

Sampling for Data Freshness Optimization: Non-linear Age Functions

Yin Sun[†] and Benjamin Cyr[‡]

[†]Dept. of ECE, Auburn University, Auburn, AL

[‡]Dept. of EECS, University of Michigan, Ann Arbor, MI

Abstract—In this paper, we study how to take samples from an information source, where the samples are forwarded to a remote receiver through a queue. The optimal sampling problem for maximizing the freshness of received samples 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 determined by 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. This solution is optimal under quite general conditions, including (i) general data freshness metrics represented by monotonic functions of the age of information, (ii) general service time distributions of the queueing server, and (iii) both continuous-time and discrete-time sampling problems. Numerical results suggest that the optimal sampling policies can be much better than zero-wait sampling and the classic uniform sampling.

I. INTRODUCTION

Information usually has the greatest value when it is fresh [1, 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 [2], [3], the concept of *Age of Information* was introduced to measure the freshness of information that a receiver has about the status of a 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 $\Delta_t = t - U_t$, which is the time elapsed since the newest sample was generated. Hence, a small age Δ_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 decisions¹, 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 both the time-varying pattern of the source and how valuable the fresh data is in the usage context. In this spirit, non-linear age functions $u(\Delta_t)$ have been employed as data freshness metrics in [4]–[10], where $u(\Delta_t)$ represents the utility value of data with age Δ_t , temporal autocorrelation function of the source, or some application-specific performance metrics. Several usage cases of non-linear age functions are summarized in Section 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 the linear age function $a\Delta_t$, e.g., [3], [11]–[37]. However, the design of data update policies for optimizing non-linear age metrics remains largely unexplored.

In this paper, we study 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 subject to a sampling rate constraint is obtained. In particular, 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 Δ_t . Accordingly, the penalty for data staleness is denoted by a *non-decreasing* function $p(\Delta_t)$ of Δ_t . The sampler design problem for optimizing these data freshness metrics 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 to the sampling rate constraint. 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. This age optimality result holds for (i) general monotonic age metrics, (ii) general service time distributions of the queueing server, and (iii) both continuous-time and discrete-time sampling. Among the technical tools used to prove these results are an extension of Dinkelbach’s method for MDP, and a geometric multiplier

Benjamin Cyr participated in this study when he was an undergraduate student at Auburn University. This work was supported in part by NSF grant CCF-1813050 and ONR grant N00014-17-1-2417.

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

technique for establishing strong duality. These technical tools were recently used in [38], [39], where a quite different sampling problem was solved.

- When there is no sampling rate constraint, a logical sampling policy is the zero-wait sampling policy [3], [9], [15], 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. This paper is an extended version of [40].

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 [2], [41]–[43]. Queueing theoretic techniques were introduced to evaluate the age of information in [3]. The average age, average peak age, and age distribution have been analyzed for various queueing systems in, e.g., [3], [11]–[13], [44]–[47]. It was observed that a Last Come, First Served (LCFS) transmission 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 [48]. This age optimality result has been generalized to queueing systems with multiple servers, multiple hops, and/or multiple sources [48]–[52].

When the transmission power of the source is subject to an energy-harvesting constraint, the age of information was minimized in, e.g., [9], [14]–[21]. Source coding and channel coding schemes for reducing the age were developed in, e.g., [22]–[25]. Age-optimal transmission scheduling of wireless networks have been investigated in, e.g., [26]–[33], [53], [54]. Game theoretical perspective of the age was studied in [34], [35], [55], [56]. The aging effect of channel state information was analyzed in, e.g., [57]–[59]. The relationship between the age of information and signal sampling was investigated in [38], [39]. The impact of the age to control systems was studied in [36], [60]. Emulations and measurements of the age of information was conducted in [37], [61], [62]. An age-based transport protocol was developed in [63].

The most relevant studies to this paper are [9], [24], [33]. The optimal sampling policies developed in this paper are simpler than those in [9], but the data freshness metrics considered in this paper are more general than those of [9]. In [24], [33], optimal sampling policies were developed to

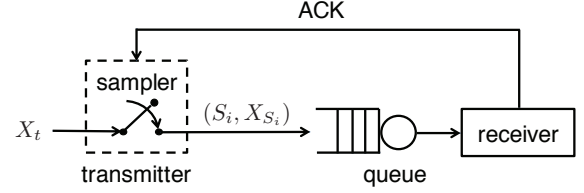


Fig. 1: System model.

minimize the time-average age for status updates over wireless channels, where the optimal sampling policies were shown to be threshold policies. The linear age function considered in [24], [33] is a special case of the monotonic age functions considered in this paper, and the channel models in [24], [33] are different from ours. In addition, we have found that the optimal threshold is equal to the optimum objective value of the MDP plus the optimal Lagrangian dual variable; see Section V-E for the details. As a result, the optimal threshold can be computed by bisection search, and the value iteration or policy iteration algorithms used for solving the MDPs in [24], [33] are not needed. Furthermore, our solution is optimal even if the MDP has a countable or uncountable state space. However, because the channel model of this paper is different from those of [24], [33], it is unclear whether our solution techniques can be applied to the MDPs in [24], [33].

III. MODEL, METRICS, AND FORMULATION

A. System Model

We consider the data 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 *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 $S_i \leq S_{i+1}$, $S_i + Y_i \leq D_i$, $D_i + Y_{i+1} \leq D_{i+1}$, and $0 < \mathbb{E}[Y_i] < \infty$ 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 $U_t = \max\{S_i : D_i \leq t\}$ be the time stamp of the freshest sample that the receiver has received by time t . Then, the *age of information*, or simply *age*, at time t is defined by [2], [3]

$$\Delta_t = t - U_t = t - \max\{S_i : D_i \leq t\}, \quad (1)$$

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

$$\Delta_t = t - S_i, \text{ if } D_i \leq t < D_{i+1}. \quad (2)$$

The initial state of the system is assumed to be $S_0 = 0$, $D_0 = Y_0$, and Δ_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, $t \in \{0, T_s, 2T_s, \dots\}$ is a multiple of period T_s ; as

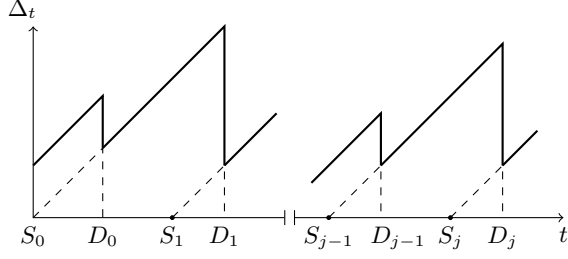


Fig. 2: Evolution of the age Δ_t over time.

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

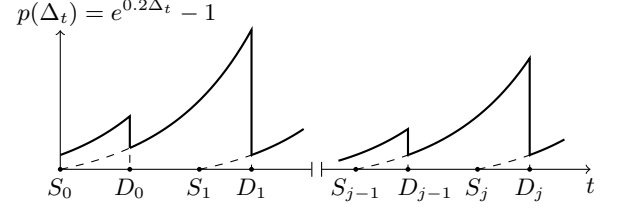
We consider a class of data staleness metrics: The dissatisfaction for data staleness (or the eagerness for data refreshing) is represented by a penalty function $p(\Delta)$ of the age Δ , where the function $p : [0, \infty) \mapsto \mathbb{R}$ is *non-decreasing*. This non-decreasing requirement on $p(\Delta)$ complies with the observations that stale data is usually less desired than fresh data [1], [4]–[8]. We further assume that $\mathbb{E}[p(\Delta + Y_i)] < \infty$ for all finite Δ . 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 [9], 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 Δ [5], [7]. One simple choice is $u(\Delta) = -p(\Delta)$. Note that because the age Δ_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. In the sequel, we discuss several usage cases of $p(\cdot)$ and $u(\cdot)$.

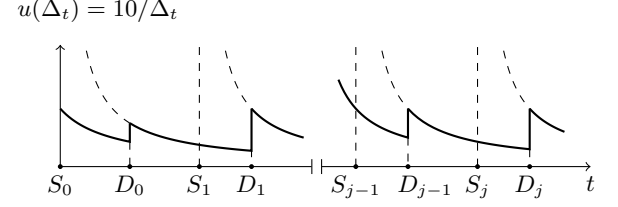
1) Temporal Auto-correlation Function of the Source:

In [10], the auto-correlation function $\mathbb{E}[X_t^* X_{t-\Delta_t}]$ is used to evaluate the freshness of the sample $X_{t-\Delta_t}$. For some stationary sources, $\mathbb{E}[X_t^* X_{t-\Delta_t}]$ is a non-increasing function of the age Δ_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 Δ_t is small, the auto-correlation function and the data rate both decay with respect to the age Δ_t [57].

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



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



(b) Age utility function $u(\Delta_t) = 10/\Delta_t$.

Fig. 3: Two examples of age penalty and utility functions.

real-time source value. If the sampling times S_i are independent of the observed source $\{X_t, t \geq 0\}$, then $\{\Delta_t, t \geq 0\}$ is independent of $\{X_t, t \geq 0\}$. In this case, the estimation error at time t can be expressed as an age penalty function $p(\Delta_t)$. For example, if the source is a Wiener process, the estimation error is exactly the age Δ_t [13], [38], [39]. If the sampling times S_i are chosen based on causal knowledge about the source, the estimation error is no longer a function of the age [38], [39].

3) Information based Data Freshness Metric: Let

$$\mathbf{W}_t = \{X_{S_i} : D_i \leq t\} \quad (3)$$

denote the samples that have been delivered to the receiver by time t . One can use the mutual information $I(X_t; \mathbf{W}_t)$ — the amount of information that the received samples \mathbf{W}_t carry about the current source value X_t — to evaluate the freshness of \mathbf{W}_t .² If $I(X_t; \mathbf{W}_t)$ is close to $H(X_t)$, the samples \mathbf{W}_t contains a lot of information about X_t and is considered to be fresh; if $I(X_t; \mathbf{W}_t)$ is almost 0, \mathbf{W}_t provides little information about X_t and is deemed to be obsolete.

One way to interpret $I(X_t; \mathbf{W}_t)$ is to consider how helpful the received samples \mathbf{W}_t are for inferring X_t . By using the Shannon code lengths [64, Section 5.4], the expected minimum number of bits L required to specify X_t satisfies

$$H(X_t) \leq L < H(X_t) + 1, \quad (4)$$

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 \mathbf{W}_t , the expected minimum number of bits L' that are required to specify X_t satisfies

$$H(X_t | \mathbf{W}_t) \leq L' < H(X_t | \mathbf{W}_t) + 1. \quad (5)$$

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 (4) and (5) is insignificant. Hence, $I(X_t; \mathbf{W}_t)$

²In this metric, the knowledge implied by the sampling times $\{S_i : D_i \leq t\}$ is neglected. One interesting future research direction is to investigate how to exploit timing information to improve data freshness.

is approximately the reduction in the description cost for inferring X_t without and with the knowledge of \mathbf{W}_t .

If X_t is a stationary, time-homogeneous Markov chain, by data processing inequality [64, Theorem 2.8.1], it is easy to prove the following lemma:

Lemma 1. *If X_t is a stationary, time-homogeneous Markov chain, \mathbf{W}_t is defined in (3), and $\{\Delta_t, t \geq 0\}$ is independent of $\{X_t, t \geq 0\}$, then the mutual information*

$$I(X_t; \mathbf{W}_t) = I(X_t; X_{t-\Delta_t}) \quad (6)$$

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

Proof. See Appendix A. \square

Lemma 1 provides an intuitive interpretation of “information aging”: The amount of information $I(X_t; \mathbf{W}_t)$ that is preserved in \mathbf{W}_t for inferring the current source value X_t decreases as the age Δ_t grows. Next, we provide the closed-form expression of $I(X_t; \mathbf{W}_t)$ for two Markov sources:

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

$$X_t = aX_{t-1} + V_t, \quad (7)$$

where $a \in (-1, 1)$ and the V_t 's are zero-mean *i.i.d.* Gaussian random variables with variance σ^2 . Because X_t is a Gauss-Markov process, one can show that [65]

$$I(X_t; \mathbf{W}_t) = I(X_t; X_{n-\Delta_t}) = -\frac{1}{2} \log_2 (1 - a^{2\Delta_t}). \quad (8)$$

Since $a \in (-1, 1)$ and $\Delta_t \geq 0$ is an integer, $I(X_t; \mathbf{W}_t)$ is a positive and decreasing function of the age Δ_t . Note that if $\Delta_t = 0$, then $I(X_t; \mathbf{W}_t) = H(X_t) = \infty$, 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

$$X_t = X_{t-1} \oplus V_t, \quad (9)$$

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

$$I(X_t; \mathbf{W}_t) = I(X_t; X_{n-\Delta_t}) = 1 - h\left(\frac{1 - (1 - 2q)^{\Delta_t}}{2}\right), \quad (10)$$

where $\Pr[X_t = 1 | X_0 = 0] = \frac{1 - (1 - 2q)^n}{2}$ 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 $x \in [0, 1]$ [64, Eq. (2.5)]. Because $h(x)$ is increasing on $[0, \frac{1}{2}]$, $I(X_t; \mathbf{W}_t)$ is a non-negative and decreasing function of the age Δ_t .

Furthermore, if the Markov chain X_t is non-stationary, $H(X_t)$ changes over time. In this case, one can use the conditional entropy $H(X_t | \mathbf{W}_t)$ to represent the staleness of \mathbf{W}_t [66]–[68]. If the sampling times S_i are independent of $\{X_t, t \geq 0\}$, $H(X_t | \mathbf{W}_t)$ is a non-decreasing function $p(\Delta_t)$ of the age Δ_t . However, if S_i is determined based on causal knowledge of X_t , $H(X_t | \mathbf{W}_t)$ is not a function of the age.

Additional usage cases of $p(\cdot)$ and $u(\cdot)$ can be found in [4]–[8]. Other data staleness and freshness metrics that cannot

be expressed as functions of Δ_t were discussed in [44], [48]–[52].

C. Formulation of Optimal Sampling Problems

Let $\pi = (S_1, S_2, \dots)$ represent a sampling policy and Π 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 $\{T_i = S_{i+1} - S_i, i = 1, 2, \dots\}$ form a regenerative process [69, Section 6.1]³: There exists an increasing sequence $0 \leq k_1 < k_2 < \dots$ of almost surely finite random integers such that the post- k_j process $\{T_{k_j+i}, i = 1, 2, \dots\}$ has the same distribution as the post- k_1 process $\{T_{k_1+i}, i = 1, 2, \dots\}$ and is independent of the pre- k_j process $\{T_i, i = 1, 2, \dots, k_j - 1\}$; in addition, $0 < \mathbb{E}[S_{k_{j+1}} - S_{k_j}] < \infty$, $j = 1, 2, \dots$. 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.

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:

$$\bar{p}_{\text{opt},1} = \inf_{\pi \in \Pi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\int_0^T p(\Delta_t) dt \right] \quad (11)$$

$$\text{s.t.} \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[S_n] \geq \frac{1}{f_{\max}}, \quad (12)$$

where $\bar{p}_{\text{opt},1}$ is the optimal value of (11) and f_{\max} is the maximum allowed sampling rate. In the discrete-time case, we need to solve the following optimal sampling problem:

$$\bar{p}_{\text{opt},2} = \inf_{\pi \in \Pi} \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\sum_{t=1}^n p(\Delta_t) \right] \quad (13)$$

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

where $\bar{p}_{\text{opt},2}$ is the optimal value of (13). 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 (11) and (13) by choosing $p(\Delta) = -u(\Delta)$. In practice, the cost for data updates increases with the average sampling rate. Therefore, Problems (11) and (13) represent a tradeoff between data staleness (freshness) and update cost.

Problems (11) and (13) are constrained MDPs, one with a continuous (uncountable) state space and the other with a countable state space. Because of the *curse of dimensionality* [70], it is quite rare that one can explicitly solve such problems and derive analytical or closed-form solutions that are arbitrarily accurate.

IV. MAIN RESULTS: OPTIMAL SAMPLING POLICIES

In this section, we present a complete characterization of the solutions to (11) and (13). Specifically, the optimal sam-

³We assume that T_i is a regenerative process because we will optimize $\limsup_{T \rightarrow \infty} \mathbb{E}[\int_0^T p(\Delta_t) dt]/T$, but operationally a nicer objective function is $\limsup_{i \rightarrow \infty} \mathbb{E}[\int_0^{D_i} p(\Delta_t) dt]/\mathbb{E}[D_i]$. These two objective functions are equivalent if $\{T_1, T_2, \dots\}$ is a regenerative process.

pling policies are either deterministic or randomized threshold policies, depending on the scenario under consideration. Efficiently computation algorithms of the thresholds and the randomization probabilities are provided. The proofs are relegated to Section V.

A. Continuous-time Sampling without Rate Constraint

We first consider the continuous-time sampling problem (11). When there is no sampling rate constraint (i.e., $f_{\max} = \infty$), an solution to (11) 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 $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \dots)$ is an optimal solution to (11), where*

$$S_{i+1}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta\}, \quad (15)$$

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

$$\beta = \frac{\mathbb{E} \left[\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta_t) dt \right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \quad (16)$$

Further, β is exactly the optimal value to (11), i.e., $\beta = \bar{p}_{\text{opt},1}$.

The optimal sampling policy in (15)-(16) 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 $\mathbb{E}[p(\Delta_{t+Y_{i+1}})]$ has grown to be no smaller than a pre-determined threshold β . 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, β is equal to the optimum objective value $\bar{p}_{\text{opt},1}$ of (11). Hence, (15)-(16) requires 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 of (16). Because the Y_i 's are i.i.d., the expectations in (16) are functions of β and are irrelevant of i . Given β , these expectations can be evaluated by Monte Carlo simulations or importance sampling. Define

$$v(s) = \int_0^s p(t) dt, \quad (17)$$

then

$$\int_{D_i(\beta)}^{D_{i+1}(\beta)} p(\Delta_t) dt = v(D_{i+1}(\beta) - S_i(\beta)) - v(Y_i). \quad (18)$$

If the analytical expression of $v(\cdot)$ is available, then (18) can be used to simplify the numerical evaluation of the expected integral in (16). As shown in Section V-F, (16) has a unique solution. We use a simple bisection method to solve (16), which is illustrated in Algorithm 1.

1) *Optimality Condition of Zero-wait Sampling:* When $f_{\max} = \infty$, one logical sampling policy is the zero-wait sampling policy [3], [9], [15], given by

$$S_{i+1} = S_i + Y_i. \quad (19)$$

This zero-wait sampling policy achieves the maximum

Algorithm 1 Bisection method for solving (16)

given l, u , tolerance $\epsilon > 0$.

repeat

$$\beta := (l + u)/2.$$

$$o := \beta - \frac{\mathbb{E}[v(D_{i+1}(\beta) - S_i(\beta)) - v(Y_i)]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}.$$

if $o \geq 0$, $u := \beta$; **else**, $l := \beta$.

until $u - l \leq \epsilon$.

return β .

throughput and the minimum queueing delay. In the special case of $p(t) = t$, Theorem 5 of [9] provided a sufficient and necessary condition for characterizing the optimality of the zero-wait sampling policy. We now generalize that result to 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 $0 < \mathbb{E}[Y_i] < \infty$, then the zero-wait sampling policy in (19) is optimal for solving (11) if and only if*

$$\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] \geq \frac{\mathbb{E} \left[\int_{Y_i}^{Y_i + Y_{i+1}} p(t) dt \right]}{\mathbb{E}[Y_{i+1}]}, \quad (20)$$

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

One can consider $\text{ess inf } Y_i$ 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 $0 < \mathbb{E}[Y_i] < \infty$, then the following assertions are true:*

- (a). *If Y_i is a constant, then (19) is optimal for solving (11).*
- (b). *If $\text{ess inf } Y_i = 0$ and $p(\cdot)$ is strictly increasing, then (19) is **not** optimal for solving (11).*

The condition $\text{ess inf } Y_i = 0$ 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 (19) is not optimal for these commonly used distributions.

B. Continuous-time Sampling with Rate Constraint

When the sampling rate constraint (12) is imposed, an solution to (11) is presented in the following theorem:

Theorem 2 (Continuous-time Sampling with Rate Constraint). *If $p(\cdot)$ is non-decreasing, $\mathbb{E}[p(t + Y_i)] < \infty$ for all finite t , and the service times Y_i are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then (15)-(16) is an optimal solution to (11), if*

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}. \quad (21)$$

Otherwise, $(S_1(\beta), S_2(\beta), \dots)$ is an optimal solution to (11), where

$$S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) & \text{with probability } \lambda, \\ T_{i,\max}(\beta) & \text{with probability } 1 - \lambda, \end{cases} \quad (22)$$

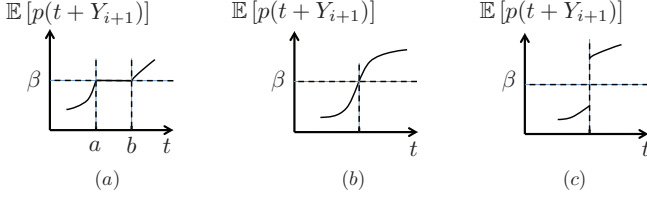


Fig. 4: Three cases of function $f(t) = \mathbb{E}[p(t + Y_{i+1})]$.

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

$$T_{i,\min}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta\}, \quad (23)$$

$$T_{i,\max}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta_{t+Y_{i+1}})] > \beta\}, \quad (24)$$

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

$$\mathbb{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)], \quad (25)$$

and λ is given by⁴

$$\lambda = \frac{\mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathbb{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}. \quad (26)$$

According to Theorem 2, the solution to (11) consists of two cases: In *Case 1*, the deterministic threshold policy in Theorem 1 is an optimal solution to (11), which needs to satisfy (21). In *Case 2*, the randomized threshold policy in (22)-(26) is an optimal solution to (11), which needs to satisfy

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] = \frac{1}{f_{\max}}. \quad (27)$$

We note that the only difference between (23) and (24) is that “ \geq ” is used in (23) while “ $>$ ” is employed in (24).⁵ If there exists a time-interval $[a, b]$ such that

$$\mathbb{E}[p(t + Y_{i+1})] = \beta \text{ for all } t \in [a, b], \quad (28)$$

as shown in Fig. 4(a), then $T_{i,\min}(\beta) < T_{i,\max}(\beta)$. In this case, $S_{i+1}(\beta) = T_{i,\min}(\beta)$ or $S_{i+1}(\beta) = T_{i,\max}(\beta)$ may not satisfy (27), but their randomized mixture in (22) can satisfy (27). In particular, if β and λ are given by (25) and (26), then (27) follows.

We provide a low-complexity algorithm to compute the randomized threshold policy in (22)-(26): As shown in Appendix C, there is a unique β satisfying (25). We use the bisection method in Algorithm 2 to solve (25) and obtain β . After that, $S_{i+1}(\beta)$ and λ can be computed by substituting β into (22)-(24) and (26). Because of the similarity between (23) and (24), $S_{i+1}(\beta)$ and λ are quite sensitive to the numerical error in β . This issue can be resolved by replacing $T_{i,\min}(\beta)$ in (22) and (26) with $T'_{i,\min}(\beta)$ and replacing $T_{i,\max}(\beta)$ in (22) and (26) with $T'_{i,\max}(\beta)$, where $T'_{i,\min}(\beta)$ and $T'_{i,\max}(\beta)$ are determined by

⁴If $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely, then (22) becomes a deterministic threshold policy and λ can be any number within $[0, 1]$.

⁵Clearly, an important issue is the optimality of such a randomized threshold policy, which is proven in Section V.

Algorithm 2 Bisection method for solving (25)

given l, u , tolerance $\epsilon > 0$.

repeat

$\beta := (l + u)/2$.

$o_1 := \mathbb{E}[T_{i,\min}(\beta) - S_i(\beta)]$.

$o_2 := \mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)]$.

if $o_1 > \frac{1}{f_{\max}}$, $u := \beta$;

else if $o_2 < \frac{1}{f_{\max}}$, $l := \beta$;

else return β .

until $u - l \leq \epsilon$.

return β .

$$T'_{i,\min}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta - \epsilon/2\}, \quad (29)$$

$$T'_{i,\max}(\beta) = \inf\{t \geq D_i(\beta) : \mathbb{E}[p(\Delta_{t+Y_{i+1}})] > \beta + \epsilon/2\}, \quad (30)$$

respectively, and $\epsilon > 0$ is the tolerance in Algorithm 2. One can improve the accuracy of this solution by (i) reducing the tolerance ϵ 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 Fig. 4(b)-(c), if $\mathbb{E}[p(t + Y_{i+1})]$ is strictly increasing on $t \in [0, \infty)$, then $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely and (22) 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 $\mathbb{E}[p(t + Y_{i+1})]$ is strictly increasing in t , then (15) is an optimal solution to (11), where $D_i(\beta) = S_i(\beta) + Y_i$, $\Delta_t = t - S_i(\beta)$, and β is determined by (16), if*

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}; \quad (31)$$

otherwise, β is determined by solving

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] = \frac{1}{f_{\max}}. \quad (32)$$

If $p(\cdot)$ is strictly increasing or the distribution of Y_i is sufficiently smooth, $\mathbb{E}[p(t + Y_{i+1})]$ 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 (11) was studied in [9], where $p(\cdot)$ was assumed to be positive and non-decreasing. There is an error in Theorem 3 of [9], because the condition “ $\mathbb{E}[p(t + Y_{i+1})]$ is strictly increasing in t ” is missing. Further, the solution in Theorem 3 of [9] is more complicated than that in Corollary 3. A special case of Corollary 3 with $p(t) = t$ was derived in Theorem 4 of [9].

C. Discrete-time Sampling

We now move on to the discrete-time sampling problem (13). When there is no sampling rate constraint (i.e., $f_{\max} = \infty$), the solution to (13) 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 $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \dots)$ is an optimal solution to (13), where

$$S_{i+1}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), \mathbb{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta\}, \quad (33)$$

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

$$\beta = \frac{\mathbb{E}\left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1} p(\Delta_t)\right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \quad (34)$$

Further, β is exactly the optimal value to (13), i.e., $\beta = \bar{p}_{\text{opt},2}$.

Theorem 3 is quite similar to Theorem 1, with two minor differences: (i) The sampling time $S_{i+1}(\beta)$ in (15) is a real number, which is restricted to an integer in (33). (ii) The integral in (16) becomes a summation in (34).

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

Corollary 4. If $f_{\max} = \infty$, $p(\cdot)$ is non-decreasing, and the service times Y_i are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then the zero-wait sampling policy (19) is optimal for solving (13) if and only if there exists $e < 1$ such that

$$\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1} + e)] \geq \frac{\mathbb{E}\left[\sum_{t=Y_i}^{Y_i+Y_{i+1}-1} p(t)\right]}{\mathbb{E}[Y_{i+1}]}, \quad (35)$$

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

When the sampling rate constraint (14) is imposed, the solution to (13) is provided in the following theorem.

Theorem 4 (Discrete-time Sampling with Rate Constraint). If $p(\cdot)$ is non-decreasing, $\mathbb{E}[p(t + Y_i)] < \infty$ for all finite t , and the service times Y_i are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then (33)-(34) is an optimal solution to (13), if

$$\mathbb{E}[S_{i+1}(\beta) - S_i(\beta)] > \frac{1}{f_{\max}}. \quad (36)$$

Otherwise, $(S_1(\beta), S_2(\beta), \dots)$ is an optimal solution to (13), where

$$S_{i+1}(\beta) = \begin{cases} T_{i,\min}(\beta) & \text{with probability } \lambda, \\ T_{i,\max}(\beta) & \text{with probability } 1 - \lambda, \end{cases} \quad (37)$$

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

$$T_{i,\min}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), \mathbb{E}[p(\Delta_{t+Y_{i+1}})] \geq \beta\}, \quad (38)$$

$$T_{i,\max}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), \mathbb{E}[p(\Delta_{t+Y_{i+1}})] > \beta\}, \quad (39)$$

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

$$\mathbb{E}[T_{i,\min}(\beta) - S_i(\beta)] \leq \frac{1}{f_{\max}} \leq \mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)], \quad (40)$$

and λ is given by

$$\lambda = \frac{\mathbb{E}[T_{i,\max}(\beta) - S_i(\beta)] - \frac{1}{f_{\max}}}{\mathbb{E}[T_{i,\max}(\beta) - T_{i,\min}(\beta)]}. \quad (41)$$

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 (23)-(24), which are restricted to integers in (38)-(39). (ii) If $\mathbb{E}[p(t + Y_{i+1})]$ is strictly increasing in t , then $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ holds almost surely in (23)-(24) and Theorem 2

can be greatly simplified. However, in the discrete-time case, even if $\mathbb{E}[p(t + Y_{i+1})]$ is strictly increasing in t , $T_{i,\min}(\beta) < T_{i,\max}(\beta)$ may still occur in (38)-(39). In fact, it is rather common that $T_{i,\min}(\beta) < T_{i,\max}(\beta)$ holds for the optimal β , because of the following reason:

If $T_{i,\min}(\beta) = T_{i,\max}(\beta)$ almost surely, then (37) becomes a deterministic threshold policy that needs to ensure (27). However, because $S_{i+1}(\beta)$ and $S_i(\beta)$ are integers, such a deterministic threshold policy is difficult to satisfy (27) for all possible values of f_{\max} . On the other hand, if $T_{i,\min}(\beta) < T_{i,\max}(\beta)$, the randomized threshold policy in (37)-(41) can satisfy (27). Hence, even though $\mathbb{E}[p(t + Y_{i+1})]$ 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 here.

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 W_t is formulated as

$$\bar{I}_{\text{opt}} = \sup_{\pi \in \Pi} \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\sum_{t=1}^n I(X_t; W_t) \right], \quad (42)$$

where \bar{I}_{opt} is the optimal value of (42). We assume that \bar{I}_{opt} is finite. Problem (42) is a special case of (13) satisfying $p(\Delta_t) = -u(\Delta_t) = -I(X_t; W_t)$ 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 $0 < \mathbb{E}[Y_i] < \infty$, then $(S_1(\beta), S_2(\beta), \dots)$ is an optimal solution to (13), where

$$S_{i+1}(\beta) = \min\{t \in \mathbb{N} : t \geq D_i(\beta), I(X_{t+Y_{i+1}}; X_{S_i(\beta)} | Y_{i+1}) \leq \beta\}, \quad (43)$$

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

$$\beta = \frac{\mathbb{E} \left[\sum_{t=D_i(\beta)}^{D_{i+1}(\beta)-1} I(X_t; X_{S_i(\beta)}) \right]}{\mathbb{E}[D_{i+1}(\beta) - D_i(\beta)]}. \quad (44)$$

Further, β is exactly the optimal value of (42), 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 (43), the $(i+1)$ -th sample is generated at the smallest

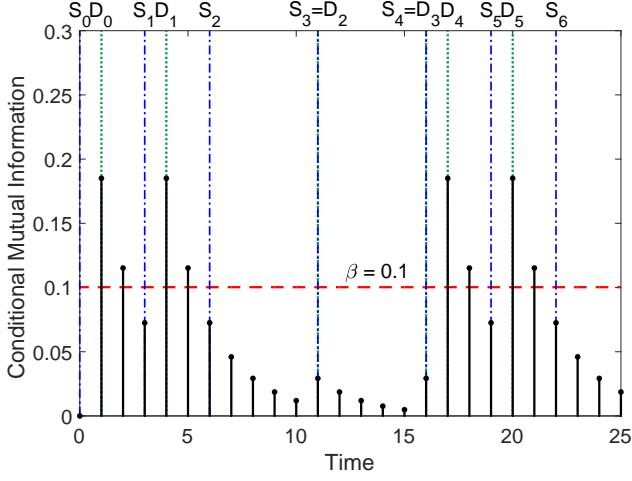


Fig. 5: A sample-path illustration of the optimal sampling policy (43) and (44), where $\beta = 0.1$, Y_i is either 1 or 5 with equal probability, S_i and D_i are sampling time and delivery time of the i -th sample. On this sample-path, the service times are $Y_0 = 1, Y_1 = 1, Y_2 = 5, Y_3 = 5, Y_4 = 1, Y_5 = 1, Y_6 = 5$.

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}_{opt} , i.e., the optimum of the time-average expected mutual information $\liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\sum_{t=1}^n I(X_t; \mathbf{W}_t)]$ 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 β and then take the next sample; if the service time of the previous sample is $Y_i = 5$, the next sample is taken once the previous sample is delivered, because the conditional mutual information is already below β then.

V. PROOFS OF THE MAIN RESULTS

In this section, we present the proofs of the main results in Section IV. We first provide 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 [9], 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 (11), it is sub-optimal to take a new sample before the previous sample is delivered.*

By Lemma 2, the queue in Figure 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.,

$$\Pi_1 = \{\pi \in \Pi : S_{i+1} \geq D_i = S_i + Y_i \text{ for all } i\}. \quad (45)$$

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 $S_i = S_0 + \sum_{j=0}^{i-1} (Y_j + Z_j) = \sum_{j=0}^{i-1} (Y_j + Z_j)$ and $D_i = S_i + Y_i$. Given (Y_1, Y_2, \dots) , (S_1, S_2, \dots) is uniquely determined by (Z_1, Z_2, \dots) . Hence, one can also use $\pi = (Z_1, Z_2, \dots)$ to represent a sampling policy in Π_1 .

Because T_i is a regenerative process, by using the renewal theory in [71] and [69, Section 6.1], one can show that $\frac{1}{i} \mathbb{E}[S_i]$ and $\frac{1}{i} \mathbb{E}[D_i]$ in (11) are convergent sequences and

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\int_0^T p(\Delta_t) dt \right] \\ &= \lim_{i \rightarrow \infty} \frac{\mathbb{E} \left[\int_0^{D_i} p(\Delta_t) dt \right]}{\mathbb{E}[D_i]} \\ &= \lim_{i \rightarrow \infty} \frac{\sum_{j=1}^i \mathbb{E} \left[\int_{D_j}^{D_{j+1}-1} p(\Delta_t) dt \right]}{\sum_{j=1}^i \mathbb{E}[Y_j + Z_j]}. \end{aligned} \quad (46)$$

On the other hand, it follows from (2) that

$$\begin{aligned} \int_{D_j}^{D_{j+1}} p(\Delta_t) dt &= \int_{D_j}^{D_{j+1}} p(t - S_j) dt \\ &= \int_{Y_j}^{Y_j + Z_j + Y_{j+1}} p(t) dt, \end{aligned} \quad (47)$$

which is a function of (Y_j, Z_j, Y_{j+1}) . Define

$$q(y_i, z, y') = \int_{y_i}^{y_i + z + y'} p(t) dt, \quad (48)$$

then Problem (11) can be rewritten as

$$\bar{p}_{\text{opt}_1} = \inf_{\pi \in \Pi_1} \lim_{i \rightarrow \infty} \frac{\sum_{j=1}^i \mathbb{E}[q(Y_j, Z_j, Y_{j+1})]}{\sum_{j=1}^i \mathbb{E}[Y_j + Z_j]} \quad (49)$$

$$\text{s.t. } \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{j=1}^i \mathbb{E}[Y_j + Z_j] \geq \frac{1}{f_{\max}}. \quad (50)$$

B. Reformulation of Problem (49)

In order to solve (49), we consider the following MDP with a parameter $c \geq 0$:

$$h(c) \triangleq \inf_{\pi \in \Pi_1} \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{j=0}^{i-1} \mathbb{E}[q(Y_j, Z_j, Y_{j+1}) - c(Y_j + Z_j)] \quad (51)$$

$$\text{s.t. } \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{j=1}^i \mathbb{E}[Y_j + Z_j] \geq \frac{1}{f_{\max}}, \quad (52)$$

where $h(c)$ is the optimum value of (51). Similar with Dinkelbach's method [72] for nonlinear fractional programming, the following lemma holds for the MDP (49):

Lemma 3. [39, Lemma 2] *The following assertions are true:*

- (a). $\bar{p}_{\text{opt}_1} \geq c$ if and only if $h(c) \geq 0$.
- (b). If $h(c) = 0$, the solutions to (49) and (51) are identical.

Hence, the solution to (49) can be obtained by solving (51) and seeking $\bar{p}_{\text{opt}_1} \in \mathbb{R}$ that satisfies

$$h(\bar{p}_{\text{opt}_1}) = 0. \quad (53)$$

C. Lagrangian Dual Problem of (51) when $c = \bar{p}_{\text{opt}_1}$

Although (51) 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 (51) and show that it admits no duality gap.

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

$$L(\pi; \alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E} [q(Y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Y_i + Z_i)] + \frac{\alpha}{f_{\max}}, \quad (54)$$

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

$$g(\alpha) \triangleq \inf_{\pi \in \Pi_1} L(\pi; \alpha). \quad (55)$$

Then, the Lagrangian dual problem of (51) is defined by

$$d \triangleq \max_{\alpha \geq 0} g(\alpha), \quad (56)$$

where d is the optimum value of (56). Weak duality [73], [74] implies that $d \leq h(\bar{p}_{\text{opt}_1})$. Next, we will solve (55) and establish strong duality, i.e., $d = h(\bar{p}_{\text{opt}_1})$.

D. Optimal Solutions to (55)

We solve (55) in two steps: First, we use a sufficient statistic argument to show that (55) 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 (55).*

Proof. In (55), the minimization of the term

$$\begin{aligned} & \mathbb{E} [q(Y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Y_i + Z_i)] \\ & \stackrel{(a)}{=} \mathbb{E} [q(Y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Z_i + Y_{i+1})] \end{aligned} \quad (57)$$

over Z_i depends on $(Y_1, \dots, Y_i, Z_1, \dots, Z_{i-1})$ via Y_i , where Step (a) is due to $\mathbb{E}[Y_i] = \mathbb{E}[Y_{i+1}]$. Hence, Y_i is a sufficient statistic for determining Z_i in (55). \square

By Lemma 4, (55) 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

$$\min_{Z_i \geq 0} \mathbb{E} [q(y_i, Z_i, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(Z_i + Y_{i+1})], \quad (58)$$

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 (58), let us consider the following problem

$$\min_{z \geq 0} \mathbb{E} [q(y_i, z, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(z + Y_{i+1})]. \quad (59)$$

Because $p(\cdot)$ is non-decreasing, the functions $z \rightarrow q(y_i, z, y')$ and $z \rightarrow \mathbb{E} [q(y_i, z, Y_{i+1})]$ are both convex. Hence, (59) is a convex optimization problem.

Lemma 5. *If $p(\cdot)$ is non-decreasing, then the set of optimal solution to (59) is $[z_{\min}(y_i, \alpha), z_{\max}(y_i, \alpha)]$ where*

$$z_{\min}(y, \alpha) = \inf \{t \geq 0 : \mathbb{E}[p(y + t + Y_{i+1})] \geq \bar{p}_{\text{opt}_1} + \alpha\}, \quad (60)$$

$$z_{\max}(y, \alpha) = \inf \{t \geq 0 : \mathbb{E}[p(y + t + Y_{i+1})] > \bar{p}_{\text{opt}_1} + \alpha\}. \quad (61)$$

Proof. See Appendix B. \square

By Lemma 5, z is an optimal solution to (59) if and only if $z \in [z_{\min}(y_i, \alpha), z_{\max}(y_i, \alpha)]$. Hence, the set of optimal solutions to (58) 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 $0 < \mathbb{E}[Y_i] < \infty$, then the set of optimal solutions to (55) is*

$$\Gamma(\alpha) = \{\pi : Z_i \in [z_{\min}(Y_i, \alpha), z_{\max}(Y_i, \alpha)] \text{ for almost all } i\}, \quad (62)$$

where $z_{\min}(y, \alpha)$ and $z_{\max}(y, \alpha)$ are given in (60)-(61).

E. Zero Duality Gap and Optimal Solution to (51)

Strong duality and an optimal solution to (51) are obtained in the following theorem:

Theorem 5. *If $c = \bar{p}_{\text{opt}_1}$, $p(\cdot)$ is non-decreasing, $\mathbb{E}[p(t + Y_i)] < \infty$ for all finite t , and the service times Y_i are i.i.d. with $0 < \mathbb{E}[Y_i] < \infty$, then the duality gap between (51) and (56) is zero. Further, $Z_i = z_{\min}(Y_i, 0)$ is an optimal solution to (51) and (56), if*

$$\mathbb{E}[Y_i + z_{\min}(Y_i, 0)] > \frac{1}{f_{\max}}. \quad (63)$$

Otherwise, (Z_1, Z_2, \dots) is an optimal solution to (51) and (56), where

$$Z_i = \begin{cases} z_{\min}(Y_i, \alpha) & \text{with probability } \lambda, \\ z_{\max}(Y_i, \alpha) & \text{with probability } 1 - \lambda, \end{cases} \quad (64)$$

$\alpha \geq 0$ is determined by solving

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

and λ is given by

$$\lambda = \frac{\mathbb{E}[Y_i + z_{\max}(Y_i, \alpha)] - \frac{1}{f_{\max}}}{\mathbb{E}[z_{\max}(Y_i, \alpha) - z_{\min}(Y_i, \alpha)]}. \quad (65)$$

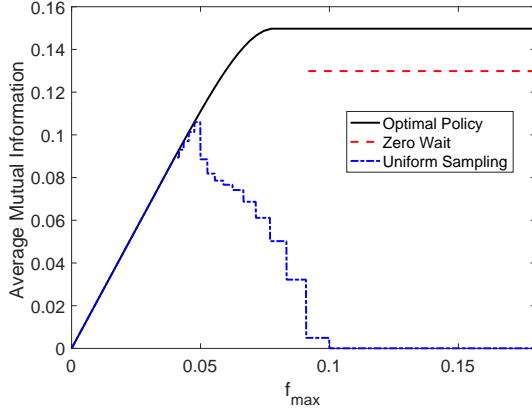


Fig. 6: Average mutual information of the Gauss-Markov source versus f_{\max} , where the service times Y_i are equal to either 1 or 21 with probability 0.5.

Proof. We use [73, Prop. 6.2.5] to find a *geometric multiplier* for (51). This suggests that the duality gap between (51) and (56) must be zero, because otherwise there exists no geometric multiplier [73, Prop. 6.2.3(b)]. The details are provided in Appendix C. \square

By choosing

$$\beta = \bar{p}_{\text{opt}_1} + \alpha, \quad (66)$$

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 (11) as the convex optimization problem in (59). These technical tools were also used in a recent study [39], where a quite different sampling problem is solved. Further, (66) implies that the optimal threshold β is equal to the optimum objective value of the MDP \bar{p}_{opt_1} plus the optimal Lagrangian dual variable α . By using these results, bisection search algorithms are developed in Section IV to compute β , 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 (11) is

$$\begin{aligned} \bar{p}_{\text{opt}_1} &= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbb{E}[q(Y_i, z_{\min}(Y_i, 0), Y_{i+1})]}{\sum_{i=1}^n \mathbb{E}[Y_i + z_{\min}(Y_i, 0)]} \\ &= \frac{\mathbb{E}[q(Y_i, z_{\min}(Y_i, 0), Y_{i+1})]}{\mathbb{E}[Y_i + z_{\min}(Y_i, 0)]} \\ &= \frac{\mathbb{E}\left[\int_{Y_i}^{Y_i + z_{\min}(Y_i, 0) + Y_{i+1}} p(t) dt\right]}{\mathbb{E}[Y_i + z_{\min}(Y_i, 0)]}, \end{aligned} \quad (67)$$

from which (16) follows. We note that the root of (16) must be unique; otherwise, one can follow the arguments in Appendix C to show that the optimal objective value to (11) is non-unique, which cannot be true. Further, as shown in Appendix C, the condition " $\mathbb{E}[p(t + Y_i)] < \infty$ for all finite t " is not needed in the case of Theorem 1.

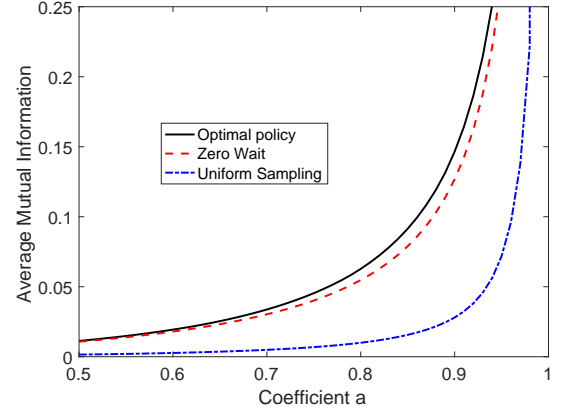


Fig. 7: Average mutual information of the Gauss-Markov source versus the coefficient a in (7), where the service times Y_i are equal to either 1 or 21 with probability 0.5.

The proof of Corollary 1 is provided in Appendix D. The proof of Corollary 2 is provided in Appendix E. The proof of Corollary 3 is omitted, because it follows immediately from Theorem 2.

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 the convex optimization problem (59) of the continuous-time case becomes the following integer optimization problem in the discrete-time case:

$$\min_{z \in \mathbb{N}} \mathbb{E}[q(y_i, z, Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)(z + Y_{i+1})], \quad (68)$$

where

$$q(y_i, z, y') = \sum_{t=y}^{y+z+y'-1} p(t). \quad (69)$$

By adopting an idea in [75, Problem 5.5.3], we obtain

Lemma 7. *If $p(\cdot)$ is non-decreasing, then the set of optimal solution to (68) is $\{z_{\min}(y_i, \alpha), z_{\min}(y_i, \alpha) + 1, z_{\min}(y_i, \alpha) + 2, \dots, z_{\max}(y_i, \alpha)\}$, where*

$$z_{\min}(y, \alpha) = \inf\{t \in \mathbb{N} : \mathbb{E}[p(y + t + Y_{i+1})] \geq \bar{p}_{\text{opt}_1} + \alpha\}, \quad (70)$$

$$z_{\max}(y, \alpha) = \inf\{t \in \mathbb{N} : \mathbb{E}[p(y + t + Y_{i+1})] > \bar{p}_{\text{opt}_1} + \alpha\}. \quad (71)$$

Proof. See Appendix F. \square

By replacing Lemma 5 with Lemma 7 and following the proof arguments in Section V.A-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 .

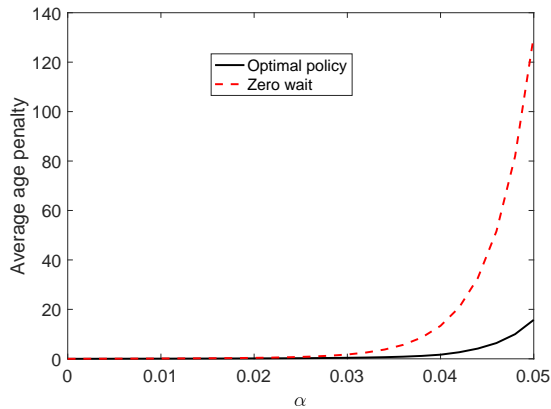


Fig. 8: Average age penalty of an exponential penalty function $p_{\text{exp}}(\Delta_t) = e^{\alpha\Delta_t} - 1$ versus the coefficient α , where the service times Y_i follow a discretized log-normal distribution.

- *Zero-wait*: The sampling policy in (19), which is infeasible when $f_{\max} < 1/\mathbb{E}[Y_i]$.
- *Optimal policy*: The sampling policy given by Theorem 2 for continuous-time sampling, or Theorem 4 for discrete-time sampling.

Figure 6 depicts the tradeoff between the time-average expected mutual information of the Gauss-Markov source in (7) and f_{\max} , where the mutual information is given by (8) with $a = 0.9$, and the service times Y_i are equal to either 1 or 21 with probability 0.5. Hence, $\mathbb{E}[Y_i] = 11$ and the zero-wait sampling policy is infeasible when $f_{\max} < 1/11$. One can observe that the average mutual information of the optimal sampling policy is higher than that of zero-waiting sampling and uniform sampling. When f_{\max} is large, the queueing 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} . In Fig. 7, we plot the time-average expected mutual information of the Gauss-Markov source versus the coefficient a in (7), where $f_{\max} = 0.095$ and Y_i is equal to either 1 or 21 with probability 0.5. We find that the average mutual information achieved by all three policies increases with a .

Figure 8 illustrates the time-average expectation of an exponential penalty function $p_{\text{exp}}(\Delta_t) = e^{\alpha\Delta_t} - 1$ versus the coefficient α , where Y_i follows a discretized log-normal distribution. In particular, Y_i can be expressed as $Y_i = \lceil e^{\sigma X_i} / \mathbb{E}[e^{\sigma X_i}] \rceil$, where the X_i 's are *i.i.d.* Gaussian random variables with zero mean and unit variance, and $\sigma = 1.5$. Figure 9 shows the time-average expectation of $p_{\text{exp}}(\Delta_t)$ versus the coefficient σ of the discretized log-normal service time distribution. One can observe that zero-wait sampling is far from optimal when α is relative large or the service times Y_i are highly random. Numerical results for continuous-time sampling can be found in our earlier work [9].

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

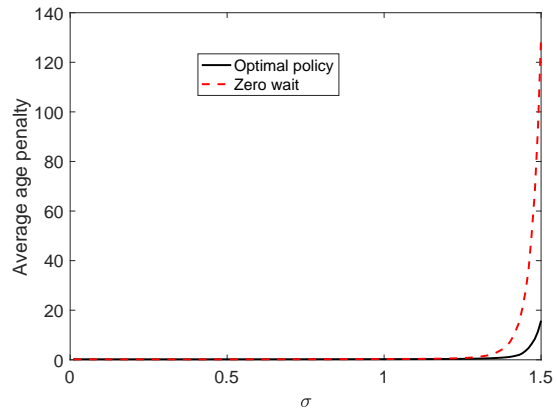


Fig. 9: Average age penalty of an exponential penalty function $p_{\text{exp}}(\Delta_t) = e^{\alpha\Delta_t} - 1$ versus the coefficient σ of discretized log-normal service time distribution.

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.

REFERENCES

- [1] C. Shapiro and H. Varian, *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press, 1999.
- [2] X. Song and J. W. S. Liu, "Performance of multiversion concurrency control algorithms in maintaining temporal consistency," in *Fourteenth Annual International Computer Software and Applications Conference*, Oct 1990, pp. 132–139.
- [3] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM*, 2012.
- [4] 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.
- [5] A. Even and G. Shankaranarayanan, "Utility-driven assessment of data quality," *SIGMIS Database*, vol. 38, no. 2, pp. 75–93, May 2007.
- [6] 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, 2009.
- [7] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *IEEE INFOCOM*, 2009.
- [8] S. Razniewski, "Optimizing update frequencies for decaying information," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 1191–1200.
- [9] 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.
- [10] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *IEEE ISIT*, June 2017, pp. 326–330.
- [11] 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, March 2016.
- [12] 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.
- [13] 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.
- [14] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Information Theory and Applications Workshop (ITA)*, 2015.

- [15] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *IEEE ISIT*, 2015.
- [16] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *IEEE ISIT*, 2017.
- [17] A. Arafat and S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *IEEE GLOBECOM*, Dec 2017, pp. 1–6.
- [18] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. and Netw.*, vol. 2, no. 1, pp. 193–204, March 2018.
- [19] A. Arafat, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," 2018, <https://arxiv.org/abs/1806.07271>.
- [20] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: With and without feedback," 2018, <https://arxiv.org/abs/1808.05141>.
- [21] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu, "Achieving the age-energy tradeoff with a finite-battery energy harvesting source," in *IEEE ISIT*, June 2018, pp. 876–880.
- [22] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *Data Compression Conference (DCC)*, March 2016.
- [23] R. D. Yates, E. Najm, E. Soljanin, and J. Zhong, "Timely updates over an erasure channel," in *IEEE ISIT*, June 2017, pp. 316–320.
- [24] E. T. Ceran, D. Gunduz, and A. Gyorgy, "Average age of information with hybrid ARQ under a resource constraint," in *IEEE WCNC*, 2018.
- [25] P. Mayekar, P. Parag, and H. Tyagi, "Optimal lossless source codes for timely updates," in *IEEE ISIT*, June 2018, pp. 1246–1250.
- [26] 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.
- [27] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Allerton*, Sept 2016, pp. 844–851.
- [28] 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.
- [29] Y. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *IEEE ISIT*, June 2017, pp. 561–565.
- [30] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *IEEE INFOCOM*, April 2018, pp. 1844–1852.
- [31] N. Lu, B. Ji, and B. Li, "Age-based scheduling: Improving data freshness for wireless real-time traffic," in *ACM MobiHoc*, 2018.
- [32] Z. Jiang, B. Krishnamachari, X. Zheng, S. Zhou, and Z. Niu, "Decentralized status update for age-of-information optimization in wireless multiaccess channels," in *IEEE ISIT*, June 2018, pp. 2276–2280.
- [33] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things," 2018, <https://arxiv.org/abs/1807.04356>.
- [34] Y. Xiao and Y. Sun, "A dynamic jamming game for real-time status updates," in *IEEE INFOCOM Aol Workshop*, April 2018, pp. 354–360.
- [35] S. Gopal and S. K. Kaul, "A game theoretic approach to DSRC and WiFi coexistence," in *IEEE INFOCOM Aol Workshop*, April 2018, pp. 565–570.
- [36] T. Soleymani, J. S. Baras, and K. H. Johansson, "Stochastic control with stale information—part i: Fully observable systems," 2018, <https://arxiv.org/abs/1810.10983>.
- [37] 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 *IEEE BlackSeaCom*, June 2018, pp. 1–5.
- [38] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *IEEE ISIT*, 2017.
- [39] —, "Sampling of the Wiener process for remote estimation over a channel with random delay," 2017, <https://arxiv.org/abs/1707.02531>.
- [40] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *IEEE SPAWC Workshop*, 2018.
- [41] A. Segev and W. Fang, "Optimal update policies for distributed materialized views," *Manage. Sci.*, vol. 37, no. 7, pp. 851–870, Jul. 1991.
- [42] B. Adelberg, H. Garcia-Molina, and B. Kao, "Applying update streams in a soft real-time database system," in *Proc. ACM SIGMOD*, 1995, pp. 245–256.
- [43] J. Cho and H. Garcia-Molina, "Synchronizing a database to improve freshness," in *Proc. ACM SIGMOD*, 2000, pp. 117–128.
- [44] 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, April 2016.
- [45] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *IEEE ISIT*, 2015.
- [46] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "The stationary distribution of the age of information in FCFS single-server queues," in *IEEE ISIT*, June 2017, pp. 571–575.
- [47] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," 2018, <https://arxiv.org/abs/1806.03487>.
- [48] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *IEEE ISIT*, 2016.
- [49] —, "Age-optimal information updates in multihop networks," in *IEEE ISIT*, 2017.
- [50] —, "Minimizing the age of information through queues," submitted to *IEEE Trans. Inf. Theory*, 2017, <http://arxiv.org/abs/1709.04956>.
- [51] —, "The age of information in multihop networks," submitted to *ACM/IEEE Trans. Netw.*, 2018, <https://arxiv.org/abs/1712.10061>.
- [52] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *IEEE INFOCOM Aol Workshop*, 2018.
- [53] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Allerton*, Oct 2017, pp. 486–493.
- [54] —, "Optimizing information freshness in wireless networks under general interference constraints," in *ACM MobiHoc*, 2018.
- [55] G. D. Nguyen, S. Kompella, C. Kam, J. E. Wieselthier, and A. Ephremides, "Impact of hostile interference on information freshness: A game approach," in *WiOpt*, May 2017, pp. 1–7.
- [56] —, "Information freshness over an interference channel: A game theoretic view," in *IEEE INFOCOM*, April 2018, pp. 908–916.
- [57] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug 2013.
- [58] M. Costa, S. Valentin, and A. Ephremides, "On the age of channel state information for non-reciprocal wireless links," in *IEEE ISIT*, June 2015, pp. 2356–2360.
- [59] 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 *IEEE ICASSP*, March 2016, pp. 3621–3625.
- [60] J. Zhang and C. Wang, "On the rate-cost of Gaussian linear control systems with random communication delays," in *IEEE ISIT*, June 2018, pp. 2441–2445.
- [61] C. Kam, S. Kompella, and A. Ephremides, "Experimental evaluation of the age of information via emulation," in *IEEE MILCOM*, Oct. 2015, pp. 1070–1075.
- [62] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Modeling the age of information in emulated ad hoc networks," in *IEEE MILCOM*, Oct. 2017, pp. 436–441.
- [63] T. Shreedhar, S. K. Kaul, and R. D. Yates, "ACP: Age control protocol for minimizing age of information over the Internet," in *ACM MobiCom*, 2018, pp. 699–701.
- [64] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
- [65] 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.
- [66] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal self-driven sampling for estimation based on value of information," in *Proceedings of the 13th International Workshop on Discrete Event Systems (WODES)*, 2016.
- [67] —, "Maximization of information in energy-limited directed communication," in *European Control Conference (ECC)*, 2016.
- [68] —, "Optimal stationary self-triggered sampling for estimation," in *IEEE CDC*, 2016.
- [69] P. J. Haas, *Stochastic Petri Nets: Modelling, Stability, Simulation*. New York, NY: Springer New York, 2002.
- [70] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
- [71] S. M. Ross, *Stochastic Processes*, 2nd ed. John Wiley & Sons, 1996.
- [72] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [73] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.
- [74] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [75] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

- [76] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*. Norwell, MA: Kluwer Academic Publisher, 2000.
- [77] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Univeristy Press, 2010.

APPENDIX A PROOF OF LEMMA 1

If $\{\Delta_t, t \geq 0\}$ is independent of $\{X_t, t \geq 0\}$, because X_t is a Markov chain, $X_{\max\{S_i; D_i \leq t\}} = X_{t-\Delta_t}$ contains all the information in $\mathbf{W}_t = \{X_{S_i} : D_i \leq t\}$ about X_t . In other words, $X_{t-\Delta_t}$ is a sufficient statistic of \mathbf{W}_t for inferring X_t . Then, (6) follows from [64, Eq. (2.124)].

Next, because X_t is stationary and time-homogeneous, $I(X_t; X_{t-\Delta}) = I(X_\Delta; X_0)$ for all t , which is a function of Δ . Further, because X_t is a Markov chain, owing to the data processing inequality [64, Theorem 2.8.1], $I(X_\Delta; X_0)$ is non-increasing in Δ . 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

$$\delta h(z; w) \triangleq \lim_{\epsilon \rightarrow 0^+} \frac{h(z + \epsilon w) - h(z)}{\epsilon}. \quad (72)$$

Because the function $h(z) = \mathbb{E}[q(y_i, z, Y_{i+1})]$ is convex, the one-sided derivative $\delta h(z; w)$ of $h(z)$ exist [75, p. 709]. Because $z \rightarrow q(y_i, z, y')$ is convex, the function $\epsilon \rightarrow [q(y_i, z + \epsilon w, y') - q(y_i, z, y')]/\epsilon$ is non-decreasing and bounded from above on $(0, a]$ for some $a > 0$ [76, Proposition 1.1.2(i)]. By monotone convergence theorem [77, Theorem 1.5.6], we can interchange the limit and integral operators in $\delta h(z; w)$ such that

$$\begin{aligned} \delta h(z; w) &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \mathbb{E}[q(y_i, z + \epsilon w, Y_{i+1}) - q(y_i, z, Y_{i+1})] \\ &= \mathbb{E} \left[\lim_{\epsilon \rightarrow 0^+} \frac{1}{\epsilon} \{q(y_i, z + \epsilon w, Y_{i+1}) - q(y_i, z, Y_{i+1})\} \right] \\ &= \mathbb{E} \left[\lim_{t \rightarrow z^+} p(y_i + t + Y_{i+1}) w 1_{\{w > 0\}} \right. \\ &\quad \left. + \lim_{t \rightarrow z^-} p(y_i + t + Y_{i+1}) w 1_{\{w < 0\}} \right] \\ &= \lim_{t \rightarrow z^+} \mathbb{E}[p(y_i + t + Y_{i+1}) w 1_{\{w > 0\}}] \\ &\quad + \lim_{t \rightarrow z^-} \mathbb{E}[p(y_i + t + Y_{i+1}) w 1_{\{w < 0\}}], \end{aligned} \quad (73)$$

where 1_A is the indicator function of event A . According to [75, p. 710] and the convexity of $h(z)$, z is an optimal solution to (59) if and only if the following assertion is true: If $z > 0$, then

$$\delta h(z; w) - (\bar{p}_{\text{opt}_1} + \alpha)w \geq 0, \quad \forall w \in \mathbb{R}, \quad (74)$$

otherwise, $z = 0$. Because w in (74) is an arbitrary real number, if we choose $w = 1$, then (74) becomes

$$\lim_{t \rightarrow z^+} \mathbb{E}[p(y_i + t + Y_{i+1})] - (\bar{p}_{\text{opt}_1} + \alpha) \geq 0. \quad (75)$$

Similarly, if we choose $w = -1$, then (74) implies

$$\lim_{t \rightarrow z^-} \mathbb{E}[p(y_i + t + Y_{i+1})] - (\bar{p}_{\text{opt}_1} + \alpha) \leq 0. \quad (76)$$

Because $p(\cdot)$ is non-decreasing, we can obtain from (74)-(76) that if $z > 0$, then z satisfies (77)-(78):

$$\mathbb{E}[p(y_i + t + Y_{i+1})] - (\bar{p}_{\text{opt}_1} + \alpha) \geq 0, \quad \text{if } t > z, \quad (77)$$

$$\mathbb{E}[p(y_i + t + Y_{i+1})] - (\bar{p}_{\text{opt}_1} + \alpha) \leq 0, \quad \text{if } t < z, \quad (78)$$

otherwise, $z = 0$. The smallest z satisfying (77)-(78) is

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

and the largest z satisfying (77)-(78) is

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

Hence, the set of optimal solutions to (59) is given by Lemma 5. This completes the proof.

APPENDIX C PROOF OF THEOREM 5

According to [73, Prop. 6.2.5], if we can find $\pi^* = (Z_1^*, Z_2^*, \dots)$ and α^* that satisfy the following conditions:

$$\pi^* \in \Pi_1, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[Y_i + Z_i^*] - \frac{1}{f_{\max}} \geq 0, \quad (79)$$

$$\alpha^* \geq 0, \quad (80)$$

$$L(\pi^*; \alpha^*) = \inf_{\pi \in \Pi_1} L(\pi; \alpha^*), \quad (81)$$

$$\alpha^* \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[Y_i + Z_i^*] - \frac{1}{f_{\max}} \right\} = 0, \quad (82)$$

then π^* is an optimal solution to (51) and α^* is a geometric multiplier [73] for (51). Further, if we can find such π^* and α^* , then the duality gap between (51) and (56) must be zero, because otherwise there is no geometric multiplier [73, Prop. 6.2.3(b)]. The remaining task is to find π^* and α^* that satisfy (79)-(82).

According to Lemma 6, the set of optimal solutions to (81) is given by $\Gamma(\alpha^*)$. Hence, we only need to find α^* and $\pi^* \in \Gamma(\alpha^*)$ that satisfy (79), (80), and (82). The search for such α^* and π^* falls into the following two cases:

Case 1: If (63) is satisfied, then $\alpha_1^* = 0$ and $\pi_1^* = (z_{\min}(Y_1, 0), z_{\min}(Y_2, 0), \dots)$ satisfy the conditions (79)-(82).

Case 2: If (63) is not satisfied, we seek $\alpha_2^* \geq 0$ and $\pi_2^* = (Z_1^*, Z_2^*, \dots) \in \Gamma(\alpha_2^*)$ that satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[Y_i + Z_i^*] = \frac{1}{f_{\max}}. \quad (83)$$

By Lemma 6, we can get from (83) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[Y_i + z_{\min}(Y_i, \alpha_2^*)] &\leq \frac{1}{f_{\max}} \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[Y_i + z_{\max}(Y_i, \alpha_2^*)]. \end{aligned} \quad (84)$$

Because the Y_i 's are *i.i.d.*, (84) is equivalent to

$$\mathbb{E}[Y_i + z_{\min}(Y_i, \alpha_2^*)] \leq \frac{1}{f_{\max}} \leq \mathbb{E}[Y_i + z_{\max}(Y_i, \alpha_2^*)]. \quad (85)$$

Next, we will find $\alpha_2^* \geq 0$ that satisfies (85). According to (60)-(61), $z_{\min}(y, \alpha)$ and $z_{\max}(y, \alpha)$ are non-decreasing in α . Hence, $\mathbb{E}[z_{\min}(Y_i, \alpha)]$ and $\mathbb{E}[z_{\max}(Y_i, \alpha)]$ are also non-decreasing in α . In addition, it holds that for all $\alpha_0 > 0$

$$\begin{aligned} \lim_{\alpha \rightarrow \alpha_0^-} z_{\max}(y, \alpha) &= z_{\min}(y, \alpha_0) \\ &\leq z_{\max}(y, \alpha_0) = \lim_{\alpha \rightarrow \alpha_0^+} z_{\min}(y, \alpha). \end{aligned} \quad (86)$$

By invoking the monotone convergence theorem [77, Theorem 1.5.6], we obtain that for all $\alpha_0 > 0$

$$\begin{aligned} \lim_{\alpha \rightarrow \alpha_0^-} \mathbb{E}[z_{\max}(Y_i, \alpha)] &= \mathbb{E}[z_{\min}(Y_i, \alpha_0)] \\ &\leq \mathbb{E}[z_{\max}(Y_i, \alpha_0)] = \lim_{\alpha \rightarrow \alpha_0^+} \mathbb{E}[z_{\min}(Y_i, \alpha)]. \end{aligned} \quad (87)$$

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

$$[\mathbb{E}[z_{\min}(Y_i, 0)], \infty) \subseteq \bigcup_{\alpha \geq 0} [\mathbb{E}[z_{\min}(Y_i, \alpha)], \mathbb{E}[z_{\max}(Y_i, \alpha)]]. \quad (88)$$

In Case 2, (63) is not satisfied, i.e.,

$$\mathbb{E}[Y_i + z_{\min}(Y_i, 0)] \leq \frac{1}{f_{\max}}, \quad (89)$$

then (87)-(88) tells us that there exists a unique $\alpha_2^* \geq 0$ satisfying (85). Further, policy $\pi^* \in \Gamma(\alpha_2^*)$ is chosen as

$$Z_i^* = \begin{cases} z_{\min}(Y_i, \alpha_2^*) & \text{with probability } \lambda, \\ z_{\max}(Y_i, \alpha_2^*) & \text{with probability } 1 - \lambda, \end{cases} \quad (90)$$

where λ is given by

$$\lambda = \frac{\mathbb{E}[Y_i + z_{\max}(Y_i, \alpha_2^*)] - \frac{1}{f_{\max}}}{\mathbb{E}[z_{\max}(Y_i, \alpha_2^*) - z_{\min}(Y_i, \alpha_2^*)]}. \quad (91)$$

By combining (85), (89), and (90), (83) follows. Hence, the α_2^* and π_2^* selected above satisfy the conditions (79)-(82).

In both cases, (79)-(82) are satisfied. By [73, Prop. 6.2.3(b)], the duality gap between (51) and (56) is zero. A solution to (51) and (56) 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 (15) with $\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] \geq \beta$.

In the one direction, if the zero-wait sampling policy is optimal, then the root of (16) must satisfy $\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] \geq \beta$. Substituting this into (15), yields

$D_{i+1}(\beta) = D_i(\beta) + Y_{i+1} = S_i(\beta) + Y_i + Y_{i+1}$. Combining this with (16), we get

$$\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] \geq \beta = \frac{\mathbb{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}} p(t) dt\right]}{\mathbb{E}[Y_{i+1}]}, \quad (92)$$

which implies (20).

In the other direction, if (20) holds, then by choosing

$$\beta = \frac{\mathbb{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}} p(t) dt\right]}{\mathbb{E}[Y_{i+1}]}, \quad (93)$$

we get $\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] \geq \beta$. According to (93), such a β is a root of (16). 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

$$\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] = p(2y) \geq \frac{\int_y^{2y} p(t) dt}{y} \quad (94)$$

holds for all non-decreasing $p(\cdot)$. Hence, (20) is satisfied and the zero-wait sampling policy is optimal.

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

$$\mathbb{E}[p(\text{ess inf } Y_i + Y_{i+1})] = \mathbb{E}[p(Y_{i+1})] = \mathbb{E}[p(Y_i)]. \quad (95)$$

Because $\mathbb{E}[Y_{i+1}] = \mathbb{E}[Y_i] > 0$, 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,

$$\begin{aligned} &\mathbb{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}} p(t) dt\right] \\ &> \mathbb{E}\left[\int_{Y_i}^{Y_i+Y_{i+1}} p(Y_i) dt\right] \\ &= \mathbb{E}[Y_{i+1}] \mathbb{E}[p(Y_i)]. \end{aligned} \quad (96)$$

By combining (95) and (96), (20) is not true and the zero-wait sampling policy is not optimal. This completes the proof.

APPENDIX F PROOF OF LEMMA 7

Using (69), (68) can be expressed as

$$\min_{z \in \mathbb{N}} \mathbb{E}\left[\sum_{t=0}^{z+Y_{i+1}-1} [p(t + y_i) - (\bar{p}_{\text{opt}_1} + \alpha)]\right]. \quad (97)$$

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

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=0}^{m+Y_{i+1}-1} [p(t + y_i) - (\bar{p}_{\text{opt}_1} + \alpha)]\right] \\ &\quad - \sum_{t=0}^{m+Y_{i+1}-1} [p(t + y_i) - (\bar{p}_{\text{opt}_1} + \alpha)] \\ &= \mathbb{E}[p(y_i + m + Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)]. \end{aligned} \quad (98)$$

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

$$\mathbb{E}[p(y_i + t + Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)] \leq 0, \quad t = 0, \dots, z - 1, \quad (99)$$

$$\mathbb{E}[p(y_i + t + Y_{i+1}) - (\bar{p}_{\text{opt}_1} + \alpha)] \geq 0, \quad t = z, z + 1, \dots \quad (100)$$

Using (98)-(100), one can see that $\{z_{\min}(y_i, \alpha), z_{\min}(y_i, \alpha) + 1, z_{\min}(y_i, \alpha) + 2, \dots, z_{\max}(y_i, \alpha)\}$ is the set of optimal solutions to (68). This completes the proof.