18) involves t, the final value $\overline{p}_\theta$ does not since $\theta(t)$ is a stationary random process.

In general, the stationary random process $\theta(t)$ has memory. We thus analyze the joint pmf of $(\theta_1, \theta_2) = (\theta(t), \theta(t+m))$ for two time instants $t_1 = t$ and $t_2 = t + m$. The expression of the joint pmf, denoted by $\overline{p}_{\theta_1,\theta_2}(m)$, contains three cases.

Case 1: If $\theta_2 > \theta_1 + m$ then $\overline{p}_{\theta_1,\theta_2}(m) = 0$. It is because even if there is no new arrival during time $(t, t+m]$, the delay θ_2 experienced at time $t + m$ is at most $\theta_1 + m$.

Case 2: If $\theta_2 = \theta_1 + m$ then

$$\overline{p}_{\theta_1,\theta_2}(m) = p_{[0,\theta_1]} \cdot \prod_{\tau=t-\theta_1+1}^{t+m} p_{(t+m-\tau, \delta_{\max}]} \quad (19)$$

It is because if $\theta_2 = \theta_1 + m$, then it means that string $\mathbf{s}(t-\theta_1)$ is received by time t and none of the strings $\mathbf{s}(\tau)$, $\forall \tau \in [t - \theta_1 + 1, t + m]$, has arrived by time $t + m$.

Case 3: If $\theta_2 < \theta_1 + m$ then

$$\overline{p}_{\theta_1,\theta_2}(m) = p_{[0,\theta_1]} \cdot \left(\prod_{\tau=t-\theta_1+1}^{t+m-\theta_2-1} p_{(t-\tau,\delta_{\max}]}\right) \cdot p_{(\theta_2-m, \theta_2]} \cdot \left(\prod_{\tau=t+m-\theta_2+1}^{t+m} p_{(t+m-\tau,\delta_{\max}]}\right) \quad (20)$$

Case 3 computes the probability that string $\mathbf{s}(t-\theta_1)$ is received by time t; none of the strings $\mathbf{s}(\tau)$, $\forall \tau \in [t-\theta_1+1, t+m-\theta_2)$, has arrived by time t; string $\mathbf{s}(t+m-\theta_2)$ has arrived between time $(t, t+m]$; and none of the strings $\mathbf{s}(\tau)$, $\forall \tau \in (t+m-\theta_2, t+m]$, has arrived by time $t + m$.

IV. Main Results

A. A new lower bound

Proposition 1. *For any given integers $m \in [1, \infty)$ and $\theta_1, \theta_2 \in [0, \delta_{\max}]$, define an $N \times N$ matrix $F_{\theta_1,\theta_2}(m)$ by*

$$F_{\theta_1,\theta_2}(m) \triangleq \sum_{i=\theta_2+1}^{\theta_1+m} A^i \cdot \Sigma_{\hat{\mathbf{w}}} \cdot (A^i)^{\mathrm{T}} \quad (21)$$

$$= V \cdot \mathsf{diag}\{\sigma_n^{(\theta_1,\theta_2,m)}\} \cdot V^{\mathrm{T}} \quad (22)$$

where (22) is the singular value decomposition of $F_{\theta_1,\theta_2}(m)$ with V being a unitary matrix and $\sigma_n^{(\theta_1,\theta_2,m)}$ being the n-th singular value of $F_{\theta_1,\theta_2}(m)$ for $n \in [1, N]$. Note that if $\theta_2 \geq \theta_1 + m$, then we simply have $F_{\theta_1,\theta_2}(m) = \mathbf{0}$ and $\sigma_n^{(\theta_1,\theta_2,m)} = 0$ per the convention adopted in this work.

Also define scalars $\mathsf{D}_{\min}^{(\theta)}$ and $\overline{\mathsf{D}}_{\min}$ by

$$\mathsf{D}_{\min}^{(\theta)} \triangleq \mathrm{tr}\left(A^{\theta+1} P (A^{\theta+1})^{\mathrm{T}}\right) + \mathrm{tr}\left(\sum_{i=0}^{\theta} A^i \Sigma_{\mathbf{w}} (A^i)^{\mathrm{T}}\right) \quad (23)$$

$$\overline{\mathsf{D}}_{\min} \triangleq \sum_{\theta \in [0,\delta_{\max}]} \overline{p}_\theta \cdot \mathsf{D}_{\min}^{(\theta)}. \quad (24)$$

If $\mathsf{D} \leq \overline{\mathsf{D}}_{\min}$*, then no scheme can achieve* $\sup_t E\left(\|\mathbf{x}(t)\|^2\right) \leq \mathsf{D}$*. If* $\mathsf{D} > \overline{\mathsf{D}}_{\min}$*, then for any scheme that satisfies* $\sup_t E\left(\|\mathbf{x}(t)\|^2\right) \leq \mathsf{D}$*, its expected bit length* L *must satisfy:*

$$L - \left(L \cdot \log_2\left(\frac{L}{L+1}\right) + \log_2\left(\frac{1}{L+1}\right)\right) \geq$$
$$\frac{1}{m}\sum_{\forall\theta_1,\theta_2} \overline{p}_{\theta_1,\theta_2}(m) \cdot \left(\sum_{n=1}^{N} \frac{1}{2}\log_2\left(\frac{\sigma_n^{(\theta_1,\theta_2,m)}}{\mathsf{D}_n^{(\theta_1,\theta_2,m)}}\right)\right) \qquad (25)$$

where the value of $\mathsf{D}_n^{(\theta_1,\theta_2,m)}$ *is the* water-filling *coefficients:*

$$\mathsf{D}_n^{(\theta_1,\theta_2,m)} = \min(\eta^{(m)}, \sigma_n^{(\theta_1,\theta_2,m)}),\ \forall n, \theta_1, \theta_2 \qquad (26)$$

and $\eta^{(m)}$ *is the largest possible value that still satisfies*

$$\sum_{\forall\theta_1,\theta_2} \overline{p}_{\theta_1,\theta_2}(m) \cdot \left(\sum_{n=1}^{N} \mathsf{D}_n^{(\theta_1,\theta_2,m)}\right) \leq \mathsf{D} - \overline{\mathsf{D}}_{\min}. \qquad (27)$$

We note that Proposition 1 makes no separate consideration of whether $|\lambda_n(A)| < 1$ or not. It thus simultaneously considers both stable and unstable directions of the system. Also since Proposition 1 holds for all m, we can further improve the bound by taking the supremum over all $m \in [1, \infty)$.

A detailed proof is provided separately in [1]. Technically, for each fixed m, a *single-shot* approach is used, which focuses exclusively on two fixed time instants t and $t+m$, rather than the entire time axis. Together with the individual stability condition in (8), this approach allows new careful analysis of the probabilistic impact of random delay at these two time instants and circumvents the difficulty, pointed out in [3, Remark 2], of applying the water-filling principle under the time average stability condition in (11).

Intuitively, Proposition 1 contains three components. Component 1 is the minimum achievable stability $\overline{\mathsf{D}}_{\min}$ in (24). Specifically (23) describes the minimum achievable $\mathsf{D}_{\min}^{(\theta)}$ when the controller is facing a deterministic delay θ, which depends on both P and $\Sigma_{\mathbf{w}}$ and grows exponentially with respect to θ. $\mathsf{D}_{\min}^{(\theta)}$ is then averaged according to $\{\overline{p}_\theta\}$.

Component 2 is the $F_{\theta_1,\theta_2}(m)$ in (21), which is based on the covariance $\Sigma_{\hat{\mathbf{w}}}$ instead of $\Sigma_{\mathbf{w}}$. Specifically, each update $\hat{\mathbf{w}}(t)$, see (15), represents the "additional information" that a new KF estimate $\hat{\mathbf{x}}(t+1)$ has over the old $\hat{\mathbf{x}}(t)$ due to the controller receiving exactly 1 additional observation $\mathbf{y}(t)$. However, suppose $\theta(t) = \theta_1$ and $\theta(t+m) = \theta_2$. Then there are roughly $(\theta_1 + m - \theta_2)$ number of new observations $\mathbf{s}(\tau)$ (and thus $\mathbf{y}(\tau)$) arriving during time $(t, t+m]$. The summation of $(\theta_1 + m - \theta_2)$ terms in (21) then represents the maximum amount of additional information that can be carried during $(t, t+m]$ if no quantization/compression is ever performed.

Component 3 then invokes the classic rate-distortion results to characterize the stability impact when quantizing the additional information $F_{\theta_1,\theta_2}(m)$ in Component 2. As a result, the average in the right-hands side of (25) is taken over the joint pmf $\overline{p}_{\theta_1,\theta_2}(m)$. Since the rate-distortion results are entropy-based, the left-hand side of (25) quantifies the largest possible entropy $H(\mathbf{s}(t))$ given $E(|\mathbf{s}(t)|) \leq L$, which further relates the expected length L to the entropy $H(\mathbf{s}(t))$.

Proposition 2. *Continue from Proposition 1. If the* A *matrix is non-defective, then we also have*

$$\lim_{m\to\infty} \frac{1}{m}\sum_{\forall\theta_1,\theta_2} \overline{p}_{\theta_1,\theta_2}(m) \cdot \left(\sum_{n=1}^{N} \frac{1}{2}\log_2\left(\frac{\sigma_n^{(\theta_1,\theta_2,m)}}{\mathsf{D}_n^{(\theta_1,\theta_2,m)}}\right)\right)$$
$$= \sum_{n=1}^{N} \max(\log_2(|\lambda_n(A)|), 0)$$

Proposition 2 shows that when using (25) as an *entropy* lower bound rather than an *expected-length* lower bound, then the right-hand side of (25) coincides with the asymptotic entropy lower bound (9) and is thus tight when $\mathsf{D} \to \infty$.

The main step of proving Proposition 2 is to note that since $\Sigma_{\hat{\mathbf{w}}}$ is positive definite, we can have $\beta_{\min} I \preceq \Sigma_{\hat{\mathbf{w}}} \preceq \beta_{\max} I$ for some $\beta_{\min}$ and $\beta_{\max}$, where I is the identity matrix. Then we replace $\Sigma_{\hat{\mathbf{w}}}$ in (21) by $\beta_{\min} I$ (or $\beta_{\max} I$) that helps crystalize the relationship between $\sigma_n^{(\theta_1,\theta_2,m)}$ and $\lambda_n(A)$.

The detailed proofs of Propositions 1 and 2 are omitted due to the space constraint.

B. A simple upper bound

The main focus of this work is on deriving the new lower bound. For upper bounds, we can adapt a straightforward KF-based quantizer in a similar way as in [8]. Specifically, first let the sensor computes the KF estimate $\hat{\mathbf{x}}(t)$. Then consider a (potentially rotated) rectangle lattice where the corresponding Voronoi regions are hyperrectangles. The $\hat{\mathbf{x}}(t)$ is then represented by the index of the corresponding Voronoi region and the index is later *entropy coded* to generate the variable length string $\mathbf{s}(t)$.

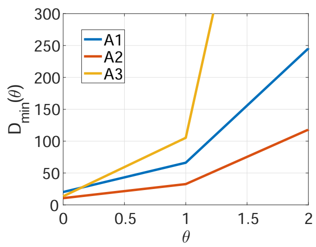
At the controller side, in time t the latest received quantized estimate, denoted by $[\hat{\mathbf{x}}(t-\theta(t))]_q$, is used to generate $\mathbf{u}(t) = -A^{\theta(t)+1}[\hat{\mathbf{x}}(t-\theta(t))]_q$, which aims to bring the state $\mathbf{x}(t+1)$ back to $\mathbf{0}$ according to the latest quantized estimate $[\hat{\mathbf{x}}(t-\theta(t))]_q$. This type of predictive controllers is relatively straightforward and the detailed description is omitted.

We use Monte-Carlo simulation to compute its numerical performance, which then serves as an upper bound $R_{\mathrm{UB}}(\mathsf{D})$.

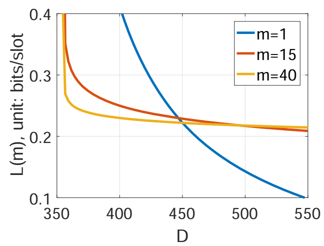
C. Discussion and further extension

Proposition 1 suggests that the impact of random ages $\theta(t)$ are two fold: The marginal distribution $\overline{p}_\theta$ directly impacts the minimum achievable stability $\overline{\mathsf{D}}_{\min}$, see (24). The joint distribution $\overline{p}_{\theta_1,\theta_2}(m)$ then decides how much information can be effectively compressed and transmitted from the sensor to the controller, see (21) and (25).

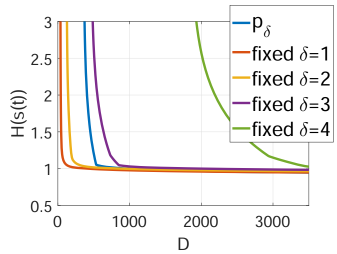
These main ideas can be readily generalized to the setting of K sensors: Each sensor k transmits its own string $\mathbf{s}_k(t)$ and delay $\delta_k(t)$ is incurred during the transmission. The controller decides $\mathbf{u}(t)$ based on all the $\mathbf{s}_k(\tau)$ it has received. Now the sensor-side delays form a K-dimensional vector $\vec{\delta}(t)$. We use its pmf $p_{\vec{\delta}}$ to compute the distribution of the K-dimensional delay vector experienced by the controller, denoted by $\overline{p}_{\vec{\theta}}$ and $\overline{p}_{\vec{\theta}_1,\vec{\theta}_2}(m)$. Finally, we apply the same three components of Proposition 1 to lower bound the traffic sum rate from the K sensors to the controller.



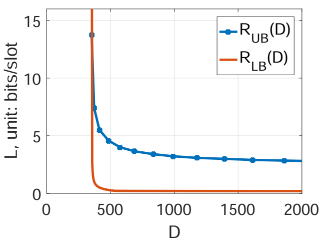
(a) $\mathsf{D}_{\min}^{(\theta)}$ versus θ



(b) Lower bounds w. different m



(c) Asymptote predicted in (9)



(d) $R_{\rm UB}(\mathsf{D})$ vs $R_{\rm LB}(\mathsf{D})$

Figure 2: Numerical evaluation.

V. A Numerical Example

We assume $\Sigma_{\mathbf{w}} = I_{3\times 3}$, $\Sigma_{\mathbf{v}} = I_{2\times 2}$, and

$$C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \tag{28}$$

Consider three different state evolution matrices A_1 to A_3:

$$A_i = \begin{bmatrix} 2 & \alpha_i & 0 \\ 0 & 1 & \alpha_i \\ 0 & 0 & 0.9 \end{bmatrix}, \ \forall i \in \{1,2,3\} \tag{29}$$

with $\alpha_1 = 0$, $\alpha_2 = 0.5$, and $\alpha_3 = 3$. Namely, system matrix A_3 has the highest coupling between the three coordinates while system matrix A_1 is diagonal and has no coupling. Fig. 2(a) plots $\mathsf{D}_{\min}^{(\theta)}$ versus θ. As one may expect, the highly coupled A_3 has the highest exponential growth rate. Surprisingly, the growth rate of the *decoupled* A_1 is higher than the slightly coupled A_2. Furthermore, when $\theta = 0$ the $\mathsf{D}_{\min}^{(\theta)}$ of A_3 is actually slightly smaller than that of A_1. Since A_1 to A_3 have the same set of eigenvalues $\lambda_n(A_i)$, their asymptotic minimum rates (9) are identical. This shows that the finite length performance depends heavily on the interplay among A, C, $\Sigma_{\mathbf{w}}$, $\Sigma_{\mathbf{v}}$ and delay distribution $\{p_\delta\}$.

In the following, we focus exclusively on A_2 and assume $p_\delta = 0.25$ for all $\delta \in [1,4]$. By (18), we have $\overline{p}_\theta$ being $1/4$, $3/8$, $9/32$, and $3/32$ for θ being 1, 2, 3, and 4, respectively. We then compute $\overline{\mathsf{D}}_{\min} = 354.12$ using (24). Fig. 2(b) plots $R_{\rm LB}(\mathsf{D})$ in Proposition 1 for the cases of $m = 1$, 15, and 40, respectively. One can see that the lower bounds of $m = 1$, 15, and 40 dominate the scenarios of small, medium, and large D values, respectively. The "envelope" (or equivalently the supremum) over all $m \in [1,\infty)$ then gives the strongest lower bound. In the remaining discussion, we take the supremum over $m \in [1, 40]$ when evaluating the lower bound.

Fig. 2(c) plots the entropy part of the lower bound (the right-hand side of (25)). For comparison, we also plot the cases of fixed deterministic delay with $\delta = 1$ to 4, respectively. In all five cases (one random and four deterministic delays), the asymptotes converge to the entropy being $H(\mathbf{s}(t)) = 1$, as predicted by (9) since the eigenvalues of A_2 are 2, 1, and 0.9.

Fig. 2(d) plots the upper bound $R_{\rm UB}(\mathsf{D})$ vs the lower bound $R_{\rm LB}(\mathsf{D})$. One can see that the vertical asymptote of our lower bound (i.e., $\overline{\mathsf{D}}_{\min} = 354.12$) is tight and matches the numerically computed upper bound $R_{\rm UB}(\mathsf{D})$.

VI. Conclusion

This work derives a lower bound $R_{\rm LB}(\mathsf{D})$ for the optimal rate-cost tradeoff with i.i.d. random delay, and uses a simple scheme to numerically compute an upper bound $R_{\rm UB}(\mathsf{D})$. The results complement the existing *age of information* problems by quantitatively bracketing the impact of (random) age of information to the optimal $R^*(\mathsf{D})$. The new $R_{\rm LB}(\mathsf{D})$ is asymptotically tight when either $\mathsf{D} \to \infty$ or $R \to \infty$.

References

[1] https://engineering.purdue.edu/~chihw/pub_pdf/ISIT2018_proof.pdf.

[2] M. Khojasteh, P. Tallapragada, J. Cortés, and M. Franceschetti, "Time-triggering versus event-triggering control over communication channels," in *Proc. 56th IEEE Conf. Decision and Contr.* Melbourne, Australia, December 2017.

[3] V. Kostina and B. Hassibi, "Rate cost tradeoffs in control," 2016, part I and Part II appeared in arXiv:1612.02126 and arXiv:1612.02128, respectively.

[4] A. Matveev and A. Savkin, "An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM J. Contr. Optim.*, vol. 46, no. 4, pp. 1323–1367, 2007.

[5] G. Nair and R. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Contr. Optim.*, vol. 43, no. 2, pp. 413–436, 2004.

[6] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback control under data rate constraints: An overview," *Proc. IEEE*, vol. 95, no. 1, pp. 108–137, January 2007.

[7] D. Quevedo and I. Jurado, "Stability of sequence-based control with random delays and dropouts," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1296–1302, May 2014.

[8] D. Quevedo, J. Østergaard, and D. Nešić, "Packetized predictive control of stochastic systems over bit-rate limited channels with packet loss," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2854–2868, December 2011.

[9] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link — part i: Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, August 2006.

[10] E. Silva, M. Derpich, J. Østergaard, and M. A. Encina, "A characterization of the minimal average data rate that guarantees a given closed-loop performance level," *IEEE Trans. Autom. Control*, vol. 61, no. 8, pp. 2171–2186, August 2016.

[11] S. Tatikonda and S. Mitter, "Control over noisy channels," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1196–1201, July 2004.

[12] ——, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, no. 7, pp. 1056–1068, July 2004.

[13] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, September 2004.

[14] K. Tsumura and J. Maciejowski, "Stability of SISO control systems under constraints of channel capacity," in *Proc. 42nd IEEE Conf. Decision and Contr.* Maui, USA, December 2003.

[15] R. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *Proc. IEEE Int'l Symp. Inform. Theory.* Aachen, Germany, June 2017.