# Sampling for Data Freshness Optimization: Non-linear Age Functions

Yin Sun and Benjamin Cyr

*Abstract:* **In this paper, we study how to take samples at a data source for improving the freshness of received data samples at a remote receiver. We use non-linear functions of the age of information to measure data freshness, and provide a survey of non-linear age functions and their applications. The sampler design problem is studied to optimize these data freshness metrics, even when there is a sampling rate constraint. This sampling problem is formulated as a constrained Markov decision process (MDP) with a possibly uncountable state space. We present a complete characterization of the optimal solution to this MDP: The optimal sampling policy is a deterministic or randomized threshold policy, where the threshold and the randomization probabilities are characterized based on the optimal objective value of the MDP and the sampling rate constraint. The optimal sampling policy can be computed by bisection search, and the curse of dimensionality is circumvented. These age optimality results hold for (i) general data freshness metrics represented by monotonic functions of the age of information, (ii) general service time distributions of the queueing server, (iii) both continuous-time and discrete-time sampling problems, and (iv) sampling problems both with and without the sampling rate constraint. Numerical results suggest that the optimal sampling policies can be much better than zero-wait sampling and the classic uniform sampling.**

*Index Terms:* **Age of information, data freshness, Markov decision process, sampling.**

## I. INTRODUCTION

INFORMATION usually has the greatest value when it is fresh [2, p. 56]. For example, real-time knowledge about the location, orientation, and speed of motor vehicles is imperative in autonomous driving, and the access to timely updates about the stock price and interest-rate movements is essential for developing trading strategies on the stock market. In [3], [4], the concept of *Age of Information* was introduced to measure the freshness of information that a receiver knows about the status of the remote source. Consider a sequence of source samples that are sent through a queue to a receiver. Let $U_t$ be the generation time of the newest sample that has been delivered to the receiver by time $t$. The age of information, as a function of $t$, is defined as

$$\Delta_t = t - U_t, \tag{1}$$

which is the time elapsed since the newest sample was generated. Hence, a small age $\Delta_t$ indicates that there exists a recently generated sample at the receiver.

In practice, some information sources (e.g., vehicle location, stock price) vary quickly over time, while others (e.g., temperature, interest-rate) change slowly. Consider again the example of autonomous driving: The location information of motor vehicles collected 0.5 seconds ago could already be quite stale for making control decisions[1], but the engine temperature measured a few minutes ago is still valid for engine health monitoring. From this example, one can observe that data freshness should be evaluated based on *(i)* the time-varying pattern of the source and *(ii)* how valuable the fresh data is in the specific application. However, the age $\Delta_t$ defined in (1) is the time difference between data generation at the transmitter and data usage at the receiver, which cannot fully describe the source pattern and application context.[2] This motivated us to seek more appropriate data freshness metrics that can interpret the role of freshness in real-time applications.

In this paper, we consider using a non-linear function $u(\Delta_t)$ of the age $\Delta_t$ as a data freshness metric, where $u(\Delta_t)$ could be the utility value of data with age $\Delta_t$, temporal autocorrelation function of the source, estimation error of signal value, or other application-specific performance metrics [1], [5]–[19]. A survey of non-linear age functions and their applications is provided in Subsection III.B. Recently, the age of information has received significant attention, because of the rapid deployment of real-time applications. A large portion of existing studies on age have been devoted to linear functions of the age $\Delta_t$, e.g., [4], [20]–[46]. However, the design of efficient data update policies for optimizing non-linear age metrics remains largely unexplored. To that end, we investigate a problem of sampling an information source, where the samples are forwarded to a remote receiver through a channel that is modeled as a FIFO queue. The optimal sampler design for optimizing non-linear age metrics is obtained. The contributions of this paper are summarized as follows:

- We consider a class of data freshness metrics, where the utility for data freshness is represented by a *non-increasing* function $u(\Delta_t)$ of the age $\Delta_t$. Accordingly, the penalty for data

Manuscript received December 15, 2018; approved for publication by Clement Kam, Guest Editor.

This paper was presented in part at IEEE SPAWC 2018 [1].

This work was supported in part by NSF grant CCF-1813050 and ONR grant N00014-17-1-2417.

Y. Sun is with the Department of ECE, Auburn University, Auburn, AL 36849 USA, email: yzs0078@auburn.edu.

B. Cyr is with the Department of EECS, University of Michigan, Ann Arbor, MI 48109 USA, email: bac0027@tigermail.auburn.edu.

B. Cyr participated in this study when he was an undergraduate student at Auburn University.

Y. Sun is the corresponding author.

Digital Object Identifier: 10.1109/JCN.2019.000035

[1] A car will travel 15 meters during 0.5 seconds at the speed of 70 mph.

[2] To the best of our knowledge, the issue that "the actual age $\Delta_t$ is not a good representation of freshness" was firstly pointed out by Anthony Ephremides in one presentation at the Information Theory and Application (ITA) Workshop in 2015.

1229-2370/19/$10.00 © 2019 KICS

staleness is denoted by a *non-decreasing* function $p(\Delta_t)$ of $\Delta_t$. The sampler design problem for optimizing these data freshness metrics, possibly with a sampling rate constraint, is considered. This sampling problem is formulated as a constrained Markov decision process (MDP) with a possibly uncountable state space.

- We prove that an optimal sampling solution to this MDP is a deterministic or randomized threshold policy, where the threshold is equal to the optimum objective value of the MDP plus the optimal Lagrangian dual variable associated with the sampling rate constraint; see Subsection V.E for the details. The threshold can be computed by bisection search, and the randomization probabilities are chosen to satisfy the sampling rate constraint. The curse of dimensionality is circumvented in this sampling solution by exploiting the structure of the MDP. These age optimality results hold for *(i)* general monotonic age metrics, *(ii)* general service time distributions of the queueing server, *(iii)* both continuous-time and discrete-time sampling problems, and *(iv)* sampling problems both with and without the sampling rate constraint. Among the technical tools used to prove these results are an extension of Dinkelbach's method for MDP and a geometric multiplier technique for establishing strong duality. These technical tools were recently used in [47], [48], where a quite different sampling problem was solved. In addition, we will also introduce some proof ideas that are specific to the sampling problem that we consider in this paper, which will be used to prove Lemma 5, Theorem 5, and Lemma 7 in Section V.
- When there is no sampling rate constraint, a logical sampling policy is the zero-wait sampling policy [4], [15], [24], which is throughput-optimal and delay-optimal, but not necessarily age-optimal. We develop sufficient and necessary conditions for characterizing the optimality of the zero-wait sampling policy for general monotonic age metrics. Our numerical results show that the optimal sampling policies can be much better than zero-wait sampling and the classic uniform sampling.

The rest of this paper is organized as follows. In Section II, we discuss some related work. In Section III, we describe the system model and the formulation of the optimal sampling problem; a short survey of non-linear age functions is also provided. In Section IV, we present the optimal sampling policy for different system settings, as well as a sufficient and necessary condition for the optimality of the zero-wait sampling policy. The proofs are provided In Section V. The numerical results and the conclusion are presented in Section VI and Section VII.

## II. RELATED WORK

The age of information was used as a data freshness metric as early as 1990s in the studies of real-time databases [3], [49]–[51]. Queueing theoretic techniques were introduced to evaluate the age of information in [4]. The average age, average peak age, and age distribution have been analyzed for various queueing systems in, e.g., [4], [16], [20]–[22], [52]–[55]. It was observed that a last-come, first-served (LCFS) scheduling policy can achieve a smaller time-average age than a few other scheduling policies. The optimality of the LCFS policy, or more generally the last-generated, first-served (LGFS) policy, was first proven in [56]. This age optimality result holds for several queueing systems with multiple servers, multiple hops, and/or multiple sources [56]–[60].

When the transmission power of the source is subject to an energy-harvesting constraint, the age of information was minimized in, e.g., [15], [23]–[30]. Source coding and channel coding schemes for reducing the age were developed in, e.g., [31]–[34]. Age-optimal transmission scheduling of wireless networks have been investigated in, e.g., [35]–[42], [61], [62]. Game theoretical perspective of the age was studied in [43], [44], [63], [64]. The aging effect of channel state information was analyzed in, e.g., [65]–[67]. An interesting connection between the age of information and remote estimation error was revealed in [17], [47], [48], where the optimal sampling policies were obtained for two continuous-time Markov processes. The impact of the age to control systems was studied in [18], [19], [45], [68]. Emulations and measurements of the age were conducted in [46], [69], [70]. An age-based transport protocol was developed in [71].

In [33], [42], optimal sampling policies were developed to minimize the time-average age for status updates over wireless channels, where the optimal sampling policies were shown to be randomized threshold policies. Structural properties of the randomized threshold policies were obtained in [33], [42] to simplify the value iteration or policy iteration algorithms therein. The linear age function considered in [33], [42] is a special case of the monotonic age functions considered in this paper, and the channel models in [33], [42] are different from ours. In our study, the optimal sampling policies are characterized semi-analytically and can be computed by bisection search. In a special case of [33], a closed-form optimal sampling solution was obtained. However, it is unclear whether (semi-)analytical or closed-form solutions can be found for the general cases considered in [33], [42].

The most relevant prior study to this paper is [15]. This paper generalizes [15] in the following aspects: *(i)* The data freshness metrics considered in this paper are more general than those of [15]. The age penalty function $p(\Delta_t)$ in [15] is assumed to be non-negative and non-decreasing, which is relaxed in this paper to be an arbitrary non-decreasing function that is more desirable for some applications. *(ii)* The optimal sampling policies developed in this paper are simpler and more insightful than those in [15]. A two-layered nested bisection search algorithm was developed to compute the optimal threshold [15]. In this paper, the optimal threshold can be computed by a single layer of bisection search. *(iii)* In [15], the optimal sampling strategy was obtained for continuous-time systems. In this paper, we also develop an optimal sampling strategy for discrete-time systems, without sacrificing from any approximation error or sub-optimality. *(iv)* It was assumed in [15] that after the previous sample was delivered, the next sample must be generated within a given amount of time. By adopting more insightful proof techniques, we are able to remove such an assumption and greatly simplify the proofs in this paper.

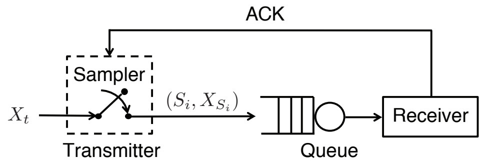



Fig. 1. System model.

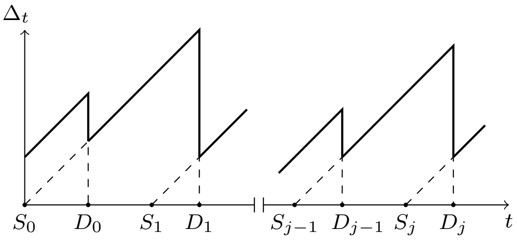



Fig. 2. Evolution of the age $\Delta_t$ over time.

## III. MODEL, METRICS, AND FORMULATION

### A. System Model

We consider the status update system illustrated in Fig. 1, where samples of a source process $X_t$ are taken and sent to a receiver through a communication channel. The channel is modeled as a single-server FIFO queue with independent and identically distributed (i.i.d.) service times. The system starts to operate at time $t = 0$. The $i$th sample is generated at time $S_i$ and is delivered to the receiver at time $D_i$ with a service time $Y_i$, which satisfy $S_i \le S_{i+1}$, $S_i + Y_i \le D_i$, $D_i + Y_{i+1} \le D_{i+1}$, and $0 < \mathrm{E}[Y_i] < \infty$ for all $i$. Each sample packet $(S_i, X_{S_i})$ contains the sampling time $S_i$ and the sample value $X_{S_i}$. Once a sample is delivered, the receiver sends an acknowledgement (ACK) back to the sampler with zero delay. Hence, the sampler has access to the idle/busy state of the server in real-time.

Let $U_t = \max\{S_i : D_i \le t\}$ be the generation time of the freshest sample that has been delivered to the receiver by time $t$. Then, the *age of information*, or simply *age*, at time $t$ is defined by [3], [4]

$$\Delta_t = t - U_t = t - \max\{S_i : D_i \le t\}, \tag{2}$$

which is plotted in Fig. 2. Because $D_i \le D_{i+1}$, $\Delta_t$ can be also written as

$$\Delta_t = t - S_i, \text{ if } D_i \le t < D_{i+1}. \tag{3}$$

The initial state of the system is assumed to be $S_0 = 0$, $D_0 = Y_0$, and $\Delta_0$ is a finite constant.

In this paper, we will consider both continuous-time and discrete-time status-update systems. In the continuous-time setting, $t \in [0, \infty)$ can take any positive value. In the discrete-time setting, $t \in \{0, T_s, 2T_s, \cdots\}$ is a multiple of period $T_s$; as a result, $S_i, D_i, Y_i, t, U_t, \Delta_t$ are all discrete-time variables. For notational simplicity, we choose $T_s = 1$ second such that all the discrete-time variables are integers. The results for other values of $T_s$ can be readily obtained by time scaling.

In practice, the continuous-time setting can be used to model status-update systems with a high clock rate, while the discrete-time setting is appropriate for characterizing sensors that have a very low energy budget and can only wake up periodically from a low-power sleep mode.

### B. Data Staleness and Freshness Metrics: A Survey

The dissatisfaction for data staleness (or the eagerness for data refreshing) is represented by a penalty function $p(\Delta)$ of the age $\Delta$, where the function $p : [0, \infty) \mapsto \mathbb{R}$ is *non-decreasing*. This non-decreasing requirement on $p(\Delta)$ complies with the observations that stale data is usually less desired than fresh data [2], [5]–[10]. This data staleness model is quite general, as it allows $p(\Delta)$ to be non-convex or dis-continuous. These data staleness metrics are clearly more general than those in [14], [15], where $p(\Delta)$ was restricted to be *non-negative* and *non-decreasing*.

Similarly, data freshness can be characterized by a *non-increasing* utility function $u(\Delta)$ of the age $\Delta$ [6], [8]. One simple choice is $u(\Delta) = -p(\Delta)$. Note that because the age $\Delta_t$ is a function of time $t$, $p(\Delta_t)$ and $u(\Delta_t)$ are both time-varying, as illustrated in Fig. 3. In practice, one can choose $p(\cdot)$ and $u(\cdot)$ based on the information source and the application under consideration, as illustrated in the following examples.[3]

#### B.1 Auto-correlation Function of the Source

The auto-correlation function $\mathrm{E}[X_t^* X_{t-\Delta_t}]$ can be used to evaluate the freshness of the sample $X_{t-\Delta_t}$ [16]. For some stationary sources, $|\mathrm{E}[X_t^* X_{t-\Delta_t}]|$ is a non-negative, non-increasing function of the age $\Delta_t$, which can be considered as an age utility function $u(\Delta_t)$. For example, in stationary ergodic Gauss-Markov block fading channels, the impact of channel aging can be characterized by the auto-correlation function of fading channel coefficients. When the age $\Delta_t$ is small, the auto-correlation function and the data rate both decay with respect to the age $\Delta_t$ [65].

#### B.2 Estimation Error of Real-time Source Value

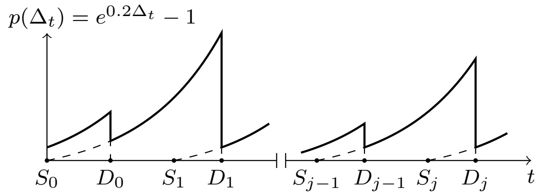
Consider a status-update system, where samples of a Markov source $X_t$ are forwarded to a remote estimator. The estimator uses causally received samples to reconstruct an estimate $\hat{X}_t$ of real-time source value. If the sampling times $S_i$ are independent of the observed source $\{X_t, t \ge 0\}$, the mean-squared estimation error at time $t$ can be expressed as an age penalty function $p(\Delta_t)$ [17], [22], [47], [48]. If the sampling times $S_i$ are chosen based on causal knowledge about the source, the estimation error is not a function of $\Delta_t$ [17], [47], [48].

The above result can be generalized to the state estimation error of feedback control systems [18], [19]. Consider a single-loop feedback control system, where a plant and a controller are governed by a linear time-invariant (LTI) system, i.e.,

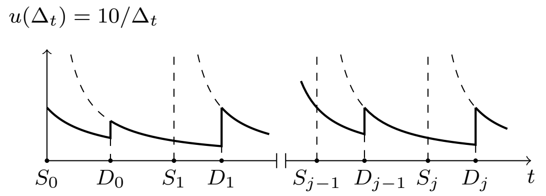
$$X_{t+1} = AX_t + BU_t + N_t, \tag{4}$$

where $X_t \in R^n$ is the state of the system at time slot $t$, $n$ is the system dimension, $U_t \in R^m$ represents the control input, and

[3]In some of these examples, the age utility function $u(\Delta_t)$ is non-negative and non-increasing. The corresponding age penalty function $p(\Delta_t) = -u(\Delta_t)$ is non-positive and non-decreasing. Hence, it is desirable to allow the age penalty function $p(\Delta_t)$ to be negative.




(a) Non-decreasing age penalty function $p(\Delta_t) = e^{0.2\Delta_t} - 1$.




(b) Non-increasing age utility function $u(\Delta_t) = 10/\Delta_t$.

Fig. 3. Two examples of non-linear age functions.

$N_t \in R^n$ is the exogenous noise vector having i.i.d. Gaussian distributed elements with zero mean and covariance $\Sigma$. The constant matrices $A \in R^{n\times n}$ and $B \in R^{n\times m}$ are the system and input matrices, respectively, where $(A, B)$ is assumed to be controllable. Samples of the state process $X_t$ are forwarded to the controller, which determines $U_t$ at time $t$ based on the samples that have been delivered by time $t$. Under some assumptions, the state estimation error can be proven to be independent of the adopted control policy [72, Lemma 5.2.1], [45]. Furthermore, if the sampling times $S_i$ are independent of the state process $X_t$, then the state estimation error is an age penalty function $p(\Delta_t)$ that is determined by the system matrix $A$ and the covariance $\Sigma$ of the exogenous noise [18], [19].

### B.3 Information based Data Freshness Metric

Let

$$\boldsymbol{W}_t = \{(X_{S_i}, S_i) : D_i \le t\} \tag{5}$$

denote the samples that have been delivered to the receiver by time $t$. One can use the mutual information $I(X_t; \boldsymbol{W}_t)$ — the amount of information that the received samples $\boldsymbol{W}_t$ carry about the current source value $X_t$ — to evaluate the freshness of $\boldsymbol{W}_t$. If $I(X_t; \boldsymbol{W}_t)$ is close to $H(X_t)$, the samples $\boldsymbol{W}_t$ contains a lot of information about $X_t$ and is considered to be fresh; if $I(X_t; \boldsymbol{W}_t)$ is almost 0, $\boldsymbol{W}_t$ provides little information about $X_t$ and is deemed to be obsolete.

One way to interpret $I(X_t; \boldsymbol{W}_t)$ is to consider how helpful the received samples $\boldsymbol{W}_t$ are for inferring $X_t$. By using the Shannon code lengths [73, Section 5.4], the expected minimum number of bits $L$ required to specify $X_t$ satisfies

$$H(X_t) \le L < H(X_t) + 1, \tag{6}$$

where $L$ can be interpreted as the expected minimum number of binary tests that are needed to infer $X_t$. On the other hand, with the knowledge of $\boldsymbol{W}_t$, the expected minimum number of bits $L'$ that are required to specify $X_t$ satisfies

$$H(X_t|\boldsymbol{W}_t) \le L' < H(X_t|\boldsymbol{W}_t) + 1. \tag{7}$$

If $X_t$ is a random vector consisting of a large number of symbols (e.g., $X_t$ represents an image containing many pixels or the coefficients of MIMO-OFDM channels), the one bit of overhead in (6) and (7) is insignificant. Hence, $I(X_t; \boldsymbol{W}_t)$ is approximately the reduction in the description cost for inferring $X_t$ without and with the knowledge of $\boldsymbol{W}_t$.

If $X_t$ is a stationary Markov chain, by data processing inequality [73, Theorem 2.8.1], it is easy to prove the following lemma:

**Lemma 1:** If $X_t$ is a stationary (continuous-time or discrete-time) Markov chain, $\boldsymbol{W}_t$ is defined in (5), and the sampling times $S_i$ are independent of $\{X_t, t \ge 0\}$, then the mutual information

$$I(X_t; \boldsymbol{W}_t) = I(X_t; X_{t-\Delta_t}) \tag{8}$$

is a non-negative and non-increasing function $u(\Delta_t)$ of $\Delta_t$.

*Proof:* See Appendix A. □

Lemma 1 provides an intuitive interpretation of "information aging": The amount of information $I(X_t; \boldsymbol{W}_t)$ that is preserved in $\boldsymbol{W}_t$ for inferring the current source value $X_t$ decreases as the age $\Delta_t$ grows. We note that Lemma 1 can be generalized to the case that $X_t$ is a stationary discrete-time Markov chain with memory $k$. In this case, each sample $\boldsymbol{V}_t = (X_t, X_{t-1}, \cdots, X_{t-k+1})$ should contain the source values at $k$ successive time instants. Let $\boldsymbol{W}_t = \{(\boldsymbol{V}_{S_i}, S_i) : D_i \le t\}$, then one can show that $\boldsymbol{V}_{t-\Delta_t}$ is a sufficient statistic of $\boldsymbol{W}_t$ for inferring $X_t$ and $I(X_t; \boldsymbol{W}_t) = I(X_t; \boldsymbol{V}_{t-\Delta_t})$ is still a non-negative and non-increasing function of $\Delta_t$.

If the sampling times $S_i$ are determined by using causal knowledge of $X_t$, $I(X_t; \boldsymbol{W}_t)$ is not necessarily a function of the age. One interesting future research direction is how to choose the sampling time $S_i$ based on the signal and utilize the timing information in $S_i$ to improve data freshness.

Next, we provide the closed-form expression of $I(X_t; \boldsymbol{W}_t)$ for two Markov sources:

*Gauss-Markov source:* Suppose that $X_t$ is a first-order discrete-time Gauss-Markov process, defined by

$$X_t = aX_{t-1} + V_t, \tag{9}$$

where $a \in (-1, 1)$ and the $V_t$'s are zero-mean i.i.d. Gaussian random variables with variance $\sigma^2$. Because $X_t$ is a Gauss-Markov process, one can show that [74]

$$I(X_t; \boldsymbol{W}_t) = I(X_t; X_{t-\Delta_t}) = -\frac{1}{2}\log_2\left(1 - a^{2\Delta_t}\right). \tag{10}$$

Since $a \in (-1, 1)$ and $\Delta_t \ge 0$ is an integer, $I(X_t; \boldsymbol{W}_t)$ is a positive and decreasing function of the age $\Delta_t$. Note that if $\Delta_t = 0$, then $I(X_t; \boldsymbol{W}_t) = H(X_t) = \infty$, because the absolute entropy of a Gaussian random variable is infinite.

*Binary Markov source:* Suppose that $X_t \in \{0, 1\}$ is a binary symmetric Markov process defined by

$$X_t = X_{t-1} \oplus V_t, \tag{11}$$

where $\oplus$ denotes binary modulo-2 addition and the $V_t$'s are i.i.d. Bernoulli random variables with mean $q \in [0, \frac{1}{2}]$. One can show that

$$I(X_t;\boldsymbol{W}_t)=I(X_t;X_{t-\Delta_t})=1-h\left(\frac{1-(1-2q)^{\Delta_t}}{2}\right), \quad (12)$$

where $\Pr[X_t=1|X_0=0]=(1-(1-2q)^t)/2$ and $h(x)$ is the binary entropy function defined by $h(x)=-x\log_2 x-(1-x)\log_2(1-x)$ with a domain $x\in[0,1]$ [73, Eq. (2.5)]. Because $h(x)$ is increasing on $[0,\frac{1}{2}]$, $I(X_t;\boldsymbol{W}_t)$ is a non-negative and decreasing function of the age $\Delta_t$.

Similarly, one can also use the conditional entropy $H(X_t|\boldsymbol{W}_t)$ to represent the staleness of $\boldsymbol{W}_t$ [11]–[13]. In particular, $H(X_t|\boldsymbol{W}_t)$ can be interpreted as the amount of uncertainty about the current source value $X_t$ after receiving the samples $\boldsymbol{W}_t$. If the $S_i$'s are independent of $\{X_t,t\geq 0\}$ and $X_t$ is a stationary Markov chain, $H(X_t|\boldsymbol{W}_t)=H(X_t|\{X_{S_i}:D_i\leq t\})=H(X_t|X_{t-\Delta_t})$ is a non-decreasing function $p(\Delta_t)$ of the age $\Delta_t$. If the sampling times $S_i$ are determined based on causal knowledge of $X_t$, $H(X_t|\boldsymbol{W}_t)$ is not necessarily a function of the age.

More usage cases of $p(\cdot)$ and $u(\cdot)$ can be found in [5]–[10]. Other data freshness metrics that cannot be expressed as functions of $\Delta_t$ were discussed in [52], [56]–[60].

### *C. Formulation of Optimal Sampling Problems*

Let $\pi=(S_1,S_2,\cdots)$ represent a sampling policy and $\Pi$ denote the set of *causal* sampling policies that satisfy the following two conditions: *(i)* Each sampling time $S_i$ is chosen based on history and current information of the idle/busy state of the channel. *(ii)* The inter-sampling times $\{T_i=S_{i+1}-S_i, i=1,2,\cdots\}$ form a regenerative process [75, Section 6.1][4]: There exists an increasing sequence $0\leq k_1<k_2<\cdots$ of almost surely finite random integers such that the post-$k_j$ process $\{T_{k_j+i}, i=1,2,\cdots\}$ has the same distribution as the post-$k_1$ process $\{T_{k_1+i}, i=1,2,\cdots\}$ and is independent of the pre-$k_j$ process $\{T_i, i=1,2,\cdots,k_j-1\}$; in addition, $\mathrm{E}[k_{j+1}-k_j]<\infty$, $\mathrm{E}[S_{k_1}]<\infty$, and $0<\mathrm{E}[S_{k_{j+1}}-S_{k_j}]<\infty$, $j=1,2,\cdots$

We assume that the sampling times $S_i$ are independent of the source process $\{X_t,t\geq 0\}$, and the service times $Y_i$ of the queue do not change according to the sampling policy. We further assume that $\mathrm{E}[p(\Delta+Y_i)]<\infty$ for all finite $\Delta$.

In this paper, we study the optimal sampling policy that minimizes (maximizes) the average age penalty (utility) subject to an average sampling rate constraint. In the continuous-time case, we will consider the following problem:

$$\bar{p}_{\text{opt},1}=\inf_{\pi\in\Pi}\ \limsup_{T\to\infty}\frac{1}{T}\,\mathrm{E}\left[\int_0^T p(\Delta_t)dt\right] \quad (13)$$

$$\text{s.t.}\ \liminf_{n\to\infty}\frac{1}{n}\mathrm{E}\left[S_n\right]\geq\frac{1}{f_{\max}}, \quad (14)$$

where $\bar{p}_{\text{opt},1}$ is the optimal value of (13) and $f_{\max}$ is the maximum allowed sampling rate. In the discrete-time case, we need to solve the following optimal sampling problem:

$$\bar{p}_{\text{opt},2}=\inf_{\pi\in\Pi}\ \limsup_{n\to\infty}\frac{1}{n}\,\mathrm{E}\left[\sum_{t=1}^{n}p(\Delta_t)\right] \quad (15)$$

$$\text{s.t.}\ \liminf_{n\to\infty}\frac{1}{n}\mathrm{E}\left[S_n\right]\geq\frac{1}{f_{\max}}, \quad (16)$$

where $\bar{p}_{\text{opt},2}$ is the optimal value of (15). We assume that $\bar{p}_{\text{opt},1}$ and $\bar{p}_{\text{opt},2}$ are finite. The problems for maximizing the average age utility can be readily obtained from (13) and (15) by choosing $p(\Delta)=-u(\Delta)$. In practice, the cost for data updates increases with the average sampling rate. Therefore, Problems (13) and (15) represent a tradeoff between data staleness (freshness) and update cost.

Problems (13) and (15) are constrained MDPs, one with a continuous (uncountable) state space and the other with a countable state space. Because of the *curse of dimensionality* [76], it is quite rare that one can explicitly solve such problems and derive analytical or closed-form solutions that are arbitrarily accurate.

[4]We assume that $T_i$ is a regenerative process because we will optimize $\limsup_{T\to\infty}\mathrm{E}[\int_0^T p(\Delta_t)dt]/T$, but operationally a nicer objective function is $\limsup_{i\to\infty}\mathrm{E}[\int_0^{D_i} p(\Delta_t)dt]/\mathrm{E}[D_i]$. These two criteria are equivalent, if $\{T_1,T_2,\cdots\}$ is a regenerative process, or more generally, if $\{T_1,T_2,\cdots\}$ has only one ergodic class. If no condition is imposed, however, they are different.

**Algorithm 1** Bisection method for solving (18).

**given** $l$, $u$, tolerance $\epsilon>0$.
**repeat**
  $\beta:=(l+u)/2$.
  $o:=\beta-\frac{\mathrm{E}[v(D_{i+1}(\beta)-S_i(\beta))-v(Y_i)]}{\mathrm{E}[D_{i+1}(\beta)-D_i(\beta)]}$.
  **if** $o\geq 0$, $u:=\beta$; **else**, $l:=\beta$.
**until** $u-l\leq\epsilon$.
**return** $\beta$.

## IV. MAIN RESULTS: OPTIMAL SAMPLING POLICIES

In this section, we present a complete characterization of the solutions to (13) and (15). Specifically, the optimal sampling policies are either deterministic or randomized threshold policies, depending on the scenario under consideration. Efficient computation algorithms of the thresholds and the randomization probabilities will be provided.

### *A. Continuous-time Sampling without Rate Constraint*

We first consider the continuous-time sampling problem (13). When there is no sampling rate constraint (i.e., $f_{\max}=\infty$), a solution to (13) is provided in the following theorem:

**Theorem 1** (Continuous-time sampling without rate constraint) If $f_{\max}=\infty$, $p(\cdot)$ is non-decreasing, and the service times $Y_i$ are i.i.d. with $0<\mathrm{E}[Y_i]<\infty$, then $(S_1(\beta),S_2(\beta),\cdots)$ with a parameter $\beta$ is an optimal solution to (13), where

$$S_{i+1}(\beta)=\inf\{t\geq D_i(\beta):\mathrm{E}[p(\Delta_{t+Y_{i+1}})]\geq\beta\}, \quad (17)$$

$D_i(\beta)=S_i(\beta)+Y_i$, $\Delta_t=t-S_i(\beta)$, and $\beta$ is the root of

$$\beta=\frac{\mathrm{E}\left[\int_{D_i(\beta)}^{D_{i+1}(\beta)}p(\Delta_t)dt\right]}{\mathrm{E}[D_{i+1}(\beta)-D_i(\beta)]}. \quad (18)$$

Further, $\beta$ is exactly the optimal value to (13), i.e., $\beta=\bar{p}_{\text{opt},1}$.

The proof of Theorem 1 is relegated to Subsection V.F. The optimal sampling policy in (17)-(18) has a nice structure. Specifically, the $(i+1)$th sample is generated at the earliest time $t$ satisfying two conditions: *(i)* the $i$th sample has already been delivered by time $t$, i.e., $t \geq D_i(\beta)$, and *(ii)* the expected age penalty $\mathrm{E}[p(\Delta_{t+Y_{i+1}})]$ has grown to be no smaller than a pre-determined threshold $\beta$. Notice that if $t = S_{i+1}(\beta)$, then $t + Y_{i+1} = S_{i+1}(\beta) + Y_{i+1} = D_{i+1}(\beta)$ is the delivery time of the $(i+1)$th sample. In addition, $\beta$ is equal to the optimum objective value $\bar{p}_{\text{opt},1}$ of (13). Hence, (17)-(18) require that the expected age penalty upon the delivery of the $(i+1)$th sample is no smaller than $\bar{p}_{\text{opt},1}$, i.e., the minimum possible time-average expected age penalty.

Next, we develop an efficient algorithm to find the root $\beta$ of (18). Because the $Y_i$'s are i.i.d., the expectations on the right-hand side of (18) are functions of $\beta$ and are irrelevant of $i$. Given $\beta$, these expectations can be evaluated by Monte Carlo simulations or importance sampling. Define

$$v(s) = \int_0^s p(t)dt, \tag{19}$$

then

$$\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta_t)dt = v(D_{i+1}(\beta) - S_i(\beta)) - v(Y_i), \tag{20}$$

which can be used to simplify the numerical evaluation of the expected integral in (18). As proven in Subsection V.F, (18) has a unique solution. We use a simple bisection method to solve (18), which is illustrated in Algorithm 1.

A.1 Optimality Condition of Zero-wait Sampling

When $f_{\max} = \infty$, one logical sampling policy is the zero-wait sampling policy [4], [15], [24], given by

$$S_{i+1} = S_i + Y_i. \tag{21}$$

This zero-wait sampling policy achieves the maximum throughput and the minimum queueing delay. In the special case of $p(\Delta_t) = \Delta_t$, Theorem 5 of [15] provided a sufficient and necessary condition for characterizing the optimality of the zero-wait sampling policy. We now generalize that result to the case of non-linear age functions in the following corollary:

**Corollary 1:** If $f_{\max} = \infty$, $p(\cdot)$ is non-decreasing, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then the zero-wait sampling policy in (21) is optimal for solving (13) if and only if

$$\mathrm{E}\left[p(\operatorname{ess\,inf} Y_i + Y_{i+1})\right] \geq \frac{\mathrm{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}} p(t)dt\right]}{\mathrm{E}[Y_{i+1}]}, \tag{22}$$

where $\operatorname{ess\,inf} Y_i = \inf\{y \in [0,\infty) : \Pr[Y_i \leq y] > 0\}$.

*Proof:* See Appendix D. □

One can consider $\operatorname{ess\,inf} Y_i$ as the minimum possible value of $Y_i$. It immediately follows from Corollary 1 that

**Corollary 2:** If $f_{\max} = \infty$, $p(\cdot)$ is non-decreasing, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then the following assertions are true:

(a) If $Y_i$ is a constant, then (21) is optimal for solving (13).

(b) If $\operatorname{ess\,inf} Y_i = 0$ and $p(\cdot)$ is strictly increasing, then (21) is **not** optimal for solving (13).

*Proof:* See Appendix E. □

The condition $\operatorname{ess\,inf} Y_i = 0$ is satisfied by many commonly used distributions, such as exponential distribution, geometric distribution, Erlang distribution, and hyperexponential distribution. According to Corollary 2(b), if $p(\cdot)$ is strictly increasing, the zero-wait sampling policy (21) is not optimal for these commonly used distributions.

**Algorithm 2** Bisection method for solving (27).

**given** $l$, $u$, tolerance $\epsilon > 0$.
**repeat**
  $\beta := (l+u)/2$.
  $o_1 := \mathrm{E}[T_{i,\min}(\beta) - S_i(\beta)]$.
  $o_2 := \mathrm{E}[T_{i,\max}(\beta) - S_i(\beta)]$.
  **if** $o_1 > \frac{1}{f_{\max}}$, $u := \beta$;
  **else if** $o_2 < \frac{1}{f_{\max}}$, $l := \beta$;
  **else return** $\beta$.
**until** $u - l \leq \epsilon$.
**return** $\beta$.

*B. Continuous-time Sampling with Rate Constraint*

When the sampling rate constraint (14) is imposed, a solution to (13) is presented in the following theorem:

**Theorem 2** (Continuous-time sampling with rate constraint) If $p(\cdot)$ is non-decreasing, $\mathrm{E}[p(t+Y_i)] < \infty$ for all finite $t$, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then (17)-(18) is an optimal solution to (13), if

$$\mathrm{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}. \tag{23}$$

Otherwise, $(S_1(\beta), S_2(\beta), \cdots)$ with a parameter $\beta$ is an optimal solution to (13), where

$$S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) \text{ with probability } \lambda, \\ T_{i,\max}(\beta) \text{ with probability } 1-\lambda. \end{cases} \tag{24}$$

$T_{i,\min}(\beta)$ and $T_{i,\max}(\beta)$ are given by

$$T_{i,\min}(\beta) = \inf\{t \geq D_i(\beta) : \mathrm{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta\}, \tag{25}$$

$$T_{i,\max}(\beta) = \inf\{t \geq D_i(\beta) : \mathrm{E}[p(\Delta_{t+Y_{i+1}})] > \beta\}. \tag{26}$$

$D_i(\beta) = S_i(\beta) + Y_i$, $\Delta_t = t - S_i(\beta)$, $\beta$ is determined by solving

$$\mathrm{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathrm{E}[T_{i,\max}(\beta) - S_i(\beta)], \tag{27}$$

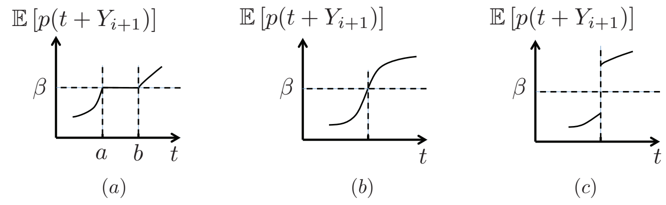



Fig. 4. Three cases of function $f(t) = \mathrm{E}[p(t+Y_{i+1})]$.

and $\lambda$ is given by[5]

$$\lambda = \frac{\mathrm{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathrm{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}. \tag{28}$$

The proof of Theorem 2 will be provided in Section V. According to Theorem 2, the solution to (13) consists of two cases: In *Case 1*, the deterministic threshold policy in Theorem 1 is an optimal solution to (13), which needs to satisfy (23). In *Case 2*, the randomized threshold policy in (24)-(28) is an optimal solution to (13), which needs to satisfy

$$\mathrm{E}[S_{i+1}(\beta) - S_i(\beta)] = \frac{1}{f_{\max}}. \tag{29}$$

We note that the only difference between (25) and (26) is that "$\geq$" is used in (25) while "$>$" is employed in (26).[6] If there exists a time-interval $[a, b]$ such that

$$\mathrm{E}[p(t + Y_{i+1})] = \beta \text{ for all } t \in [a, b] \tag{30}$$

As shown in Fig. 4(a), then $T_{i,\min}(\beta) < T_{i,\max}(\beta)$. In this case, the choices $S_{i+1}(\beta) = T_{i,\min}(\beta)$ and $S_{i+1}(\beta) = T_{i,\max}(\beta)$ may not satisfy (29), but their randomized mixture in (24) can satisfy (29). In particular, if $\beta$ and $\lambda$ are given by (27) and (28), then (29) is satisfied.

We provide a low-complexity algorithm to compute the randomized threshold policy in (24)-(28): As shown in Appendix C, there is a unique $\beta$ satisfying (27). We use the bisection method in Algorithm 2 to solve (27) and obtain $\beta$. After that, $S_{i+1}(\beta)$ and $\lambda$ can be computed by substituting $\beta$ into (24)-(26) and (28). Because of the similarity between (25) and (26), $S_{i+1}(\beta)$ and $\lambda$ are quite sensitive to the numerical error in $\beta$. This issue can be resolved by replacing $T_{i,\min}(\beta)$ in (24) and (28) with $T'_{i,\min}(\beta)$ and replacing $T_{i,\max}(\beta)$ in (24) and (28) with $T'_{i,\max}(\beta)$, where $T'_{i,\min}(\beta)$ and $T'_{i,\max}(\beta)$ are determined by

$$T'_{i,\min}(\beta) = \inf\{t \geq D_i(\beta) : \mathrm{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta - \epsilon/2\}, \tag{31}$$
$$T'_{i,\max}(\beta) = \inf\{t \geq D_i(\beta) : \mathrm{E}[p(\Delta_{t+Y_{i+1}})] > \beta + \epsilon/2\}, \tag{32}$$

respectively, and $\epsilon > 0$ is the tolerance in Algorithm 2. One can improve the accuracy of this solution by *(i)* reducing the tolerance $\epsilon$ and *(ii)* computing the expectations more accurately by increasing the number of Monte Carlo realizations or using advanced techniques such as importance sampling.

As depicted in Figs. 4(b) and 4(c), if $\mathrm{E}[p(t + Y_{i+1})]$ is strictly increasing on $t \in [0, \infty)$, then $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely and (24) reduces to a deterministic threshold policy. In this case, Theorem 2 can be greatly simplified, as stated in the following corollary:

**Corollary 3:** In Theorem 2, if $\mathrm{E}[p(t + Y_{i+1})]$ is strictly increasing in $t$, then (17) is an optimal solution to (13), where $D_i(\beta) = S_i(\beta) + Y_i$, $\Delta_t = t - S_i(\beta)$, and $\beta$ is determined by (18), if

$$\mathrm{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}. \tag{33}$$

Otherwise, $\beta$ is determined by solving

$$\mathrm{E}[S_{i+1}(\beta) - S_i(\beta)] = \frac{1}{f_{\max}}. \tag{34}$$

The proof of Corollary 3 is omitted, because it follows immediately from Theorem 2. If $p(\cdot)$ is strictly increasing or the distribution of $Y_i$ is sufficiently smooth, $\mathrm{E}[p(t+Y_{i+1})]$ is strictly increasing in $t$. Hence, the extra condition in Corollary 3 is satisfied for a broad class of age penalty functions and service time distributions.

A restrictive case of problem (13) was studied in [15], where $p(\cdot)$ was assumed to be positive and non-decreasing. There is an error in Theorem 3 of [15], because the condition "$\mathrm{E}[p(t+Y_{i+1})]$ is strictly increasing in $t$" is missing. Further, the solution in Theorem 3 of [15] is more complicated than that in Corollary 3. A special case of Corollary 3 with $p(t) = t$ was derived in Theorem 4 of [15].

### *C. Discrete-time Sampling*

We now move on to the discrete-time sampling problem (15). When there is no sampling rate constraint (i.e., $f_{\max} = \infty$), the solution to (15) is provided in the following theorem:

**Theorem 3** (Discrete-time sampling without rate constraint) If $f_{\max} = \infty$, $p(\cdot)$ is non-decreasing, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \cdots)$ is an optimal solution to (15), where

$$S_{i+1}(\beta) = \min\{t \in \mathrm{N} : t \geq D_i(\beta), \mathrm{E}\left[p(\Delta_{t+Y_{i+1}})\right] \geq \beta\}, \tag{35}$$

$\mathrm{N} = \{0, 1, 2, \cdots\}$ denotes the set of non-negative integers, $D_i(\beta) = S_i(\beta) + Y_i$, $\Delta_t = t - S_i(\beta)$, and $\beta$ is the root of

$$\beta = \frac{\mathrm{E}\left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1} p(\Delta_t)\right]}{\mathrm{E}[D_{i+1}(\beta) - D_i(\beta)]}. \tag{36}$$

Further, $\beta$ is exactly the optimal value to (15), i.e., $\beta = \bar{p}_{\text{opt},2}$.

The proofs of the discrete-time sampling results will be discussed in Subsection V.G. Theorem 3 is quite similar to Theorem 1, with two minor differences: *(i)* The sampling time $S_{i+1}(\beta)$ in (17) is a real number, which is restricted to an integer in (35). *(ii)* The integral in (18) becomes a summation in (36).

In the discrete-time case, the optimality of the zero-wait sampling policy is characterized as follows.

[5] If $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely, then (24) becomes a deterministic threshold policy and $\lambda$ can be any number within $[0, 1]$.

[6] Clearly, an important issue is the optimality of such a randomized threshold policy, which is proven in Section V.

**Corollary 4:** If $f_{\max} = \infty$, $p(\cdot)$ is non-decreasing, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then the zero-wait sampling policy (21) is optimal for solving (15) if and only if there exists $e < 1$ such that

$$\mathrm{E}\left[p(\operatorname{ess\,inf} Y_i + Y_{i+1} + e)\right] \geq \frac{\mathrm{E}\left[\sum_{t=Y_i}^{Y_i+Y_{i+1}-1} p(t)\right]}{\mathrm{E}[Y_{i+1}]}, \quad (37)$$

where $\operatorname{ess\,inf} Y_i = \min\{y \in \mathrm{N} : \Pr[Y_i \leq y] > 0\}$.

When the sampling rate constraint (16) is imposed, the solution to (15) is provided in the following theorem.

**Theorem 4** (Discrete-time sampling with rate constraint) If $p(\cdot)$ is non-decreasing, $\mathrm{E}[p(t+Y_i)] < \infty$ for all finite $t$, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then (35)-(36) is an optimal solution to (15), if

$$\mathrm{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}. \quad (38)$$

Otherwise, $(S_1(\beta), S_2(\beta), \ldots)$ is an optimal solution to (15), where

$$S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) \text{ with probability } \lambda, \\ T_{i,\max}(\beta) \text{ with probability } 1-\lambda. \end{cases} \quad (39)$$

$T_{i,\min}(\beta)$ and $T_{i,\max}(\beta)$ are given by

$$T_{i,\min}(\beta) = \min\{t \in \mathrm{N} : t \geq D_i(\beta), \mathrm{E}\left[p(\Delta_{t+Y_{i+1}})\right] \geq \beta\}, \quad (40)$$

$$T_{i,\max}(\beta) = \min\{t \in \mathrm{N} : t \geq D_i(\beta), \mathrm{E}\left[p(\Delta_{t+Y_{i+1}})\right] > \beta\}, \quad (41)$$

$D_i(\beta) = S_i(\beta) + Y_i$, $\Delta_t = t - S_i(\beta)$, $\beta$ is determined by solving

$$\mathrm{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathrm{E}[T_{i,\max}(\beta) - S_i(\beta)], \quad (42)$$

and $\lambda$ is given by

$$\lambda = \frac{\mathrm{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathrm{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}. \quad (43)$$

Theorem 4 is similar to Theorem 2, but there are two differences: *(i)* $T_{i,\min}(\beta)$ and $T_{i,\max}(\beta)$ are real numbers in (25)-(26), which are restricted to integers in (40)-(41). *(ii)* If $\mathrm{E}[p(t+Y_{i+1})]$ is strictly increasing in $t$, then $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ holds almost surely in (25)-(26) and Theorem 2 can be greatly simplified. However, in the discrete-time case, even if $\mathrm{E}[p(t+Y_{i+1})]$ is strictly increasing in $t$, $T_{i,\min}(\beta) < T_{i,\max}(\beta)$ may still occur in (40)-(41). In fact, it is rather common that $T_{i,\min}(\beta) < T_{i,\max}(\beta)$ holds for the optimal $\beta$, because of the following reason: If $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely, then (39) becomes a deterministic threshold policy that needs to ensure (29). However, because $S_{i+1}(\beta)$ and $S_i(\beta)$ are integers, such a deterministic threshold policy is difficult to satisfy (29) for certain values of $f_{\max}$. On the other hand, if $T_{i,\min}(\beta) < T_{i,\max}(\beta)$, the randomized threshold policy in (39)-(43) can satisfy (29). Hence, even though $\mathrm{E}[p(t+Y_{i+1})]$ is strictly increasing in $t$, Theorem 4 cannot be further simplified. This is a key difference between continuous-time and discrete-time sampling.

The computation algorithms of the optimal discrete-time sampling policies are similar to their counterparts in the continuous-time case, and hence are omitted.

### D. An Example: Mutual Information Maximization

Next, we provide an example to illustrate the above theoretical results. Suppose that $X_t$ is a stationary, time-homogeneous Markov chain and the sampling times $S_i$ are independent of $\{X_t, t \geq 0\}$. The optimal sampling problem that maximizes the time-average expected mutual information between $X_t$ and $\boldsymbol{W}_t$ is formulated as

$$\bar{I}_{\text{opt}} = \sup_{\pi \in \Pi} \liminf_{n \to \infty} \frac{1}{n} \mathrm{E}\left[\sum_{t=1}^{n} I(X_t; \boldsymbol{W}_t)\right], \quad (44)$$

where $\bar{I}_{\text{opt}}$ is the optimal value of (44). We assume that $\bar{I}_{\text{opt}}$ is finite. Problem (44) is a special case of (15) satisfying $p(\Delta_t) = -u(\Delta_t) = -I(X_t; \boldsymbol{W}_t)$ and $f_{\max} = \infty$. The following result follows immediately from Theorem 3.

**Corollary 5:** If the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \cdots)$ is an optimal solution to (15), where

$$S_{i+1}(\beta) = \min\{t \in \mathrm{N} : t \geq D_i(\beta), \; I(X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}) \leq \beta\}. \quad (45)$$

$D_i(\beta) = S_i(\beta) + Y_i$, and $\beta \geq 0$ is the root of

$$\beta = \frac{E\left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1} I(X_t; X_{S_i(\beta)})\right]}{E[D_{i+1}(\beta) - D_i(\beta)]}. \quad (46)$$

Further, $\beta$ is exactly the optimal value of (44), i.e., $\beta = \bar{I}_{\text{opt}}$.

In Corollary 5, the next sampling time $S_{i+1}(\beta)$ is determined based on the mutual information between the freshest received sample $X_{S_i(\beta)}$ and the source value $X_{D_{i+1}(\beta)}$, where $D_{i+1}(\beta) = S_{i+1}(\beta) + Y_{i+1}$ is the delivery time of the $(i+1)$th sample. Because $Y_{i+1}$ will be known by both the transmitter and receiver at time $D_{i+1}(\beta) = S_{i+1}(\beta) + Y_{i+1}$, $Y_{i+1}$ is the side information in the conditional mutual information $I[X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}]$. The conditional mutual information $I[X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}]$ decreases as time $t$ grows. According to (45), the $(i+1)$-th sample is generated at the smallest integer $t$ satisfying two conditions: *(i)* the $i$th sample has already been delivered by time $t$ and *(ii)* the conditional mutual information $I[X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}]$ has reduced to be no greater than $\bar{I}_{\text{opt}}$, i.e., the optimum of the time-average expected mutual information $\liminf_{n\to\infty}(1/n)\,\mathrm{E}\left[\sum_{t=1}^{n} I(X_t; \boldsymbol{W}_t)\right]$ that we are maximizing.

The optimal sampling policy is illustrated in Fig. 5, where $\beta = 0.1$ and $Y_i$ is equal to either 1 or 5 with equal probability. The sampling time $S_i(\beta)$, delivery time $D_i(\beta)$, and conditional mutual information $I[X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}]$ are depicted in the figure. One can observe that if the service time of the previous sample is $Y_i = 1$, the sampler will wait until the conditional mutual information drops below the threshold $\beta$ and then take the next sample; if the service time of the previous sample is

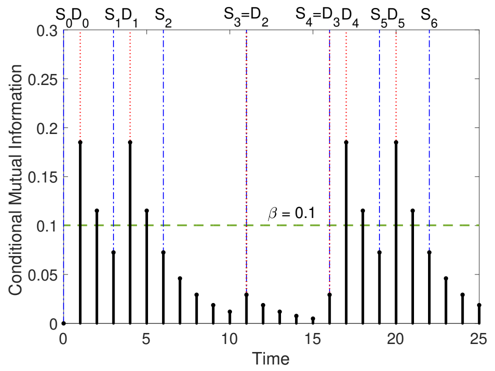



Fig. 5. A sample-path illustration of the optimal sampling policy (45) and (46), where $\beta = 0.1$, $Y_i$ is either 1 or 5 with equal probability, $S_i$ and $D_i$ are sampling time and delivery time of the $i$-th sample. On this sample-path, the service times are $Y_0 = 1, Y_1 = 1, Y_2 = 5, Y_3 = 5, Y_4 = 1, Y_5 = 1, Y_6 = 5$.

$Y_i = 5$, the next sample is taken once the previous sample is delivered, because the conditional mutual information is already below $\beta$ then.

### E. Alternative Expressions of the Threshold Sampling Policy

Finally, we present two alternative expressions of the sampling policy (17). Define

$$w(\beta) = \inf\{\Delta \geq 0 : \mathrm{E}[p(\Delta + Y_{i+1})] \geq \beta\}, \tag{47}$$

then (17) can be rewritten as

$$S_{i+1}(\beta) = \inf\{t \geq D_i(\beta) : \Delta_t \geq w(\beta)\}, \tag{48}$$

which is a threshold policy on the age $\Delta_t$. Threshold policies similar to (48) were discussed in age minimization for status update systems with an energy harvesting constraint, e.g., [23], [25], [27], [28], [30]. The technical tools therein are significantly different from ours, because of the energy harvesting constraint. Further, from (3) and (48), we get

$$\begin{aligned} S_{i+1}(\beta) &= \inf\{t \geq D_i(\beta) : \Delta_t \geq w(\beta)\}, \\ &= \inf\{t \geq D_i(\beta) : t \geq w(\beta) + D_i(\beta) - Y_i\}, \\ &= D_i(\beta) + \max\{w(\beta) - Y_i, 0\}. \end{aligned} \tag{49}$$

We use $Z_i(\beta) = S_{i+1}(\beta) - D_i(\beta) \geq 0$ to denote the waiting time from the delivery time of the $i$th sample to the generation time of the $(i+1)$th sample. By (49), $Z_i(\beta)$ can be expressed as a simple water-filling solution, i.e.,

$$Z_i(\beta) = \max\{w(\beta) - Y_i, 0\}, \tag{50}$$

where $w(\beta)$ is the water level. Hence, the waiting time $Z_i(\beta)$ decreases linearly with the service time $Y_i$, until $Z_i(\beta)$ drops to zero. The water-filling solution was shown to be age-optimal for a special case that $p(\Delta_t) = \Delta_t$ [15], [24]. Recently, it was observed via simulations that the water-filling solution comes very close to the optimal age performance in symmetric multi-source networks [77].

## V. PROOFS OF THE MAIN RESULTS

In this section, we prove the main results in Section IV, by using the technical tools recently developed in [47], [48], as well as some additional proof ideas that are needed for showing Lemma 5, Theorem 5, and Lemma 7 below.

We begin with the proof of Theorem 2, because its proof procedure is helpful for presenting and understanding the other proofs.

### A. Suspend Sampling when the Server is Busy

In [15], it was shown that *no new sample should be taken when the server is busy*. The reason is as follows: If a sample is taken when the server is busy, it has to wait in the queue for its transmission opportunity, during which time the sample is becoming stale. A better strategy is to take a new sample just when the server becomes idle, which yields a smaller age process on sample path. This comparison leads to the following lemma:

**Lemma 2:** In the optimal sampling problem (13), it is sub-optimal to take a new sample before the previous sample is delivered.

By Lemma 2, the queue in Fig. 1 should be always kept empty. In addition, we only need to consider a sub-class of sampling policies $\Pi_1 \subset \Pi$ in which each sample is generated after the previous sample is delivered, i.e.,

$$\Pi_1 = \{\pi \in \Pi : S_{i+1} \geq D_i = S_i + Y_i \text{ for all } i\}. \tag{51}$$

Let $Z_i = S_{i+1} - D_i \geq 0$ represent the waiting time between the delivery time $D_i$ of the $i$th sample and the generation time $S_{i+1}$ of the $(i+1)$th sample. Since $S_0 = 0$, we have $S_i = S_0 + \sum_{j=0}^{i}(Y_j + Z_j) = \sum_{j=0}^{i}(Y_j + Z_j)$ and $D_i = S_i + Y_i$. Given $(Y_1, Y_2, \cdots)$, $(S_1, S_2, \cdots)$ is uniquely determined by $(Z_1, Z_2, \cdots)$. Hence, one can also use $\pi = (Z_1, Z_2, \cdots)$ to represent a sampling policy in $\Pi_1$.

Because $T_i$ is a regenerative process, by using the renewal theory in [78] and [75, Section 6.1], one can show that $(1/i)\mathrm{E}[S_i]$ and $(1/i)\mathrm{E}[D_i]$ in (13) are convergent sequences and

$$\begin{aligned} &\limsup_{T\to\infty} \frac{1}{T}\,\mathrm{E}\left[\int_0^T p(\Delta_t)dt\right] \\ =& \lim_{i\to\infty} \frac{\mathrm{E}\left[\int_0^{D_i} p(\Delta_t)dt\right]}{\mathrm{E}[D_i]} \\ =& \lim_{i\to\infty} \frac{\sum_{j=1}^{i} \mathrm{E}\left[\int_{D_j}^{D_{j+1}-1} p(\Delta_t)dt\right]}{\sum_{j=1}^{i} \mathrm{E}\left[Y_j + Z_j\right]}. \end{aligned} \tag{52}$$

On the other hand, it follows from (3) that

$$\begin{aligned} \int_{D_j}^{D_{j+1}} p(\Delta_t)dt =& \int_{D_j}^{D_{j+1}} p(t - S_i)dt \\ =& \int_{Y_i}^{Y_i+Z_i+Y_{i+1}} p(t)dt, \end{aligned} \tag{53}$$

which is a function of $(Y_i, Z_i, Y_{i+1})$. Define

$$q(y_i, z, y') = \int_{y}^{y+z+y'} p(t)dt, \tag{54}$$

then (13) can be rewritten as

$$\bar{p}_{\text{opt}_1} = \inf_{\pi\in\Pi_1} \lim_{i\to\infty} \frac{\sum_{j=1}^{i} \mathrm{E}\left[q(Y_j, Z_j, Y_{j+1})\right]}{\sum_{j=1}^{i} \mathrm{E}\left[Y_j + Z_j\right]} \tag{55}$$

$$\text{s.t. } \lim_{i\to\infty} \frac{1}{i}\sum_{j=1}^{i} \mathrm{E}\left[Y_j + Z_j\right] \geq \frac{1}{f_{\max}}. \tag{56}$$

### *B. Reformulation of Problem* (55)

In order to solve (55), we consider the following MDP with a parameter $c \geq 0$:

$$h(c) \triangleq \inf_{\pi\in\Pi_1} \lim_{i\to\infty} \frac{1}{i}\sum_{j=0}^{i-1} \mathrm{E}\left[q(Y_j, Z_j, Y_{j+1}) - c(Y_j + Z_j)\right] \tag{57}$$

$$\text{s.t. } \lim_{i\to\infty} \frac{1}{i}\sum_{j=1}^{i} \mathrm{E}\left[Y_j + Z_j\right] \geq \frac{1}{f_{\max}}, \tag{58}$$

where $h(c)$ is the optimum value of (57). Similar with Dinkelbach's method [79] for nonlinear fractional programming, the following lemma holds for the MDP (55):

**Lemma 3:** [48, Lemma 2] The following assertions are true:
(a) $\bar{p}_{opt_1} \gtreqless c$ if and only if $h(c) \gtreqless 0$.
(b) If $h(c) = 0$, the solutions to (55) and (57) are identical.

Hence, the solution to (55) can be obtained by solving (57) and seeking $\bar{p}_{\text{opt}_1} \in \mathbb{R}$ that satisfies

$$h(\bar{p}_{\text{opt}_1}) = 0. \tag{59}$$

### *C. Lagrangian Dual Problem of* (57) *when* $c = \bar{p}_{\text{opt}_1}$

Although (57) is a continuous-time MDP with a continuous state space, rather than a convex optimization problem, it is possible to use the Lagrangian dual approach to solve (57) and show that it admits no duality gap.

When $c = \bar{p}_{\text{opt}_1}$, define the following Lagrangian

$$L(\pi;\alpha) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left[q(Y_j, Z_j, Y_{j+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Y_i + Z_i)\right] + \frac{\alpha}{f_{\max}}, \tag{60}$$

where $\alpha \geq 0$ is the dual variable. Let

$$g(\alpha) \triangleq \inf_{\pi\in\Pi_1} L(\pi;\alpha). \tag{61}$$

Then, the Lagrangian dual problem of (57) is defined by

$$d \triangleq \max_{\alpha\geq 0} g(\alpha), \tag{62}$$

where $d$ is the optimum value of (62). Weak duality [80], [81] implies that $d \leq h(\bar{p}_{\text{opt}_1})$. Next, we will solve (61) and establish strong duality, i.e., $d = h(\bar{p}_{\text{opt}_1})$.

### *D. Optimal Solutions to* (61)

We solve (61) in two steps: First, we use a sufficient statistic argument to show that (61) can be decomposed into a series of per-sample optimization problems. Second, each per-sample optimization problem is reformulated as a convex optimization problem, which is solved in closed-form. The details are provided as follows.

**Lemma 4:** If the service times $Y_i$ are i.i.d., then $Y_i$ is a sufficient statistic for determining the optimal $Z_i$ in (61).

*Proof:* In (61), the minimization of the term

$$\begin{aligned} &\mathrm{E}\left[q(Y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Y_i + Z_i)\right] \\ \overset{(a)}{=} &\mathrm{E}\left[q(Y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Z_i + Y_{i+1})\right] \end{aligned} \tag{63}$$

over $Z_i$ depends on $(Y_1, \cdots, Y_i, Z_1, \cdots, Z_{i-1})$ via $Y_i$, where Step (a) is due to $\mathrm{E}[Y_i] = \mathrm{E}[Y_{i+1}]$. Hence, $Y_i$ is a sufficient statistic for determining $Z_i$ in (61). □

By Lemma 4, (61) can be decomposed into a series of per-sample optimization problems. In particular, after observing the realization $Y_i = y_i$, $Z_i$ is determined by solving

$$\min_{\substack{\Pr[Z_i\in A|Y_i=y_i]\\ Z_i\geq 0}} \mathrm{E}\left[q(y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Z_i + Y_{i+1})\right], \tag{64}$$

where the rule for determining $Z_i$ is represented by $\Pr[Z_i \in A|Y_i = y_i]$, i.e., the conditional distribution of $Z_i$ given the occurrence of $Y_i = y_i$. To find all possible solutions to (64), let us consider the following problem

$$\min_{z\geq 0} \mathrm{E}\left[q(y_i, z, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(z + Y_{i+1})\right]. \tag{65}$$

Because $p(\cdot)$ is non-decreasing, the functions $z \to q(y_i, z, y')$ and $z \to \mathrm{E}\left[q(y_i, z, Y_{i+1})\right]$ are both convex. Hence, (65) is a convex optimization problem.

**Lemma 5:** If $p(\cdot)$ is non-decreasing, then the set of optimal solution to (65) is $[z_{\min}(y_i, \alpha), z_{\max}(y_i, \alpha)]$ where

$$z_{\min}(y, \alpha) = \inf\{t \geq 0 : \mathrm{E}[p(y + t + Y_{i+1}] \geq \bar{p}_{\text{opt}_1} + \alpha\}, \tag{66}$$

$$z_{\max}(y, \alpha) = \inf\{t \geq 0 : \mathrm{E}[p(y + t + Y_{i+1}] > \bar{p}_{\text{opt}_1} + \alpha\}. \tag{67}$$

*Proof:* See Appendix B. □

By Lemma 5, $z$ is an optimal solution to (65) if and only if $z \in [z_{\min}(y_i, \alpha), z_{\max}(y_i, \alpha)]$. Hence, the set of optimal solutions to (64) is

$$\{\Pr[Z_i \in A|Y_i = y_i] : Z_i \in [z_{\min}(y_i, \alpha), z_{\max}(y_i, \alpha)] \text{ almost surely}\}.$$

Combining this with Lemma 4, yields

**Lemma 6:** If $p(\cdot)$ is non-decreasing and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then the set of optimal solutions to (61) is

$$\Gamma(\alpha) = \{\pi : Z_i \in [z_{\min}(Y_i, \alpha), z_{\max}(Y_i, \alpha)] \text{ for almost all } i\}, \tag{68}$$

where $z_{\min}(y, \alpha)$ and $z_{\max}(y, \alpha)$ are given in (66)-(67).

### *E. Zero Duality Gap and Optimal Solution to* (57)

Strong duality and an optimal solution to (57) are obtained in the following theorem:

**Theorem 5:** If $c = \bar{p}_{opt_1}$, $p(\cdot)$ is non-decreasing, $\mathrm{E}[p(t+Y_i)] < \infty$ for all finite $t$, and the service times $Y_i$ are i.i.d. with $0 < \mathrm{E}[Y_i] < \infty$, then the duality gap between (57) and (62) is zero. Further, $Z_i = z_{\min}(Y_i, 0)$ is an optimal solution to (57) and (62), if

$$\mathrm{E}[Y_i + z_{\min}(Y_i, 0)] > \frac{1}{f_{\max}}. \tag{69}$$

Otherwise, $(Z_1, Z_2, \ldots)$ is an optimal solution to (57) and (62), where

$$Z_i = \begin{cases} z_{\min}(Y_i, \alpha) \text{ with probability } \lambda, \\ z_{\max}(Y_i, \alpha) \text{ with probability } 1-\lambda. \end{cases} \tag{70}$$

$\alpha \geq 0$ is determined by solving

$$\mathrm{E}\left[Y_i + z_{\min}(Y_i, \alpha)\right] \leq \frac{1}{f_{\max}} \leq \mathrm{E}\left[Y_i + z_{\max}(Y_i, \alpha)\right],$$

and $\lambda$ is given by

$$\lambda = \frac{\mathrm{E}[Y_i + z_{\max}(Y_i, \alpha)] - \frac{1}{f_{\max}}}{\mathrm{E}[z_{\max}(Y_i, \alpha) - z_{\min}(Y_i, \alpha)]}. \tag{71}$$

*Proof:* We use [80, Prop. 6.2.5] to find a *geometric multiplier* for (57). This suggests that the duality gap between (57) and (62) must be zero, because otherwise there exists no geometric multiplier [80, Prop. 6.2.3(b)]. The details are provided in Appendix C. □

By choosing

$$\beta = \bar{p}_{\mathrm{opt}_1} + \alpha. \tag{72}$$

Theorem 2 follows from Theorem 5.

We note that the extension of Dinkelbach's method in Lemma 3 and the geometric multiplier technique used in Theorem 5 are the key technical tools that make it possible to simplify (13) as the convex optimization problem in (65). These technical tools were also used in a recent study [48], where a quite different sampling problem is solved. Further, (72) implies that the optimal threshold $\beta$ is equal to the optimum objective value of the MDP $\bar{p}_{\mathrm{opt}_1}$ plus the optimal Lagrangian dual variable $\alpha$. By using these results, bisection search algorithms are developed in Section IV to compute $\beta$, and the curse of dimensionality is circumvented.

### *F. Proofs of Other Continuous-time Sampling Results*

Theorem 1 follows immediately from Theorem 2, because it is a special case of Theorem 2. In particular, because the $Y_i$'s are i.i.d., the optimal objective value to (13) is

$$\begin{aligned}\bar{p}_{\mathrm{opt}_1} &= \lim_{n\to\infty} \frac{\sum_{i=1}^n \mathrm{E}\left[q(Y_i, z_{\min}(Y_i,0), Y_{i+1})\right]}{\sum_{i=1}^n \mathrm{E}\left[Y_i + z_{\min}(Y_i,0)\right]} \\ &= \frac{\mathrm{E}\left[q(Y_i, z_{\min}(Y_i,0), Y_{i+1})\right]}{\mathrm{E}\left[Y_i + z_{\min}(Y_i,0)\right]} \\ &= \frac{\mathrm{E}\left[\int_{Y_i}^{Y_i+z_{\min}(Y_i,0)+Y_{i+1}} p(t)dt\right]}{\mathrm{E}\left[Y_i + z_{\min}(Y_i,0)\right]},\end{aligned} \tag{73}$$

from which (18) follows. We note that the root of (18) must be unique; otherwise, one can follow the arguments in Appendix C to show that the optimal objective value to (13) is non-unique, which cannot be true. Further, as shown in Appendix C, the condition "$\mathrm{E}[p(t+Y_i)] < \infty$ for all finite $t$" is not needed in the case of Theorem 1.

### *G. Proofs of Discrete-time Sampling Results*

The proofs of the discrete-time sampling results are quite similar to their continuous-time counterparts. One difference is that (65) of the continuous-time case becomes the following integer optimization problem in the discrete-time case:

$$\min_{z \in \mathrm{N}} \mathrm{E}\left[q(y_i, z, Y_{i+1}) - (\bar{p}_{\mathrm{opt}_1} + \alpha)(z + Y_{i+1})\right], \tag{74}$$

where

$$q(y_i, z, y') = \sum_{t=y}^{y+z+y'-1} p(t). \tag{75}$$

By adopting an idea in [82, Problem 5.5.3], we obtain

**Lemma 7:** If $p(\cdot)$ is non-decreasing, then the set of optimal solution to (74) is $\{z_{\min}(y_i, \alpha), z_{\min}(y_i, \alpha)+1, z_{\min}(y_i, \alpha)+2, \cdots, z_{\max}(y_i, \alpha)\}$, where

$$z_{\min}(y, \alpha) = \inf\{t \in \mathrm{N} : \mathrm{E}[p(y+t+Y_{i+1}] \geq \bar{p}_{\mathrm{opt}_1} + \alpha\}, \tag{76}$$

$$z_{\max}(y, \alpha) = \inf\{t \in \mathrm{N} : \mathrm{E}[p(y+t+Y_{i+1}] > \bar{p}_{\mathrm{opt}_1} + \alpha\}. \tag{77}$$

*Proof:* See Appendix F. □

By replacing Lemma 5 with Lemma 7 and following the proof arguments in Subsections V.A–V.F, the discrete-time optimal sampling results can be proven readily.

## VI. NUMERICAL RESULTS

In this section, we compare the age performance of the following three sampling policies:

- *Uniform sampling*: Periodic sampling with a period given by $S_{i+1} - S_i = 1/f_{\max}$ for continuous-time sampling, or $S_{i+1} - S_i = \lceil 1/f_{\max} \rceil$ for discrete-time sampling where $\lceil x \rceil$ is the smallest integer larger than or equal to $x$.
- *Zero-wait*: The sampling policy in (21), which is infeasible when $f_{\max} < 1/\mathrm{E}[Y_i]$.
- *Optimal policy*: The sampling policy given by Theorem 2 for continuous-time sampling, or Theorem 4 for discrete-time sampling.

As the numerical results for continuous-time sampling have been reported in our earlier work [15], we will focus on the case of discrete-time sampling.

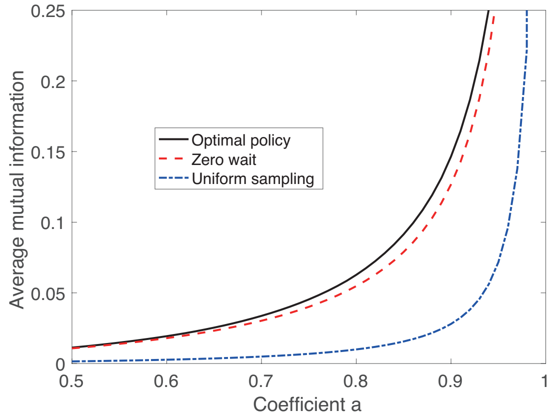



Fig. 6. Average mutual information of the Gauss-Markov source versus the coefficient $a$ in (9), where the service times $Y_i$ are equal to either 1 or 21 with probability 0.5.

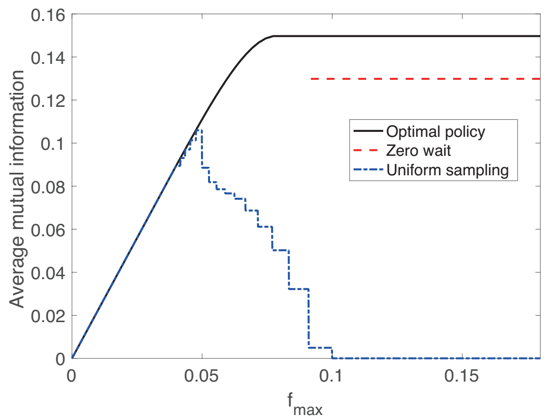



Fig. 7. Average mutual information of the Gauss-Markov source versus $f_{\max}$, where the service times $Y_i$ are equal to either 1 or 21 with probability 0.5.

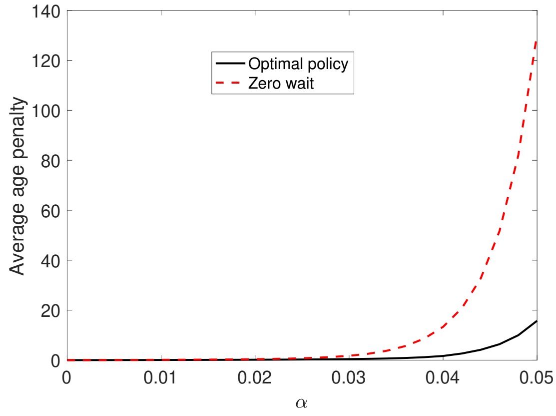



Fig. 8. Average age penalty of an exponential penalty function $p_{\exp}(\Delta_t) = e^{\alpha \Delta_t} - 1$ versus the coefficient $\alpha$, where the service times $Y_i$ follow a discretized log-normal distribution.

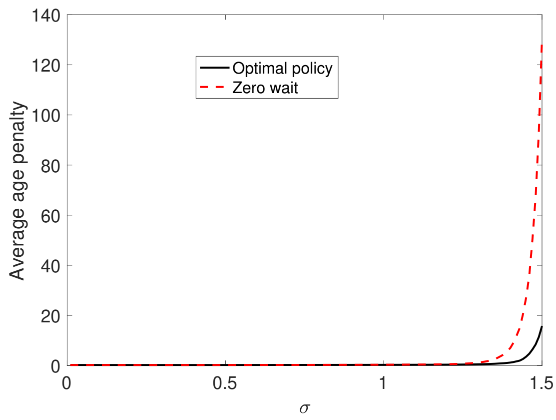



Fig. 9. Average age penalty of an exponential penalty function $p_{\exp}(\Delta_t) = e^{\alpha \Delta_t} - 1$ versus the coefficient $\sigma$ of discretized log-normal service time distribution.

In Fig. 6, we plot the time-average expected mutual information of the Gauss-Markov source versus the coefficient $a$ in (9), where $f_{\max} = 0.095$ and $Y_i$ is equal to either 1 or 21 with probability 0.5. Hence, $\mathrm{E}[Y_i] = 11$ and the zero-wait sampling policy is infeasible when $f_{\max} < 1/11$. Fig. 7 depicts the tradeoff between the time-average expected mutual information of the Gauss-Markov source $X_t$ in (9) and $f_{\max}$, where the mutual information is given by (10) with $a = 0.9$. As the coefficient $a$ grows from 0 to 1, the source $X_t$ becomes more correlated over time. Therefore, the amount of mutual information grows with respect to $a$. In addition, the mutual information of the optimal sampling policy is higher than that of zero-waiting sampling and uniform sampling. When $f_{\max}$ is large, the queue length is high and the samples become stale during the long waiting time in the queue. As a result, uniform sampling is far from optimal for large $f_{\max}$.

Fig. 8 illustrates the time-average expectation of an exponential penalty function $p_{\exp}(\Delta_t) = e^{\alpha \Delta_t} - 1$ versus the coefficient $\alpha$, where $Y_i$ follows a discretized log-normal distribution. In particular, $Y_i$ can be expressed as $Y_i = \lceil e^{\sigma X_i}/\mathrm{E}[e^{\sigma X_i}] \rceil$, where the $X_i$'s are i.i.d. Gaussian random variables with zero mean and unit variance, and $\sigma = 1.5$. Fig. 9 shows the time-average expectation of $p_{\exp}(\Delta_t)$ versus the coefficient $\sigma$ of the discretized log-normal service time distribution. If $\alpha = 0$, $p_{\exp}(\Delta_t)$ is a constant function. If $\sigma = 0$, the service time $Y_i$ is constant. Corollary 4 tells us that zero-waiting sampling is optimal in these two cases, which is in consistent with Figs. 8 and 9. On the other hand, if $\alpha$ and $\sigma$ are large, one can observe from Figs. 8 and 9 that zero-waiting sampling is far from optimal. Hence, zero-wait sampling is far from optimal if the age penalty function grows quickly with the age (i.e., $\alpha$ relatively is large) or the service times $Y_i$ are highly random.

## VII. CONCLUSION

In this paper, we have studied a sampling problem, where samples are taken from a data source and sent to a remote receiver that is in need of fresh data. We have developed the optimal sampling policies that maximize various data freshness metrics subject to a sampling rate constraint. These sampling policies have nice structures and are easy to compute. Their optimality is established under quite general conditions. Our numerical results show that the optimal sampling policies can be much better than zero-wait sampling and the classic uniform sampling.

## APPENDIX A PROOF OF LEMMA 1

If the $S_i$'s are independent of $\{X_t, t \geq 0\}$, the sampling times $\{S_i : D_i \leq t\}$ of delivered packet contain no information about $X_t$. In addition, because $X_t$ is a Markov chain, $X_{\max\{S_i:D_i\leq t\}} = X_{t-\Delta_t}$ contains all the information in $\boldsymbol{W}_t = \{(X_{S_i}, S_i) : D_i \leq t\}$ about $X_t$. In other words, $X_{t-\Delta_t}$ is a sufficient statistic of $\boldsymbol{W}_t$ for inferring $X_t$. Then, (8) follows from [ [73], (2.124)].

Next, because $X_t$ is stationary, $I(X_t; X_{t-\Delta}) = I(X_\Delta; X_0)$ for all $t$, which is a function of $\Delta$. Further, because $X_t$ is a Markov chain, owing to the data processing inequality [73, Theorem 2.8.1], $I(X_\Delta; X_0)$ is non-increasing in $\Delta$. Finally, mutual information is non-negative. This completes the proof.

## APPENDIX B PROOF OF LEMMA 5

The one-sided derivative of a function $h$ in the direction of $w$ at $z$ is denoted as

$$\delta h(z; w) \triangleq \lim_{\epsilon\to 0^+} \frac{h(z+\epsilon w) - h(z)}{\epsilon}. \tag{78}$$

Because the function $h(z) = \mathrm{E}[q(y_i, z, Y_{i+1})]$ is convex, the one-sided derivative $\delta h(z; w)$ of $h(z)$ exist [82, p. 709]. Because $z \to q(y_i, z, y')$ is convex, the function $\epsilon \to [q(y_i, z+\epsilon w, y') - q(y_i, z, y')]/\epsilon$ is non-decreasing and bounded from above on $(0, a]$ for some $a > 0$ [83, Proposition 1.1.2(i)]. By monotone convergence theorem [84, Theorem 1.5.6], we can interchange the limit and integral operators in $\delta h(z; w)$ such that

$$\begin{aligned}
\delta h(z; w) &= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\mathrm{E}\left[q(y_i, z+\epsilon w, Y_{i+1}) - q(y_i, z, Y_{i+1})\right] \\
&= \mathrm{E}\bigg[\lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\big\{q(y_i, z+\epsilon w, Y_{i+1}) - q(y_i, z, Y_{i+1})\big\}\bigg] \\
&= \mathrm{E}\bigg[\lim_{t\to z^+} p(y_i + t + Y_{i+1}) w 1_{\{w>0\}} \\
&\qquad + \lim_{t\to z^-} p(y_i + t + Y_{i+1}) w 1_{\{w<0\}}\bigg] \\
&= \lim_{t\to z^+} \mathrm{E}\left[p(y_i + t + Y_{i+1}) w 1_{\{w>0\}}\right] \\
&\qquad + \lim_{t\to z^-} \mathrm{E}\left[p(y_i + t + Y_{i+1}) w 1_{\{w<0\}}\right],
\end{aligned} \tag{79}$$

where $1_A$ is the indicator function of event $A$. According to [82, p. 710] and the convexity of $h(z)$, $z$ is an optimal solution to (65) if and only if the following assertion is true: If $z > 0$, then

$$\delta h(z; w) - (\bar{p}_{\text{opt}_1} + \alpha) w \geq 0, \ \forall\, w \in \mathbb{R}, \tag{80}$$

otherwise, $z = 0$. Because $w$ in (80) is an arbitrary real number, if we choose $w = 1$, then (80) becomes

$$\lim_{t\to z^+} \mathrm{E}\left[p(y_i + t + Y_{i+1})\right] - (\bar{p}_{\text{opt}_1} + \alpha) \geq 0. \tag{81}$$

Similarly, if we choose $w = -1$, then (80) implies

$$\lim_{t\to z^-} \mathrm{E}\left[p(y_i + t + Y_{i+1})\right] - (\bar{p}_{\text{opt}_1} + \alpha) \leq 0. \tag{82}$$

Because $p(\cdot)$ is non-decreasing, we can obtain from (80)-(82) that if $z > 0$, then $z$ satisfies (83)-(84):

$$\mathrm{E}\left[p(y_i + t + Y_{i+1})\right] - (\bar{p}_{\text{opt}_1} + \alpha) \geq 0, \ \text{if } t > z, \tag{83}$$

$$\mathrm{E}\left[p(y_i + t + Y_{i+1})\right] - (\bar{p}_{\text{opt}_1} + \alpha) \leq 0, \ \text{if } t < z, \tag{84}$$

otherwise, $z = 0$. The smallest $z$ satisfying (83)-(84) is

$$z_{\min}(y_i, \alpha) = \inf\{t \geq 0 : \mathrm{E}\left[p(y_i + t + Y_{i+1})\right] \geq \bar{p}_{\text{opt}_1} + \alpha\},$$

and the largest $z$ satisfying (83)-(84) is

$$\begin{aligned}
z_{\max}(y, \alpha) &= \sup\{t \geq 0 : \mathrm{E}\left[p(y_i + t + Y_{i+1})\right] \leq \bar{p}_{\text{opt}_1} + \alpha\} \\
&= \inf\{t \geq 0 : \mathrm{E}\left[p(y_i + t + Y_{i+1})\right] > \bar{p}_{\text{opt}_1} + \alpha\}.
\end{aligned}$$

Hence, the set of optimal solutions to (65) is given by Lemma 5. This completes the proof.

## APPENDIX C PROOF OF THEOREM 5

According to [80, Prop. 6.2.5], if we can find $\pi^\star = (Z_1^\star, Z_2^\star, \cdots)$ and $\alpha^\star$ that satisfy the following conditions:

$$\pi^\star \in \Pi_1, \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left[Y_i + Z_i^\star\right] - \frac{1}{f_{\max}} \geq 0, \tag{85}$$

$$\alpha^\star \geq 0, \tag{86}$$

$$L(\pi^\star; \alpha^\star) = \inf_{\pi\in\Pi_1} L(\pi; \alpha^\star), \tag{87}$$

$$\alpha^\star \left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left[Y_i + Z_i^\star\right] - \frac{1}{f_{\max}}\right\} = 0, \tag{88}$$

then $\pi^\star$ is an optimal solution to (57) and $\alpha^\star$ is a geometric multiplier [80] for (57). Further, if we can find such $\pi^\star$ and $\alpha^\star$, then the duality gap between (57) and (62) must be zero, because otherwise there is no geometric multiplier [80, Prop. 6.2.3(b)]. The remaining task is to find $\pi^\star$ and $\alpha^\star$ that satisfy (85)-(88).

According to Lemma 6, the set of optimal solutions to (87) is given by $\Gamma(\alpha^\star)$. Hence, we only need to find $\alpha^\star$ and $\pi^\star \in \Gamma(\alpha^\star)$ that satisfy (85), (86), and (88). The search for such $\alpha^\star$ and $\pi^\star$ falls into the following two cases:

**Case 1:** If (69) is satisfied, then $\alpha_1^\star = 0$ and $\pi_1^\star = (z_{\min}(Y_1, 0), z_{\min}(Y_2, 0), \cdots)$ satisfy the conditions (85)-(88).

**Case 2:** If (69) is not satisfied, we seek $\alpha_2^\star \geq 0$ and $\pi_2^\star = (Z_1^\star, Z_2^\star, \cdots) \in \Gamma(\alpha_2^\star)$ that satisfy

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left[Y_i + Z_i^\star\right] = \frac{1}{f_{\max}}. \tag{89}$$

By Lemma 6, we can get from (89) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left[Y_i+z_{\min}(Y_i,\alpha_2^\star)\right]\leq\frac{1}{f_{\max}}\leq\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left[Y_i+z_{\max}(Y_i,\alpha_2^\star)\right].\quad(90)$$

Because the $Y_i$'s are i.i.d., (90) is equivalent to

$$\mathrm{E}\left[Y_i+z_{\min}(Y_i,\alpha_2^\star)\right]\leq\frac{1}{f_{\max}}\leq\mathrm{E}\left[Y_i+z_{\max}(Y_i,\alpha_2^\star)\right].\quad(91)$$

Next, we will find $\alpha_2^\star\geq 0$ that satisfies (91). According to (66)-(67), $z_{\min}(y,\alpha)$ and $z_{\max}(y,\alpha)$ are non-decreasing in $\alpha$. Hence, $\mathrm{E}[z_{\min}(Y_i,\alpha)]$ and $\mathrm{E}[z_{\max}(Y_i,\alpha)]$ are also non-decreasing in $\alpha$. In addition, it holds that for all $\alpha_0>0$

$$\lim_{\alpha\to\alpha_0^-}z_{\max}(y,\alpha)=z_{\min}(y,\alpha_0)\leq z_{\max}(y,\alpha_0)=\lim_{\alpha\to\alpha_0^+}z_{\min}(y,\alpha).\quad(92)$$

By invoking the monotone convergence theorem [84, Theorem 1.5.6], we obtain that for all $\alpha_0>0$

$$\lim_{\alpha\to\alpha_0^-}\mathrm{E}[z_{\max}(Y_i,\alpha)]=\mathrm{E}[z_{\min}(Y_i,\alpha_0)]\leq\mathrm{E}[z_{\max}(Y_i,\alpha_0)]=\lim_{\alpha\to\alpha_0^+}\mathrm{E}[z_{\min}(Y_i,\alpha)].\quad(93)$$

Because $\mathrm{E}[p(t+Y_i)]<\infty$ for all finite $t$, it holds for all $y\geq 0$ that $z_{\max}(y,\alpha)$ will increase to $\infty$ as $\alpha$ grows from 0 to $\infty$. By invoking the monotone convergence theorem again, we obtain that $\mathrm{E}[z_{\max}(Y_i,\alpha)]$ will increase to $\infty$ as $\alpha$ grows from 0 to $\infty$. Hence,

$$[\mathrm{E}[z_{\min}(Y_i,0)],\infty)=\bigcup_{\alpha\geq 0}\Big[\mathrm{E}[z_{\min}(Y_i,\alpha)],\mathrm{E}[z_{\max}(Y_i,\alpha)]\Big].\quad(94)$$

In Case 2, (69) is not satisfied, which implies

$$\frac{1}{f_{\max}}\in[\mathrm{E}[z_{\min}(Y_i,0)],\infty).\quad(95)$$

Hence, (93)-(95) tell us that there exists a unique $\alpha_2^\star\geq 0$ satisfying (91). Further, policy $\pi^\star\in\Gamma(\alpha_2^\star)$ is chosen as

$$Z_i^\star=\begin{cases}z_{\min}(Y_i,\alpha_2^\star)\text{ with probability }\lambda,\\ z_{\max}(Y_i,\alpha_2^\star)\text{ with probability }1-\lambda,\end{cases}\quad(96)$$

where $\lambda$ is given by

$$\lambda=\frac{\mathrm{E}[Y_i+z_{\max}(Y_i,\alpha_2^\star)]-\frac{1}{f_{\max}}}{\mathrm{E}[z_{\max}(Y_i,\alpha_2^\star)-z_{\min}(Y_i,\alpha_2^\star)]}.\quad(97)$$

By combining (91), (95), and (96), (89) follows. Hence, the $\alpha_2^\star$ and $\pi_2^\star$ selected above satisfy the conditions (85)-(88).

In both cases, (85)-(88) are satisfied. By [80, Prop. 6.2.3(b)], the duality gap between (57) and (62) is zero. A solution to (57) and (62) is provided in the arguments above. This completes the proof.

## APPENDIX D
## PROOF OF COROLLARY 1

We note that the zero-wait sampling policy can be expressed as (17) with $\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]\geq\beta$.

In the one direction, if the zero-wait sampling policy is optimal, then the root of (18) must satisfy $\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]\geq\beta$. Substituting this into (17), yields $D_{i+1}(\beta)=D_i(\beta)+Y_{i+1}=S_i(\beta)+Y_i+Y_{i+1}$. Combining this with (18), we get

$$\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]\geq\beta=\frac{\mathrm{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}}p(t)dt\right]}{\mathrm{E}[Y_{i+1}]},\quad(98)$$

which implies (22).

In the other direction, if (22) holds, then by choosing

$$\beta=\frac{\mathrm{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}}p(t)dt\right]}{\mathrm{E}[Y_{i+1}]},\quad(99)$$

we get $\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]\geq\beta$. According to (99), such a $\beta$ is a root of (18). Therefore, the zero-wait sampling policy is optimal. This completes the proof.

## APPENDIX E
## PROOF OF COROLLARY 2

We first prove Part (a). If $Y_i=y$ almost surely, then

$$\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]=p(2y)\geq\frac{\int_y^{2y}p(t)dt}{y}\quad(100)$$

holds for all non-decreasing $p(\cdot)$. Hence, (22) is satisfied and the zero-wait sampling policy is optimal.

Next, we consider Part (b). If $\operatorname{ess\,inf}Y_i=0$, then

$$\mathrm{E}\left[p(\operatorname{ess\,inf}Y_i+Y_{i+1})\right]=\mathrm{E}\left[p(Y_{i+1})\right]=\mathrm{E}\left[p(Y_i)\right].\quad(101)$$

Because $\mathrm{E}[Y_{i+1}]=\mathrm{E}[Y_i]>0$, then the event $Y_{i+1}>0$ has a non-zero probability. Further, because $p(\cdot)$ is strictly increasing, the event $p(t)>p(Y_i)$ for $t\in(Y_i,Y_i+Y_{i+1})$ has a non-zero probability. Hence,

$$\mathrm{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}}p(t)dt\right]>\mathrm{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}}p(Y_i)dt\right]=\mathrm{E}[Y_{i+1}]\mathrm{E}\left[p(Y_i)\right].\quad(102)$$

By combining (101) and (102), (22) is not true and the zero-wait sampling policy is not optimal. This completes the proof.

## APPENDIX F
## PROOF OF LEMMA 7

Using (75), (74) can be expressed as

$$\min_{z\in\mathbb{N}}\mathrm{E}\left[\sum_{t=0}^{z+Y_{i+1}-1}\left[p(t+y_i)-(\bar{p}_{\mathrm{opt}_1}+\alpha)\right]\right].\quad(103)$$

It holds that for $m = 1, 2, 3, \cdots$.

$$\begin{aligned}&\mathrm{E}\Bigg[\sum_{t=0}^{m+Y_{i+1}}[p(t+y_i)-(\bar{p}_{\mathrm{opt}_1}+\alpha)]\\&\qquad-\sum_{t=0}^{m+Y_{i+1}-1}[p(t+y_i)-(\bar{p}_{\mathrm{opt}_1}+\alpha)]\Bigg]\\&=\mathrm{E}\big[p(y_i+m+Y_{i+1})-(\bar{p}_{\mathrm{opt}_1}+\alpha)\big]\,. \end{aligned}\tag{104}$$

Because $p(\cdot)$ is non-decreasing, if $z$ is chosen according to Lemma 7, we can obtain

$$\mathrm{E}\big[p(y_i+t+Y_{i+1})-(\bar{p}_{\mathrm{opt}_1}+\alpha)\big]\leq 0,\ t=0,\cdots,z-1,\tag{105}$$

$$\mathrm{E}\big[p(y_i+t+Y_{i+1})-(\bar{p}_{\mathrm{opt}_1}+\alpha)\big]\geq 0,\ t=z,z+1,\cdots.\tag{106}$$

Using (104)-(106), one can see that $\{z_{\min}(y_i,\alpha), z_{\min}(y_i,\alpha)+1,\ z_{\min}(y_i,\alpha)+2,\cdots,z_{\max}(y_i,\alpha)\}$ is the set of optimal solutions to (74). This completes the proof.

## REFERENCES


[1] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *Proc. IEEE SPAWC Workshop*, June 2018.

[2] C. Shapiro and H. Varian, *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press, 1999.

[3] X. Song and J. W. S. Liu, "Performance of multiversion concurrency control algorithms in maintaining temporal consistency," in *Proc. COMSAC*, Oct. 1990.

[4] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Mar. 2012.

[5] J. Cho and H. Garcia-Molina, "Effective page refresh policies for web crawlers," *ACM Trans. Database Syst.*, vol. 28, no. 4, pp. 390–426, Dec. 2003.

[6] A. Even and G. Shankaranarayanan, "Utility-driven assessment of data quality," *SIGMIS Database*, vol. 38, no. 2, pp. 75–93, May 2007.

[7] B. Heinrich, M. Klier, and M. Kaiser, "A procedure to develop metrics for currency and its application in CRM," *J. Data and Information Quality*, vol. 1, no. 1, pp. 5:1–5:28, June 2009.

[8] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, Apr. 2009.

[9] E. Altman, R. El-Azouzi, D. S. Menasche, and Y. Xu, "Forever young: Aging control for smartphones in hybrid networks," 2011, Available: https://arxiv.org/abs/1009.4733.

[10] S. Razniewski, "Optimizing update frequencies for decaying information," in *Proc. ACM CIKM*, Oct. 2016.

[11] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal self-driven sampling for estimation based on value of information," in *Proc. WODES*, May 2016.

[12] T. Soleymani, S. Hirche, and J. S. Baras, "Maximization of information in energy-limited directed communication," in *Proc. ECC*, June 2016.

[13] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal stationary self-triggered sampling for estimation," in *Proc. IEEE CDC*, Dec. 2016.

[14] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," in *Proc. IEEE INFOCOM*, Apr. 2016.

[15] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[16] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *Proc. IEEE ISIT*, June 2017.

[17] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *Proc. IEEE WiOpt*, June 2019.

[18] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *Proc. IEEE INFOCOM AoI Workshop*, Apr. 2019.

[19] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "AoI-penalty minimization for networked control systems with packet loss," in *Proc. IEEE INFOCOM*, Apr. 2019.

[20] C. Kam, S. Kompella, G. D. Nguyen, and A. Ephremides, "Effect of message transmission path diversity on status age," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1360–1374, Mar. 2016.

[21] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6419–6428, Sept. 2018.

[22] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, in press, 2018.

[23] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. IEEE ITA*, Feb. 2015.

[24] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE ISIT*, June 2015.

[25] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *Proc. IEEE ISIT*, June 2017.

[26] A. Arafa and S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *Proc. IEEE GLOBECOM*, Dec. 2017.

[27] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 193–204, Mar. 2018.

[28] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," 2018, Available: https://arxiv.org/abs/1806.07271.

[29] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: With and without feedback," 2018, Available: https://arxiv.org/abs/1808.05141.

[30] B. T. Bacinoglu, Y. Sun, E. Uysal-Bivikoglu, and V. Mutlu, "Achieving the age-energy tradeoff with a finite-battery energy harvesting source," in *Proc. IEEE ISIT*, June 2018.

[31] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *Proc. IEEE DCC*, Mar. 2016.

[32] R. D. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *Proc. IEEE ISIT*, June 2017.

[33] E. T. Ceran, D. Gunduz, and A. Gyorgy, "Average age of information with hybrid ARQ under a resource constraint," in *Proc. IEEE WCNC*, Apr. 2018.

[34] P. Mayekar, P. Parag, and H. Tyagi, "Optimal lossless source codes for timely updates," in *Proc. IEEE ISIT*, June 2018.

[35] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381–5394, July 2018.

[36] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Proc. IEEE Allerton*, Sept. 2016.

[37] C. Joo and A. Eryilmaz, "Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2556–2568, Dec. 2018.

[38] Y. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE ISIT*, June 2017.

[39] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE INFOCOM*, Apr. 2018.

[40] N. Lu, B. Ji, and B. Li, "Age-based scheduling: Improving data freshness for wireless real-time traffic," in *Proc. ACM MobiHoc*, June 2018.

[41] Z. Jiang, B. Krishnamachari, X. Zheng, S. Zhou, and Z. Niu, "Decentralized status update for age-of-information optimization in wireless multiaccess channels," in *Proc. IEEE ISIT*, June 2018.

[42] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the Internet of Things," 2018, Available: https://arxiv.org/abs/1807.04356.

[43] Y. Xiao and Y. Sun, "A dynamic jamming game for real-time status updates," in *Proc. IEEE INFOCOM*, Apr. 2018.

[44] S. Gopal and S. K. Kaul, "A game theoretic approach to DSRC and WiFi coexistence," in *Proc. IEEE INFOCOM*, Apr. 2018.

[45] T. Soleymani, J. S. Baras, and K. H. Johansson, "Stochastic control with stale information–part i: Fully observable systems," 2018, Available: https://arxiv.org/abs/1810.10983.

[46] C. Sonmez, S. Baghaee, A. Ergisi, and E. Uysal-Biyikoglu, "Age-of-information in practice: Status age measured over TCP/IP connections through WiFi, Ethernet and LTE," in *Proc. IEEE BlackSeaCom*, June 2018.

[47] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *Proc. IEEE ISIT*, June 2017.

[48] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Sampling of the Wiener process for remote estimation over a channel with random delay," 2017, Available: https://arxiv.org/abs/1707.02531.

[49] A. Segev and W. Fang, "Optimal update policies for distributed materialized views," *Manage. Sci.*, vol. 37, no. 7, pp. 851–870, July 1991.

[50] B. Adelberg, H. Garcia-Molina, and B. Kao, "Applying update streams in a soft real-time database system," in *Proc. ACM SIGMOD*, May 1995.
[51] J. Cho and H. Garcia-Molina, "Synchronizing a database to improve freshness," in *Proc. ACM SIGMOD*, 2000, pp. 117–128.
[52] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897–1910, Apr. 2016.
[53] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE ISIT*, June 2015.
[54] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *Proc. IEEE ISIT*, June 2017.
[55] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," 2018, Available: https://arxiv.org/abs/1806.03487.
[56] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. IEEE ISIT*, June 2016.
[57] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, June 2017.
[58] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," accepted by *IEEE Trans. Inf. Theory*, 2019.
[59] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *ACM/IEEE Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, June 2019.
[60] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *Proc. IEEE INFOCOM*, Apr. 2018.
[61] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. IEEE Allerton*, Oct. 2017.
[62] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. ACM MobiHoc*, June 2018.
[63] G. D. Nguyen, S. Kompella, C. Kam, J. E. Wieselthier, and A. Ephremides, "Impact of hostile interference on information freshness: A game approach," in *Proc. IEEE WiOpt*, May 2017.
[64] G. D. Nguyen, S. Kompella, C. Kam, J. E. Wieselthier, and A. Ephremide, "Information freshness over an interference channel: A game theoretic view," in *Proc. IEEE INFOCOM*, Apr. 2018.
[65] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.
[66] M. Costa, S. Valentin, and A. Ephremides, "On the age of channel state information for non-reciprocal wireless links," in *Proc. IEEE ISIT*, June 2015.
[67] S. Farazi, A. G. Klein, and D. R. Brown, "On the average staleness of global channel state information in wireless networks with random transmit node selection," in *Proc. IEEE ICASSP*, Mar. 2016.
[68] J. Zhang and C. Wang, "On the rate-cost of Gaussian linear control systems with random communication delays," in *Proc. IEEE ISIT*, June 2018.
[69] C. Kam, S. Kompella, and A. Ephremides, "Experimental evaluation of the age of information via emulation," in *Proc. IEEE MILCOM*, Oct. 2015.
[70] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Modeling the age of information in emulated ad hoc networks," in *Proc. IEEE MILCOM*, Oct. 2017.
[71] T. Shreedhar, S. K. Kaul, and R. D. Yates, "ACP: Age control protocol for minimizing age of information over the Internet," in *Proc. ACM MobiCom*, Oct. 2018.
[72] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA: Athena Scientific, 2005, vol. 1.
[73] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
[74] I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *American Mathematical Society Translations*, vol. 12, pp. 199–246, 1959.
[75] P. J. Haas, *Stochastic Petri Nets: Modelling, Stability, Simulation*. New York, NY: Springer New York, 2002.
[76] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
[77] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in *Proc. ACM MobiHoc*, July 2019.
[78] S. M. Ross, *Stochastic Processes*, 2nd ed. John Wiley & Sons, 1996.
[79] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
[80] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.
[81] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univerisity Press, 2004.
[82] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
[83] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*. Norwell, MA: Kluwer Academic Publisher, 2000.
[84] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Univeristy Press, 2010.

**Yin Sun** received his B.Eng. and Ph.D. degrees in Electronic Engineering from Tsinghua University, in 2006 and 2011, respectively. At Tsinghua, he received the Excellent Doctoral Thesis Award of Tsinghua University, among many awards and scholarships. He was a postdoctoral scholar and research associate at the Ohio State University during 2011-2017. Since fall 2017, he joined Auburn University as an assistant professor in the Department of Electrical and Computer Engineering. His research interests include wireless communications, communication networks, information freshness, information theory, and machine learning. He is the founding co-chair of the first and second Age of Information Workshops, in conjunction with the IEEE INFOCOM 2018 and 2019. His papers received the best student paper award at IEEE/IFIP WiOpt 2013 and the best paper award at IEEE/IFIP WiOpt 2019.

**Benjamin Cyr** received his B.E. in 2017 from the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA. He is pursuing a Ph.D. degree from the Computer Science and Engineering Department at the University of Michigan, Ann Arbor, MI, USA. His current research area is in cyber-physical security.