Update or Wait: How to Keep Your Data Fresh

Yin Sun[†], Elif Uysal-Biyikoglu[‡], Roy Yates[⋆], C. Emre Koksal[†], and Ness B. Shroff[†§]
[†]Dept. of ECE, [§]Dept. of CSE, The Ohio State University, Columbus, OH
[‡]Dept. of EEE, Middle East Technical University, Ankara, Turkey
[⋆]Dept. of ECE, Rutgers University, North Brunswick, NJ

***Abstract*—In this work we study how to manage the freshness of status updates sent from a source to a remote monitor via a network server. A proper metric of data freshness at the monitor is the *age-of-information*, which is defined as how old the freshest update is since the moment this update was generated at the source. A logical policy is the *zero-wait* policy, i.e., the source submits a fresh update once the server is free, which achieves the maximum throughput and the minimum average delay. Surprisingly, this zero-wait policy does not always minimize the average age. This motivates us to study how to optimally control the status updates to keep data fresh and to understand when the zero-wait policy is optimal. We introduce a penalty function to characterize the level of "dissatisfaction" on data staleness, and formulate the average age penalty minimization problem as a constrained semi-Markov decision process (SMDP) with an uncountable state space. Despite of the difficulty of this problem, we develop efficient algorithms to find the optimal status update policy. We show that, in many scenarios, the optimal policy is to wait for a certain amount of time before submitting a new update. In particular, the zero-wait policy can be far from the optimum if (i) the penalty function grows quickly with respect to the age, and (ii) the update service times are highly random and positive correlated. To the best of our knowledge, this is the first optimal control policy which is proven to minimize the age-of-information in status update systems.**

I. Introduction

In recent years, the proliferation of mobile devices and applications has significantly boosted the need for real-time information updates, such as news, weather reports, email notifications, stock quotes, social updates, mobile ads, etc. Timely status updates are also critical in network-based monitoring and control systems, including sensor networks used in temperature and air pollution monitoring, surround monitoring in autonomous vehicles, and phasor data updates in power grid stabilization systems.

A common need in these applications is to maximize the freshness of the data at the monitor. In light of this, a metric called the *age-of-information*, or simply *age*, was defined in [1]. At time t, if the freshest update at the monitor has a time stamp $U(t)$, the age is $\Delta(t) = t - U(t)$. Hence, the age is the time elapsed since the freshest packet was generated.

Most existing research on the age-of-information focuses on an "enqueue-and-forward" model [1]–[7], where status updates are randomly generated or arrive at a source node. The source enqueues these updates and forwards them later to a remote monitor through a network. It is worth noting that the goal of age minimization differs from those of throughput maximization and delay minimization: A high update frequency improves the system throughput, but may also induce a large waiting time in the queue which in turn increases the age; on the other hand, a low update frequency can reduce the queueing delay, but the monitor may end up having stale status information due to not enough updates [1], [5], [7]. In [6], it was found that a good policy is to discard the old packets waiting in the queue if a new sample arrives, which can greatly reduce the impact of queueing delay.

This work was supported in part by DTRA grant HDTRA1-14-1-0058, NSF grants CNS-1446582, CNS-1409336, CNS-1518829, CNS-1514260, CNS-1422988, and CNS-1054738, ARO grant W911NF-14-1-0368, and ONR grant N00014-15-1-2166. E. Uysal-Biyikoglu was supported in part by TUBITAK and in part by a Science Academy BAGEP award.

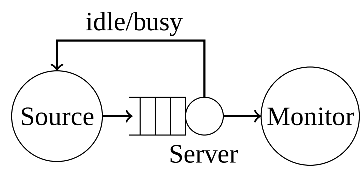


Fig. 1. System model.

In this paper, we study a "generate-at-will" model depicted in Fig. 1. In this model, the source keeps monitoring the network server's idle/busy state, and in contrast to [1]–[7], is able to generate status updates at any time by its own will. Hence, no updates need to be generated when the server is busy, which completely eliminates the waiting time in the queue and hence the queue in Fig. 1 is always empty. A simple *zero-wait* policy, also known as the *work-conserving* policy in queueing theory, that submits a fresh update once the server becomes idle, achieves the maximum throughput and the minimum average delay. Surprisingly, this zero-wait policy does **not** always minimize the average age of the information [8]. The following example reveals the reason behind this phenomenon:

Example: *Suppose that the source submits a stream of update packets to a remote monitor. The service times of these updates form a periodic sequence*

$$0, 0, 2, 2, 0, 0, 2, 2, 0, 0, 2, 2, \ldots$$

Suppose that update 1 is generated and submitted at time 0 and delivered at time 0. Under the zero-wait policy, update 2 is also generated at time 0 and delivered at time 0. However, despite of its short service time, update 2 has not brought any fresher information to the monitor after update 1 is delivered, because both updates are sampled at time 0. Therefore, the potential benefit of the zero service time of update 2 is wasted! This issue occurs periodically over time: Whenever two consecutive

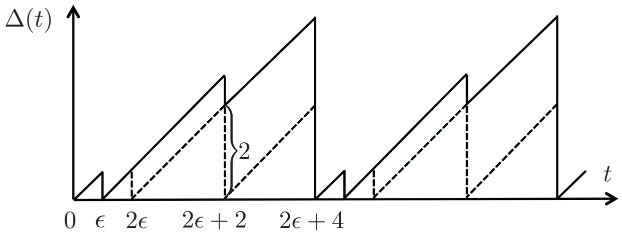


Fig. 2. Evolution of the age $\Delta(t)$ in the non-zero-wait policy in the example.

updates have zero service time, the second update of the two is wasted. Therefore, 1/4 updates in this sequence are wasted in the zero-wait policy!

For comparison, consider a non-zero-wait policy that waits for ϵ seconds after each update with a zero service time, and does not wait after each update with a service time of 2 seconds. The time-evolution of the age $\Delta(t)$ in the non-zero-wait policy is shown in Fig. 2. Update 1 is generated and delivered at time 0. Update 2 is generated and delivered at time ϵ. Update 3 is generated at time 2ϵ and is delivered at time $2\epsilon + 2$. Because the service time of update 3 is 2 seconds, the latest delivered update at time $2\epsilon + 2$ is of the age 2 seconds. Hence, the age $\Delta(t)$ drops to 2 seconds at time $2\epsilon + 2$. Update 4 is generated at time $2\epsilon + 2$ and is delivered at time $2\epsilon + 4$. At time $2\epsilon + 4$, the age drops to zero because update 5 is generated at this time and is delivered immediately.

The time-average age of the non-zero-wait policy is

$$\begin{aligned}&(\epsilon^2/2 + \epsilon^2/2 + 2\epsilon + 4^2/2)/(2\epsilon + 4)\\ =&(\epsilon^2 + 2\epsilon + 8)/(2\epsilon + 4) \text{ seconds.}\end{aligned}$$

If the waiting time is $\epsilon = 0.5$, the time-average age of the non-zero-wait policy is 1.85 seconds. If the waiting time is $\epsilon = 0$, it reduces to the zero-wait policy, whose time-average age is 2 seconds. Hence, the zero-wait policy is not optimal!

Our investigation suggests that the zero-wait policy is sub-optimal in many scenarios with various service time distributions. In particular, if the sequence of service times in this example becomes $0.2, 0.2, 2, 2, 0.2, 0.2, 2, 2, \ldots$, one can plot the time-evolution of the age $\Delta(t)$ and show that the time-average age of the non-zero-wait policy is $(\epsilon^2+2.6\epsilon+8.88)/(2\epsilon+4.4)$ seconds. If the waiting time is $\epsilon = 0.5$, the time-average age of the non-zero-wait policy is 1.93 seconds. If the waiting time is $\epsilon = 0$, we obtain the time-average age of the zero-wait policy, which is 2.02 seconds. Hence, the zero-wait policy is still not optimal. More examples with continuous service time distributions are provided in Section IV, where the sub-optimality gap of the zero-wait policy can be as large as several times of the optimum time-average age.

These examples point out a key difference between status update systems and data communication systems: In status update systems, an update packet is useful only if it carries some fresh information to the monitor; however, in communication systems, all packets are equally important. While the theory of data communications is quite mature, the optimal control of status updates remains open.

For a source that can insert waiting times between updates, the aim of this paper is to answer the following questions: How to optimally submit update packets to maximize data freshness at the monitor? When is the zero-wait policy optimal? To that end, the following are the key contributions of this paper:

- We generalize existing status update studies by introducing two new features: age penalty functions and non-*i.i.d.* service processes. We define an age penalty function $g(\Delta)$ to characterize the level of "dissatisfaction" for data staleness, where $g(\cdot)$ is *measurable, non-negative, and non-decreasing*, which is determined by the specific application. The update service process is modeled as a stationary ergodic Markov chain with an uncountable state space, which generalizes the *i.i.d.* service processes studied in previous work [1]–[9].
- We formulate the average age penalty minimization problem as a constrained semi-Markov decision process (SMDP) with an uncountable state space. Despite of the difficulty of this problem, we manage to solve it by a divide-and-conquer approach: We first prove that there exists a stationary randomized policy that is optimal for this problem (Theorem 1). Further, we prove that there exists a stationary deterministic policy that is optimal for this problem (Theorem 2). Finally, we develop a low-complexity algorithm to find the optimal stationary deterministic policy (Theorem 3). To the best of our knowledge, this is the first optimal control policy which is proven to minimize the age-of-information (i.e., maximize data freshness) in status update systems.
- We further investigate when the zero-wait policy is optimal. For the special case of proportional penalty function and *i.i.d.* service times, we devise a simpler solution to minimize the average age (Theorem 4). This solution explicitly characterizes when the zero-wait policy is optimal, and when it is not. For general age penalty functions and correlated service processes, sufficient conditions for the optimality of the zero-wait policy are provided (Lemma 3).
- Our theoretical and simulation results demonstrate that, in many scenarios, the optimal policy is to wait for a certain amount of time before submitting a new update. In particular, the zero-wait policy can be far from optimality if (i) the penalty function grows quickly with respect to the age, and (ii) the update service times are highly random and positive correlated.

II. System Model and Problem Formulation

We consider a system depicted in Fig. 1, where a source generates status update packets and sends them to a remote monitor through a network server. This server provides First-Come First-Served (FCFS) service to the submitted update packets. The service of an update packet is considered complete, when it is successfully received by the monitor. After that, the server becomes available for sending the next packet.

The source generates and submits updates at successive times $S_0, S_1, \ldots$ Update i, submitted at time S_i, is delivered at time $D_i = S_i + Y_i$, where $Y_i \geq 0$ is the service time of update i. Suppose that update 0 is submitted to an idle server at time $S_0 = -Y_0$ and delivered at $D_0 = 0$, as shown

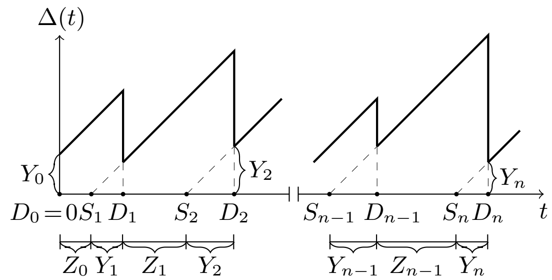


Fig. 3. Evolution of the age-of-information $\Delta(t)$.

in Fig. 3. The source has access to the idle/busy state of the server and is able to generate updates at any time by its own will. Hence, the source should not generate update $i+1$ when the server is busy processing update i, because this will incur an unnecessary waiting time in the queue. After update i is delivered at time D_i, the source may introduce a *waiting time* $Z_i \in [0, M]$ before submitting update $i+1$ at time $S_{i+1} = D_i + Z_i$, where M represents the maximum amount of waiting time allowed by the system. The source can switch to a low-power sleep mode during $[D_i, S_{i+1})$. We assume that the service process $(Y_0, Y_1, \ldots)$ is ***a stationary and ergodic Markov chain*** with a possibly uncountable state space and a positive mean $0 < \mathbb{E}[Y_i] < \infty$, which generalizes the *i.i.d.* service processes in previous studies [1]–[9]. The ergodicity of this Markov chain is assumed in the sense of ergodic theory [10], which allows the Markov chain to be periodic.[1] This Markovian service process model is introduced to study the impact of temporal-correlation on the optimality of the zero wait policy. In Section IV, we will see that the zero wait policy is close to the optimum when the service process is negative correlated; and can be far from the optimum when the service process is positive correlated. More general correlated service process model will be considered in our future work.

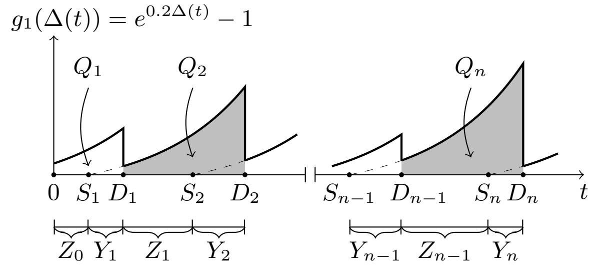
At any time t, the monitor's most recently received update packet is time-stamped with $U(t) = \max\{S_i : D_i \le t\}$. The *age-of-information* $\Delta(t)$ is defined as [1]

$$\Delta(t) = t - U(t), \tag{1}$$

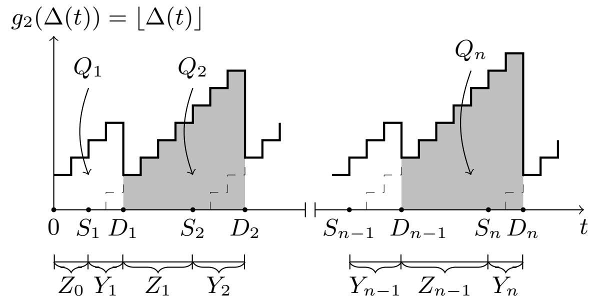
which is also referred to as *age*. As shown in Fig. 3, the age $\Delta(t)$ is a stochastic process that increases linearly with t between updates, with downward jumps occurring when updates are delivered. Specifically, when update i is sent at time $t = S_i$, it is delivered at time $D_i = S_i + Y_i$ with age $\Delta(D_i) = D_i - S_i = Y_i$. After that, the age increases linearly and reaches $\Delta(D_{i+1}^-) = Y_i + Z_i + Y_{i+1}$ just before update $i+1$ is delivered. Then, at time D_{i+1}, the age drops to $\Delta(D_{i+1}) = Y_{i+1}$.

We introduce an age penalty function $g(\Delta)$ to represent the level of "dissatisfaction" for data staleness or the "need" for new information update, where the function $g : [0, \infty) \to [0, \infty)$ is assumed to be *measurable, non-negative, and non-decreasing*. Some examples of $g(\cdot)$ are power function $g(\Delta) = \Delta^a$, exponential function $g(\Delta) = e^{a\Delta}$, and the stair-step function $g(\Delta) = \lfloor a\Delta \rfloor$, where $a \ge 0$ and $\lfloor x \rfloor$ is the largest integer no greater than x. Two examples of age penalty functions are depicted in Figure 4. Note that this age penalty model is quite general, which allows $g(\cdot)$ to be discontinuous and non-convex.

[1]The results in this paper can be readily extended to a more general model where the Markov chain $(Y_0, Y_1, \ldots)$ has a longer memory, i.e., the sequence $(W_0, W_1, \ldots)$ forms a Markov chain, with W_i defined as $W_i = (Y_i, Y_{i+1}, \ldots, Y_{i+k})$ for some finite k.



(a) An exponential age penalty function $g_1(\Delta) = e^{0.2\Delta} - 1$.



(b) A stair-step age penalty function $g_2(\Delta) = \lfloor \Delta \rfloor$.

Fig. 4. Evolutions of two age penalty functions.

To analyze the average age penalty, we decompose the area under the curve $g(\Delta(t))$ into a sum of disjoint components: Consider the time interval $[0, D_n]$, where $D_n = \sum_{i=0}^{n-1}(Y_i + Z_i)$. In this interval, the area under $g(\Delta(t))$ can be seen as the concatenation of the areas Q_i, $0 \le i \le n-1$, such that

$$\int_0^{D_n} g(\Delta(t))dt = \sum_{i=0}^{n-1} Q_i,$$

where

$$Q_i = \int_{Y_i}^{Y_i+Y_i+Z_i} g(\tau)d\tau.$$

Let us define

$$q(y, z, y') = \int_y^{y+z+y'} g(\tau)d\tau. \tag{2}$$

Then, Q_i can be expressed as $Q_i = q(Y_i, Z_i, Y_{i+1})$. We assume that

$$\mathbb{E}\left[q(Y_i, M, Y_{i+1})\right] < \infty. \tag{3}$$

Our goal is to minimize the average age penalty by controlling the sequence of waiting times $(Z_0, Z_1, \ldots)$. Let $\pi \triangleq (Z_0, Z_1, \ldots)$ denote a status update policy. We consider the class of *causally feasible* policies, in which control decisions are made based on history and current information of the system, as well as the distribution of the service process $(Y_0, Y_1, \ldots)$. Specifically, Z_i is determined based on the past realizations of $(Y_0, Y_1, \ldots, Y_i)$, without using the realizations of future service times $(Y_{i+1}, Y_{i+2}, \ldots)$; but the conditional distribution of $(Y_{i+1}, Y_{i+2}, \ldots)$ based on $(Y_0, Y_1, \ldots, Y_i)$ is available. Let Π denote the set of all causally feasible policies satisfying $Z_i \in [0, M]$ for all i.

The average age penalty can be represented by

$$\limsup_{n\to\infty} \frac{\mathbb{E}\left[\int_0^{D_n} g(\Delta(t))dt\right]}{\mathbb{E}\left[D_n\right]}.$$

Using this, the stochastic optimization problem for minimizing the average age penalty can be formulated as

$$\begin{aligned} \overline{g}_{\text{opt}} = \min_{\pi\in\Pi} \ & \limsup_{n\to\infty} \frac{\mathbb{E}\left[\sum_{i=0}^{n-1} q(Y_i, Z_i, Y_{i+1})\right]}{\mathbb{E}[\sum_{i=0}^{n-1}(Y_i + Z_i)]} \\ \text{s.t.} \ & \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1}(Y_i + Z_i)\right] \geq T_{\min}, \end{aligned} \tag{4}$$

where $\overline{g}_{\text{opt}}$ is the optimum objective value of Problem (4), the expectation $\mathbb{E}$ is taken over the stochastic service process $(Y_0, Y_1, \ldots)$ for given policy π, and $T_{\min}$ is the minimum average update period of the source due to hardware and physical constraints (e.g., limited power resource and cooling capacity). We assume $M > T_{\min}$ such that Problem (4) is feasible and $\overline{g}_{\text{opt}} < \infty$.

Problem (4) belongs to the class of constrained semi-Markov decision processes (SMDP) with a possibly uncountable state space, which is well-known for its difficulty. In this problem, Y_i is the state of the embedded Markov chain, Z_i is the control action taken after observing Y_i, Y_i+Z_i is the update period, and $q(Y_i, Z_i, Y_{i+1})$ is the reward related to both stage i and $i+1$. The class of SMDPs include Markov decision problems (MDPs) [13], [15] and optimization problems of renewal processes [16] as special cases. Most existing studies on SMDPs deal with (i) unconstrained SMDPs, e.g., [11], [13], [17], [18], or (ii) constrained SMDPs with a countable state space, e.g., [19]–[22]. However, the optimality equations (e.g., Bellman's equation) for solving unconstrained SMDPs are not applied to constrained SMDPs [23], and the studies for problems with a countable state space cannot be directly applied to Problem (4), which has an uncountable state space.

III. Optimal Status Update Policy

In this section, we develop a chain of novel theoretical results to solve Problem (4): First, we prove that there exists a *stationary randomized* policy that is optimal for Problem (4). Further, we prove that there exists a *stationary deterministic* policy that is optimal for Problem (4). Finally, we develop a low-complexity algorithm to find the *optimal* stationary deterministic policy that solves Problem (4).

A. Optimality of Stationary Randomized Policies

A policy $\pi \in \Pi$ is said to be a *stationary randomized* policy, if it observes Y_i and then chooses a waiting time $Z_i \in [0, M]$ based only on the observed value of Y_i. In this case, Z_i is determined according to a conditional probability measure $p(y, A) \triangleq \Pr[Z_i \in A | Y_i = y]$ that is invariant for all $i = 0, 1, \ldots$ We use Π_{SR} ($\Pi_{\text{SR}} \subseteq \Pi$) to denote the set of stationary randomized policies such that

$$\begin{aligned} \Pi_{\text{SR}} = \{\pi : & \text{ After observing } Y_i = y_i,\ Z_i \in [0, M] \text{ is chosen} \\ & \text{according to probability measure } p(y_i, A), i = 0, 1, \ldots\}. \end{aligned}$$

Note that (Y_i, Z_i, Y_{i+1}) is stationary and ergodic for all stationary randomized policies. In the sequel, when we refer to the stationary distribution of a stationary randomized policy $\pi \in \Pi_{\text{SR}}$, we will remove subscript i. In particular, the random variables (Y_i, Z_i, Y_{i+1}) are replaced by (Y, Z, Y'), where Z is chosen based on the conditional probability measure $\Pr[Z \in A | Y = y] = p(y, A)$ after observing $Y = y$, and (Y, Y') have the same joint distribution as (Y_0, Y_1). The first key result of this paper is stated as follows:

Theorem 1: (Optimality of Stationary Randomized Policies) If $g(\cdot)$ is measurable and non-negative, $(Y_0, Y_1, \ldots)$ is a stationary ergodic Markov chain with $Y_i \geq 0$ and $0 < \mathbb{E}[Y_i] < \infty$, condition (3) is satisfied, then there exists a stationary randomized policy that is optimal for Problem (4).

Proof sketch of Theorem 1: For any policy $\pi \in \Pi$, define finite horizon average occupations

$$a_{n,\pi} \triangleq \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} q(Y_i, Z_i, Y_{i+1})\right] - \frac{\overline{g}_{\text{opt}}}{n}\mathbb{E}\left[\sum_{i=0}^{n-1}(Y_i + Z_i)\right], \tag{5}$$

$$b_{n,\pi} \triangleq \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1}(Y_i + Z_i)\right]. \tag{6}$$

Let Γ_{SR} be the set of limit points of sequences $((a_{n,\pi}, b_{n,\pi}),\ n = 1, 2, \ldots)$ associated with stationary randomized policies in Π_{SR}. We first prove that Γ_{SR} is convex and compact. Then, we show that there exists an optimal policy π_{opt} of Problem (4), such that the sequence $((a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}}), n = 1, 2, \ldots)$ associated with policy π_{opt} has a limit point (a^*, b^*) satisfying $(a^*, b^*) \in \Gamma_{\text{SR}}$, $a^* \leq 0$, and $b^* \geq T_{\min}$. Since $(a^*, b^*) \in \Gamma_{\text{SR}}$, there exists a stationary randomized policy π^* achieving this limit point (a^*, b^*). Finally, we show that policy π^* is optimal for Problem (4), which completes the proof. The details are available in Appendix A. ■

The convexity and compactness properties of the set of occupation measures are essential in the study of constrained MDPs [24, Sec. 1.5], which dates back to Derman's monograph in 1970 [25]. Recently, it was used in stochastic optimization for discrete-time queueing systems and renewal processes, e.g., [16], [26]. The techniques in these studies cannot directly handle constrained SMDPs with an uncountable state space, like Problem (4). One crucial novel idea in our proof is to introduce $\overline{g}_{\text{opt}}$ in the definition of average occupation in (5), which turns out to be essential in later steps for showing the optimality of the stationary randomized policy π^*.

By Theorem 1, we only need to consider the class of stationary randomized policies Π_{SR}. Therefore, Problem (4) can be simplified to the following functional optimization problem (as shown in Appendix A):

$$\begin{aligned} \min_{p(y,A)} \ & \frac{\mathbb{E}[q(Y, Z, Y')]}{\mathbb{E}[Y + Z]} \\ \text{s.t.} \ & \mathbb{E}[Y + Z] \geq T_{\min} \\ & 0 \leq Z \leq M \end{aligned} \tag{7}$$

where $p(y, A) = \Pr[Z \in A | Y = y]$ is the conditional probability measure of some stationary randomized policy, and (Y, Y') have the same distribution as (Y_0, Y_1).

Algorithm 1 Two-layer bisection method for Problem (8)

given $l = 0$, sufficiently large $u > \overline{g}_{\text{SD}}$, tolerance ϵ_1.
repeat
  $c := (l + u)/2$.
  given $\zeta_l = 0$, sufficiently large $\zeta_u > 0$, tolerance ϵ_2.
  $\zeta := \zeta_l,\ \nu := \zeta + c$.
  Compute $z_\nu(\cdot)$ in (11).
  if $\mathbb{E}[z_\nu(Y)] + \mathbb{E}[Y] < T_{\min}$ **then**
    repeat
      $\zeta := (\zeta_l + \zeta_u)/2,\ \nu := \zeta + c$.
      Compute $z_\nu(\cdot)$ in (11).
      if $\mathbb{E}[z_\nu(Y)] + \mathbb{E}[Y] \geq T_{\min}$, $\zeta_u := \zeta$; **else**, $\zeta_l := \zeta$.
    until $\zeta_u - \zeta_l \leq \epsilon_2$.
  end if
  if $f(c) \leq 0$, $u := c$; **else**, $l := c$.
until $u - l \leq \epsilon_1$.
return $z(\cdot) := z_\nu(\cdot)$.

B. Optimality of Stationary Deterministic Policies

A policy $\pi \in \Pi_{\text{SR}}$ is said to be a *stationary deterministic* policy if $Z_i = z(Y_i)$ for all $i = 0, 1, \ldots$, where $z : [0, \infty) \to [0, M]$ is a deterministic function. We use Π_{SD} ($\Pi_{\text{SD}} \subseteq \Pi_{\text{SR}}$) to denote the set of stationary deterministic policies such that

$$\Pi_{\text{SD}} = \{\pi : Z_i = z(Y_i) \text{ for all } i, 0 \leq z(y) \leq M, \forall\ y \geq 0\}.$$

Theorem 2: (Optimality of Stationary Deterministic Policies) If $g(\cdot)$ is measurable and non-decreasing, then there exists a stationary deterministic policy that is optimal for Problem (7).

Proof sketch of Theorem 2: Since $g(\cdot)$ is non-decreasing, $q(y, \cdot, y')$ is convex. Using Jensen's inequality, we can show that for any feasible stationary randomized policy $\pi_1 \in \Pi_{\text{SR}}$, there is a feasible stationary deterministic policy that is no worse than policy π_1. The details are provided in Appendix B. ■

Let μ_Y be the probability measure of Y_i, then any bounded measurable function $z : [0, \infty) \to [0, M]$ belongs to the Lebesgue space $L^2(\mu_Y)$ [27, Section 3], because

$$\int_0^\infty |z(y)|^2 d\mu_Y(y) \leq \int_0^\infty M^2 d\mu_Y(y) = M^2 < \infty.$$

By Theorems 1 and 2, we only need to consider the class of stationary deterministic policies Π_{SD} and Problem (4) is simplified to the following functional optimization problem:

$$\min_{z(\cdot) \in L^2(\mu_Y)} \frac{\mathbb{E}\left[q(Y, z(Y), Y')\right]}{\mathbb{E}[Y + z(Y)]} \tag{8}$$
$$\text{s.t.} \quad \mathbb{E}[Y + z(Y)] \geq T_{\min} \tag{9}$$
$$0 \leq z(y) \leq M,\ \forall\ y \geq 0,$$

where $z(\cdot)$ is the function associated with a stationary deterministic policy $\pi \in \Pi_{\text{SD}}$. The optimum objective value of Problem (8) is equal to $\overline{g}_{\text{opt}}$.

C. A Low Complexity Solution to Problem (8)

Lemma 1: If $g(\cdot)$ is measurable, non-negative, and non-decreasing, then the functional $h : L^2(\mu_Y) \to [0, \infty)$ defined by

$$h(z) = \frac{\mathbb{E}\left[q(Y, z(Y), Y')\right]}{\mathbb{E}[Y + z(Y)]}$$

is quasi-convex.

Proof: See Appendix C. ■

Therefore, Problem (8) is a functional quasi-convex optimization problem. In order to solve Problem (8), we consider the following functional convex optimization problem with a parameter c:

$$f(c) = \min_{z(\cdot) \in L^2(\mu_Y)} \mathbb{E}\left[q(Y, z(Y), Y')\right] - c\mathbb{E}[Y + z(Y)] \tag{10}$$
$$\text{s.t.} \quad \mathbb{E}[Y + z(Y)] \geq T_{\min}$$
$$0 \leq z(y) \leq M,\ \forall\ y \geq 0.$$

It is easy to show that $\overline{g}_{\text{opt}} \leq c$ if and only if $f(c) \leq 0$ [28]. Therefore, we can solve Problem (8) by a two-layer nested algorithm, such as Algorithm 1. In the inner layer, we use *bisection* to solve Problem (10) for given parameter c; in the outer layer, we employ *bisection* again to search for a c^* such that $f(c^*) = 0$ and thus $\overline{g}_{\text{opt}} = c^*$. Algorithm 1 has low complexity. It requires at most $\lceil \log_2((u - l)/\epsilon_1) \rceil \times \lceil \log_2((\zeta_u - \zeta_l)/\epsilon_2) \rceil$ iterations to terminate. Each iteration involves computing $\mathbb{E}[z_\nu(Y)]$ based on (11). The optimality of Algorithm 1 is guaranteed by the following theorem:

Theorem 3: If $g(\cdot)$ is measurable, non-negative, and non-decreasing, then an optimal solution π_{opt} to Problem (8) is obtained by Algorithm 1, where the function $z_\nu(\cdot)$ is determined by

$$z_\nu(y) = \sup\{z \in [0, M] : \mathbb{E}\left[g(y + z + Y') | Y = y\right] \leq \nu\}. \tag{11}$$

Proof Sketch of Theorem 3: We use Lagrangian duality theory to solve Problem (8). Different from traditional finite dimensional optimization problems [28], Problem (8) is an infinite dimensional functional optimization problem. Therefore, the Karush-Kuhn-Tucker (KKT) theorem for infinite dimensional space [29], [30] and the calculus of variations are required in the analysis. In particular, since the Lagrangian may not be strictly convex for some penalty functions, one-sided Gâteaux derivative (similar to sub-gradient in finite dimensional space) is used to solve the KKT conditions in Lebesgue space $L^2(\mu_Y)$. The proof details are provided in Appendix D. ■

The policy spaces Π, Π_{SR}, Π_{SD}, and the obtained optimal policy π_{opt} are depicted in Fig. 5.

IV. When Is It Better to Wait than to Update?

When $T_{\min} \leq \mathbb{E}[Y]$, a logical policy is the *zero wait* policy: the source submits a fresh update once the prior update completes service, i.e., $\pi_{\text{zero wait}} = (0, 0, \ldots)$. As mentioned before, this zero wait policy is not always optimal to keep data fresh. When $T_{\min} > \mathbb{E}[Y]$, due to the constraint (9), the minimum possible average waiting time is $\mathbb{E}[z(Y)] = T_{\min} - \mathbb{E}[Y]$. However, even in this case, the optimal policy may have additional waiting time such that $\mathbb{E}[z(Y)] > T_{\min} - \mathbb{E}[Y]$.

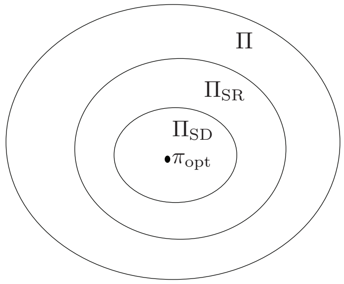


Fig. 5. Illustration of the set of causally feasible policies Π, the set of stationary randomized policies Π_{SR}, the set of stationary deterministic policies Π_{SD}, and the obtained optimal policy π_{opt}.

In this section, we will study when it is optimal to submit updates with the minimum wait and when it is not.

A. A Special Case of $g(\Delta)=\Delta$ with i.i.d. Service Times

Consider the case that $g(\Delta)=\Delta$ and the Y_i's are *i.i.d.*. In this case, Problem (8) has a simpler solution than that provided by Algorithm 1. Interestingly, this solution explicitly characterizes whether the optimal control $z(\cdot)$ can have minimum wait such that $\mathbb{E}[z(Y)]=T_{\min}-\mathbb{E}[Y]$.

As shown in Fig. 3, $q(y,z,y')$ is the area of a trapezoid, computed as

$$q(y,z,y')=\frac{1}{2}(2y+z+y')(z+y').$$

Because the Y_i's are *i.i.d.*, Y and Y' in Problem (8) are also *i.i.d.* Using this, we can obtain

$$\begin{aligned}&\mathbb{E}[q(Y,z(Y),Y')]\\ =&\mathbb{E}\left[\frac{1}{2}(Y+z(Y))^2+(Y+z(Y))Y'\right] \qquad (12)\\ =&\frac{1}{2}\mathbb{E}\left[(Y+z(Y))^2\right]+\mathbb{E}\left[Y+z(Y)\right]\mathbb{E}\left[Y'\right],\end{aligned}$$

where in (12) we have used that $\mathbb{E}[Y^2]=\mathbb{E}[Y'^2]$. Hence, Problem (8) can be reformulated as

$$\min_{z\in L^2(\mu_Y)} \frac{\mathbb{E}[(Y+z(Y))^2]}{2\mathbb{E}[Y+z(Y)]}+\mathbb{E}[Y] \qquad (13)$$

$$\text{s.t.}\quad \mathbb{E}[Y+z(Y)]\geq T_{\min} \qquad (14)$$

$$0\leq z(y)\leq M,\ \forall\, y\geq 0.$$

The following lemma tells us that Problem (13) is a functional convex optimization problem.

Lemma 2: The functional $h_1: L^2(\mu_Y)\to\mathbb{R}$ defined by

$$h_1(z)=\frac{\mathbb{E}[(Y+z(Y))^2]}{\mathbb{E}[Y+z(Y)]}$$

is convex on the domain

$$\textbf{dom}\ h_1=\left\{z: z(y)\in[0,M],\ \forall y\geq 0, z\in L^2(\mu_Y)\right\}.$$

Proof: See Appendix E. ■

Using the KKT theorem for infinite dimensional space and the calculus of variations, we can obtain

Theorem 4: The optimal solution to Problem (13) is

$$z(y)=(\beta-y)_0^M, \qquad (15)$$

where $(x)_a^b\triangleq\min[\max[x,a],b]$ and $\beta>0$ is the root of the following equation:

$$\mathbb{E}\left[(\beta)_Y^{M+Y}\right]=\max\left(T_{\min},\frac{\mathbb{E}[((\beta)_Y^{M+Y})^2]}{2\beta}\right). \qquad (16)$$

Proof: See Appendix F. ■

Algorithm 2 Bisection method for solving Problem (13)

given $l=0$, sufficiently large u, tolerance ϵ.
repeat
  $\beta:=(l+u)/2$.
  $o:=\mathbb{E}\left[(\beta)_Y^{M+Y}\right]-\max\left(T_{\min},\frac{\mathbb{E}\left[((\beta)_Y^{M+Y})^2\right]}{2\beta}\right)$, where $(x)_a^b=\min[\max[x,a],b]$.
  if $o\geq 0$, $u:=\beta$; **else**, $l:=\beta$.
until $u-l\leq\epsilon$.
Compute $z(\cdot)$ by (15).
return $z(\cdot)$.

Equation (15) has the form of a water-filling solution, where the water-level β is given by the root of equation (16). One can observe that (11) reduces to (15) if $g(\Delta)=\Delta$, the Y_i's are *i.i.d.*, and τ is replaced by $\beta+\mathbb{E}[Y]$. The root β of equation (16) can be simply solved by the bisection search method in Algorithm 2. We note that Algorithm 2 has a lower complexity than Algorithm 1 in the special case of $g(\Delta)=\Delta$ and *i.i.d.* service process, while Algorithm 1 can obtain the optimal policy in more general scenarios.

Interestingly, Theorem 4 provides a closed-form criterion on whether the optimal $z(\cdot)$ satisfies $\mathbb{E}[z(Y)]=T_{\min}-\mathbb{E}[Y]$. Specifically, (15) implies $(\beta)_Y^{M+Y}=Y+z(Y)$. This and (16) tell us that if $T_{\min}\geq\frac{\mathbb{E}[(Y+z(Y))^2]}{2\beta}$, then the optimal control $z(\cdot)$ satisfies

$$\mathbb{E}[Y+z(Y)]=T_{\min}\geq\frac{\mathbb{E}[(Y+z(Y))^2]}{2\beta}, \qquad (17)$$

such that the optimal policy has the minimum average waiting time $\mathbb{E}[z(Y)]=T_{\min}-\mathbb{E}[Y]$; otherwise, if $T_{\min}<\frac{\mathbb{E}[(Y+z(Y))^2]}{2\beta}$, the optimal control $z(\cdot)$ satisfies

$$\mathbb{E}[Y+z(Y)]=\frac{\mathbb{E}[(Y+z(Y))^2]}{2\beta}>T_{\min}, \qquad (18)$$

such that the average waiting time $\mathbb{E}[z(Y)]$ of the optimal policy is larger than $T_{\min}-\mathbb{E}[Y]$.

Furthermore, we consider the case $T_{\min}=0$, where the constraint (14) is always satisfied and can be removed. Note that we have the same problem for all $T_{\min}\leq\mathbb{E}[Y]$ and hence we can pick $T_{\min}=\mathbb{E}[Y]$. By substituting $T_{\min}=\mathbb{E}[Y]$ into (17) and (18), we can obtain the criterion on whether the zero waiting policy is optimal.

1) Simulation Results: We use "Optimal policy" to refer to the policy provided in Theorem 3 (or its special case in Theorem 4), and compare it with two reference policies:

- "Constant wait": Each update is followed by a constant wait $Z_i=const$ before submitting the next update with $const=T_{\min}-\mathbb{E}[Y]$.
- "Minimum wait": The update waiting time is given by a

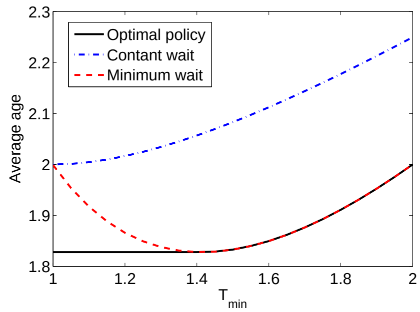


Fig. 6. Average age vs. $T_{\min}$ with $i.i.d.$ discrete service times.

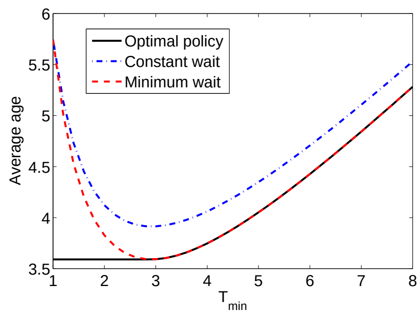


Fig. 7. Average age vs. $T_{\min}$ with $i.i.d.$ log-normal distributed service times.

deterministic function $Z_i = z(Y_i)$, where $z(\cdot)$ is given by (15) and β is chosen to satisfy $\mathbb{E}[z(Y)] = T_{\min} - \mathbb{E}[Y]$.[2]

When $\mathbb{E}[Y] = T_{\min}$, both the constant wait and minimum wait policies reduce to the zero wait policy. Two models of the service processes are considered: The first one is a discrete Markov chain with a probability mass function $\Pr[Y_i = 0] = \Pr[Y_i = 2] = 0.5$ and a transition matrix

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Hence, the Y_i's are *i.i.d.* when $p = 0.5$, and the correlation coefficient between Y_i and Y_{i+1} is $\rho = 2p - 1$. The second one is a log-normal distributed Markov chain, where $Y_i = e^{\sigma X_i}/\mathbb{E}[e^{\sigma X_i}]$ and $(X_0, X_1, \ldots)$ is a Gaussian Markov process satisfying the first-order AR equation

$$X_{i+1} = \eta X_i + \sqrt{1-\eta^2} W_i,$$

where $\sigma > 0$ is the scale parameter of log-normal distribution, $\eta \in [-1, 1]$ is the parameter of the AR model, and the W_i's are *i.i.d.* Gaussian random variables with zero mean and unit variance. The log-normal distributed Markov chain is normalized such that $\mathbb{E}[Y_i] = 1$. According to the properties of log-normal distribution, the correlation coefficient between Y_i and Y_{i+1} is $\rho = (e^{\eta} - 1)/(e - 1)$. Then, the Y_i's are *i.i.d.* when $\eta = 0$. The value of M is set to be 10.

Figures 6 and 7 illustrate the average age vs. $T_{\min}$ for *i.i.d.* discrete and log-normal distributed service times, respectively. In both figures, one can observe that the constant wait policy always incurs a larger average age than the optimal policy. In addition, as expected from (17) and (18), as $T_{\min}$ exceeds a certain threshold, the optimal policy meets the constraint (14) with equality. For smaller values of $T_{\min}$, the constraint is not active for the optimal solution. Consequently, the average age achieved by the minimum wait policy deviates from the optimum for small values of $T_{\min}$.

[2]This policy was called "β-minimum" in [8].

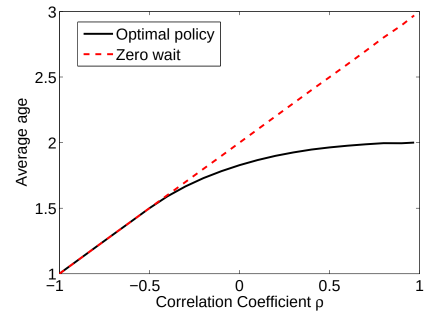


Fig. 8. Average age vs. the correlation coefficient ρ between Y_i and Y_{i+1} for discrete service times, where $T_{\min} \leq \mathbb{E}[Y]$.

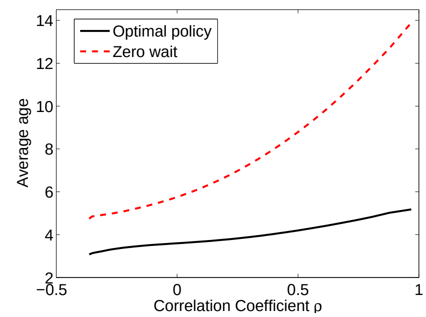


Fig. 9. Average age vs. the correlation coefficient ρ between Y_i and Y_{i+1} for log-normal distributed service times, where $T_{\min} \leq \mathbb{E}[Y]$.

B. General Age Penalties and Correlated Service Processes

For general age penalties and correlated service processes, it is essentially difficult to find closed-form characterization on whether the optimal control $z(\cdot)$ can have minimum wait such that $\mathbb{E}[z(Y)] = T_{\min} - \mathbb{E}[Y]$. Therefore, we focus on the case of $T_{\min} \leq \mathbb{E}[Y]$ and study when the zero wait policy minimizes the average age penalty. Sufficient conditions for the optimality of the zero wait policy are provided as follows:

Lemma 3: Suppose that $T_{\min} \leq \mathbb{E}[Y]$, $g(\cdot)$ is measurable, non-negative, and non-decreasing. The zero wait policy is optimal for Problem (8) if one of the following is satisfied:

1). The correlation coefficient between Y_i and Y_{i+1} is -1;
2). The Y_i's are equal to a constant value;
3). $g(\cdot)$ is a constant function.

Proof: See Appendix G. ∎

1) Simulation Results: We now provide some simulation results for general age penalties and/or correlated service processes. Figures 8 and 9 depict the average age vs. the correlation coefficient ρ between Y_i and Y_{i+1} for discrete and log-normal distributed service times, respectively. In Fig. 8, the regime of ρ is $[-1, 1]$. We observe that the zero wait policy is optimal when $\rho \in [-1, -0.5]$, and the performance gap between the optimal policy and the zero wait policy grows with ρ when $\rho \geq -0.5$. This is in accordance with the

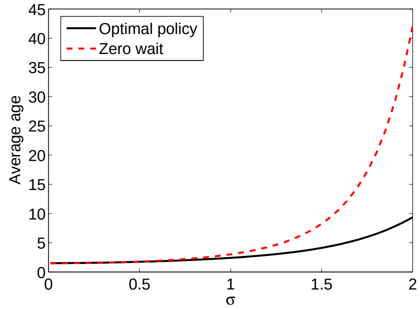


Fig. 10. Average age vs. the distribution parameter σ of $i.i.d.$ log-normal distributed service times, where $T_{\min} \leq \mathbb{E}[Y]$.

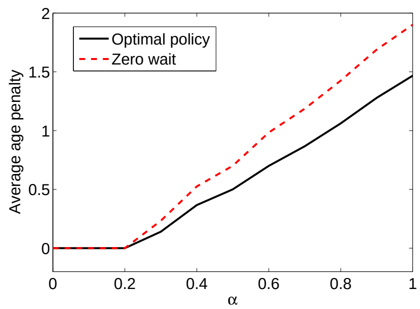


Fig. 11. Average age penalty vs. the parameter α of stair-step penalty functions with discrete service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = \lfloor \alpha\Delta \rfloor$, and $\rho = 0.4$.

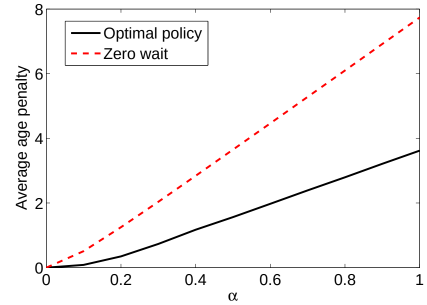


Fig. 12. Average age penalty vs. the parameter α of stair-step penalty functions with log-normal distributed service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = \lfloor \alpha\Delta \rfloor$, and $\rho = (e^{0.5} - 1)/(e - 1)$.

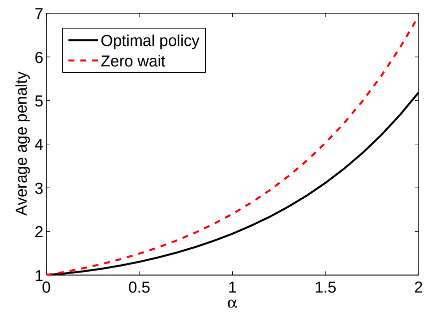


Fig. 13. Average age penalty vs. the parameter α of power penalty functions with discrete service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = \Delta^{\alpha}$, and $\rho = 0.4$.

example in the introduction: As ρ grows, the occurrence of two consecutive zero service times (i.e., $(Y_i, Y_{i+1}) = (0, 0)$) increases. Therefore, more and more updates are wasted in the zero wait policy, leading to a larger gap from the optimum. In Fig. 9, the regime of ρ is $[(e^{-1} - 1)/(e - 1), 1]$. In this case, the sub-optimality gap of the zero wait policy also increases with ρ. The point $\rho = 1$ is not plotted in these figures because the corresponding Markov chains are not ergodic.

Figure 10 considers the average age vs. the parameter σ of log-normal distributed service times, where $\rho = (e^{0.5} - 1)/(e - 1)$. We observe that the zero wait policy is optimal for small σ and is not optimal for large σ. When $\sigma = 0$, the service times are constant, i.e., $Y_i = 1$ for all i, and hence by Lemma 3, the zero wait policy is optimal. For large σ, the time-average age of the zero wait policy is significantly larger than the optimum. This implies that the sub-optimality gap of the zero wait policy can be quite large for heavy-tail service time distributions.

Figures 11-16 show the average age penalty vs. the parameter α of three types of age penalty functions, where the stair-step function $g(\Delta) = \lfloor \alpha\Delta \rfloor$ is considered in Fig. 11 and 12, the power function $g(\Delta) = \Delta^{\alpha}$ is considered in Fig. 13 and 14, and the exponential function $g(\Delta) = e^{\alpha\Delta} - 1$ is considered in Fig. 15 and 16. The correlation coefficient is $\rho = 0.4$ for discrete service times, and is $\rho = (e^{0.5} - 1)/(e - 1)$ for log-normal distributed service times. We find that the zero wait policy is optimal if $\alpha = 0$, in which case $g(\Delta)$ is a constant function. When $\alpha > 0$, the zero wait policy may not be optimal.

These simulation results suggest that the conditions in Lemma 3 are sufficient but not necessary.

V. Conclusion

We studied the optimal control of status updates from a source to a remote monitor via a FCFS network server. We generalized the existing studies on the age-of-information to cover general age penalty functions and non-$i.i.d.$ service processes. We developed efficient algorithms to find the optimal status update policy for minimizing the average age penalty. We showed that, surprisingly, in many scenarios, the optimal policy is to wait for a certain time before submitting a new update. In particular, the widely-adopted zero wait policy can be far from the optimum if (i) the penalty function grows quickly with respect to the age, and (ii) the update service times are highly random and positive correlated.

Appendix A
Proof of Theorem 1

A. An Upper Bound of $\overline{g}_{opt}$

By restricting Π in Problem (4) to Π_{SR}, we obtain the following problem:

$$\overline{g}_{\text{SR}} = \min_{\pi \in \Pi_{\text{SR}}} \limsup_{n \to \infty} \frac{\mathbb{E}\left[\sum_{i=0}^{n-1} q(Y_i, Z_i, Y_{i+1})\right]}{\mathbb{E}[\sum_{i=0}^{n-1}(Y_i + Z_i)]} \tag{19}$$
$$\text{s.t.} \quad \liminf_{n \to \infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1}(Y_i + Z_i)\right] \geq T_{\min},$$

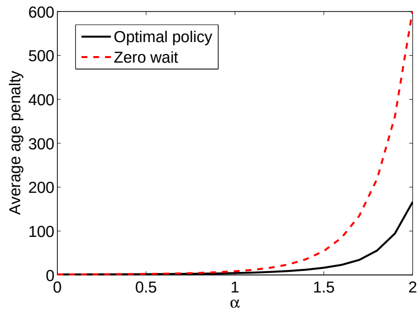


Fig. 14. Average age penalty vs. the parameter α of power penalty functions with log-normal distributed service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = \Delta^{\alpha}$, and $\rho = (e^{0.5} - 1)/(e - 1)$.

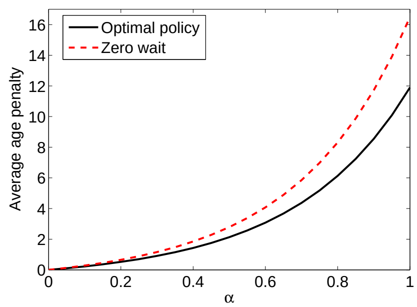


Fig. 15. Average age penalty vs. the parameter α of exponential penalty functions with discrete service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = e^{\alpha\Delta} - 1$, and $\rho = 0.4$.

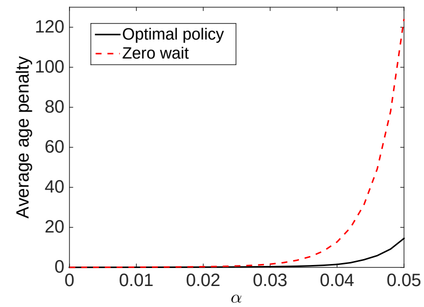


Fig. 16. Average age penalty vs. the parameter α of exponential penalty functions with log-normal distributed service times, where $T_{\min} \leq \mathbb{E}[Y]$, $g(\Delta) = e^{\alpha\Delta} - 1$, and $\rho = (e^{0.5} - 1)/(e - 1)$.

where $\overline{g}_{\text{SR}}$ is the optimum objective value of Problem (19). Since $\Pi_{\text{SR}} \subseteq \Pi$, we can obtain

$$\overline{g}_{\text{SR}} \geq \overline{g}_{\text{opt}}. \tag{20}$$

It is easy to show that the (Y_i, Z_i, Y_{i+1})'s are stationary and ergodic for all stationary randomized policies. This, together with the condition that $g(\cdot)$ is measurable, tells us that $q(Y_i, Z_i, Y_{i+1})$ is stationary and ergodic [10, Theorems 7.1.1 and 7.1.3]. For any stationary randomized policy $\pi = (Z_0, Z_1, \ldots) \in \Pi_{\text{SR}}$, we obtain

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} q(Y_i, Z_i, Y_{i+1})\right] = \mathbb{E}[q(Y_0, Z_0, Y_1)], \tag{21}$$

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} (Y_i + Z_i)\right] = \mathbb{E}[Y_0 + Z_0]. \tag{22}$$

Hence, Problem (19) can be reformulated as Problem (7).

B. The Upper Bound of $\overline{g}_{opt}$ is Tight, i.e., $\overline{g}_{SR} = \overline{g}_{opt}$

We will show $\overline{g}_{\text{SR}} = \overline{g}_{\text{opt}}$ in 4 steps. The following definitions are needed: Since $\overline{g}_{\text{opt}}$ is finite, for each causally feasible policy $\pi = (Z_0, Z_1, \ldots) \in \Pi$ we can define $a_{n,\pi}$ and $b_{n,\pi}$ as in (5) and (6), respectively.

Further, define Γ_{SR} as the set of limit points of sequences $\left((a_{n,\pi}, b_{n,\pi}), n = 1, 2, \ldots\right)$ associated with all stationary randomized policies $\pi \in \Pi_{\text{SR}}$. Because the renewal reward $q(Y_i, Z_i, Y_{i+1})$ and renewal interval $Y_i + Z_i$ are stationary and ergodic for all stationary randomized policies $\pi \in \Pi_{\text{SR}}$, the sequence $(a_{n,\pi}, b_{n,\pi})$ has a unique limit point in the form of

$$\left(\mathbb{E}[q(Y, Z, Y')] - \overline{g}_{\text{opt}}\mathbb{E}[Y + Z], \mathbb{E}[Y + Z]\right). \tag{23}$$

Hence, Γ_{SR} is the set of all points $(\mathbb{E}[q(Y, Z, Y')] - \overline{g}_{\text{opt}}\mathbb{E}[Y + Z], \mathbb{E}[Y + Z])$, where each point is associated with a conditional probability measure $p(y, A) = \Pr[Z \in A | Y = y]$, and the measure of (Y, Y') is the same as that of (Y_0, Y_1). Note that $\mathbb{E}[Y] = \mathbb{E}[Y']$.

Step 1: We will show that Γ_{SR} is a convex and compact set.

It is easy to show that Γ_{SR} is convex by considering a stationary randomized policy that is a mixture of two stationary randomized policies.

For compactness, let $((d_j, e_j), j = 1, 2, \cdots)$ be any sequence of points in Γ_{SR}, we need to show that there is a convergent subsequence (d_{j_k}, e_{j_k}) whose limit is also in Γ_{SR}. Since $(d_j, e_j) \in \Gamma_{\text{SR}}$, there must exist $(Y, Z_{(j)}, Y')$ with conditional probability $p_j(y, A) = \Pr[Z_{(j)} \in A | Y = y]$, such that $d_j = \mathbb{E}[q(Y, Z_{(j)}, Y')] - \overline{g}_{\text{opt}}\mathbb{E}[Y + Z_{(j)}]$, $e_j = \mathbb{E}[Y + Z_{(j)}]$. Let μ_j be the joint probability measure of $(Y, Z_{(j)}, Y')$, then (d_j, e_j) is uniquely determined by μ_j. For any L satisfying $L \geq M$, we can obtain

$$\begin{aligned}
&\mu_j(Y \leq L, Z_{(j)} \leq L, Y' \leq L) \\
=&\Pr(Y \leq L, Y' \leq L) \\
\geq&\Pr(Y + Y' \leq L) \\
\geq&1 - \frac{\mathbb{E}[Y + Y']}{L}, \ \forall\ j,
\end{aligned}$$

where the equality is due to the fact that $Z_{(j)} \leq M \leq L$ and the last inequality is due to Markov's inequality. Therefore, for any ϵ, there is an L such that

$$\liminf_{j \to \infty} \mu_j(|Y| \leq L, |Z_{(j)}| \leq L, |Y'| \leq L) \geq 1 - \epsilon.$$

Hence, the sequence of measures μ_j is tight. By Helly's selection theorem [10, Theorem 3.9.2], there is a subsequence of measures μ_{j_k} that converges weakly to a limit measure μ_∞.

Let $(Y, Z_{(\infty)}, Y')$ and $p_\infty(y, A) = \Pr[Z_\infty \in A | Y = y]$ denote the random vector and conditional probability corresponding to the limit measure μ_∞, respectively. We can define $d_\infty = \mathbb{E}[q(Y, Z_{(\infty)}, Y')] - \overline{g}_{\text{opt}}\mathbb{E}[Y + Z_{(\infty)}]$, $e_\infty = \mathbb{E}[Y + Z_{(\infty)}]$. Since the function $q(y, z, y')$ is in the form of an integral, it is continuous and thus measurable. Using the continuous mapping theorem [10, Theorem 3.2.4], we can obtain that $q(Y, Z_{(j_k)}, Y')$ converges weakly to $q(Y, Z_{(\infty)}, Y')$. Then, using the condition (3), together with the dominated convergence theorem (Theorem 1.6.7 of [10]) and Theorem 3.2.2 of [10], we can obtain $\lim_{k\to\infty}(d_{j_k}, e_{j_k}) = (d_\infty, e_\infty)$. Hence, $((d_j, e_j), j = 1, 2, \cdots)$ has a convergent subsequence. Further, we can generate a stationary randomized policy $\pi_{\infty,\text{SR}}$ by using the conditional probability $p_\infty(y, A)$ corresponding to μ_∞. Then, (d_∞, e_∞) is the limit point generated by the stationary randomized policy $\pi_{\infty,\text{SR}}$, which implies $(d_\infty, e_\infty) \in \Gamma_{\text{SR}}$. In summary, any sequence (d_j, e_j) in Γ_{SR} has a convergent subsequence (d_{j_k}, e_{j_k}) whose limit (d_∞, e_∞) is also in Γ_{SR}. Therefore, Γ_{SR} is a compact set.

Step 2: We will show that there exists an optimal policy $\pi_{opt} \in \Pi$ of Problem (4) *such that the sequence $(a_{n,\pi_{opt}}, b_{n,\pi_{opt}})$ associated with policy π_{opt} has at least one limit point in Γ_{SR}.*

Since the sequence $(Y_0, Y_1, \ldots)$ is a Markov chain, the observation Y_{i+1} depends only on the immediately preceding state Y_i and not on the history state and control $Y_0, \ldots, Y_{i-1}$, $Z_0, \ldots, Z_{i-1}$. Therefore, Y_i is the *sufficient statistic* [13, p. 252] for solving Problem (4). This tells us that there exists an optimal policy $\pi_{\text{opt}} = (Z_0, Z_1, \ldots) \in \Pi$ of Problem (4) in which the control action Z_i is determined based on only Y_i, but not the history state and control $Y_0, \ldots, Y_{i-1}$, $Z_0, \ldots, Z_{i-1}$ [13]. We will show that the sequence $(a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}})$ associated with this policy π_{opt} has at least one limit point in Γ_{SR}.

It is known that Z_i takes values in the standard Borel space $(\mathbb{R}, \mathcal{R})$, where $\mathcal{R}$ is the Borel σ-field. According to [10, Thoerem 5.1.9], for each i there exists a conditional probability measure $p'_i(y, A)$ such that $p'_i(y, A) = \Pr(Z_i \in A | Y_i = y)$ for almost all y. One can use this conditional probability $p'_i(y, A)$ to generate a stationary randomized policy $\pi'_{i,\text{SR}} \in \Pi_{\text{SR}}$. Then, the one-stage expectation $(\mathbb{E}[q(Y_i, Z_i, Y_{i+1})] - \overline{g}_{\text{opt}}\mathbb{E}[Y_i + Z_i], \mathbb{E}[Y_i + Z_i])$ is exactly the limit point generated by the stationary randomized policy $\pi'_{i,\text{SR}}$. Thus, $(\mathbb{E}[q(Y_i, Z_i, Y_{i+1})] - \overline{g}_{\text{opt}}\mathbb{E}[Y_i + Z_i], \mathbb{E}[Y_i + Z_i]) \in \Gamma_{\text{SR}}$ for all $i = 0, 1, 2, \ldots$ Using (5), (6), and the fact that Γ_{SR} is convex, we can obtain $(a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}}) \in \Gamma_{\text{SR}}$ for all $n = 1, 2, 3 \ldots$ In other words, the sequence $(a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}})$ is within Γ_{SR}. Since Γ_{SR} is a compact set, the sequence $(a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}})$ must have a convergent subsequence, whose limit is in Γ_{SR}.

Step 3: Let $(a^*, b^*) \in \Gamma_{SR}$ be one limit point of the sequence $(a_{n,\pi_{opt}}, b_{n,\pi_{opt}})$ associated with policy π_{opt}. We will show that $a^* \leq 0$ and $b^* \geq T_{\min}$.

Policy π_{opt} is feasible for Problem (4) and meanwhile achieves the optimum objective value $\overline{g}_{\text{opt}}$. Hence,

$$\limsup_{n\to\infty} \frac{c_{n,\pi_{\text{opt}}}}{b_{n,\pi_{\text{opt}}}} = \overline{g}_{\text{opt}}, \tag{24}$$

$$\liminf_{n\to\infty} b_{n,\pi_{\text{opt}}} \geq T_{\min}, \tag{25}$$

where

$$c_{n,\pi_{\text{opt}}} \triangleq \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1} q(Y_i, Z_i, Y_{i+1})\right].$$

By (6), $b_{n,\pi_{\text{opt}}}$ is upper bounded by

$$b_{n,\pi_{\text{opt}}} \leq \frac{1}{n}\mathbb{E}\left[\sum_{i=0}^{n-1}(M + Y_{i+1})\right] = M + \mathbb{E}[Y] < \infty.$$

Hence, by (5), we have

$$\begin{aligned} a_{n,\pi_{\text{opt}}} &= c_{n,\pi_{\text{opt}}} - \overline{g}_{\text{opt}} b_{n,\pi_{\text{opt}}} \\ &\leq \max\{c_{n,\pi_{\text{opt}}} - \overline{g}_{\text{opt}} b_{n,\pi_{\text{opt}}}, 0\} \\ &= \max\{\frac{c_{n,\pi_{\text{opt}}}}{b_{n,\pi_{\text{opt}}}} - \overline{g}_{\text{opt}}, 0\} b_{n,\pi_{\text{opt}}} \\ &\leq \max\{\frac{c_{n,\pi_{\text{opt}}}}{b_{n,\pi_{\text{opt}}}} - \overline{g}_{\text{opt}}, 0\}(M + \mathbb{E}[Y]). \end{aligned}$$

Taking the $\limsup$ in this inequality and using (24), yields

$$\limsup_{n\to\infty} a_{n,\pi_{\text{opt}}} \leq 0. \tag{26}$$

Because (a^*, b^*) is one limit point of $(a_{n,\pi_{\text{opt}}}, b_{n,\pi_{\text{opt}}})$, we have

$$a^* \leq \limsup_{n\to\infty} a_{n,\pi_{\text{opt}}}, b^* \geq \liminf_{n\to\infty} b_{n,\pi_{\text{opt}}}. \tag{27}$$

By (25)-(27), we have $a^* \leq 0$ and $b^* \geq T_{\min}$.

Step 4: We will show that there exists a stationary randomized policy that is optimal for Problems (4) *and* (7), *and thus $\overline{g}_{SR} = \overline{g}_{opt}$.* By the definition of Γ_{SR}, $(a^*, b^*) \in \Gamma_{\text{SR}}$ must be the limit point generated by a stationary randomized policy $\pi^* \in \Pi_{\text{SR}}$. Let (Y, Z^*, Y') be a random vector with the stationary distribution of policy π^*. Then, (23) implies

$$(a^*, b^*) = \left(\mathbb{E}[q(Y, Z^*, Y')] - \overline{g}_{\text{opt}}\mathbb{E}[Y + Z^*], \mathbb{E}[Y + Z^*]\right).$$

Using $a^* \leq 0$ and $b^* \geq T_{\min}$, we can obtain

$$\mathbb{E}[q(Y, Z^*, Y')] - \mathbb{E}[Y + Z^*]\overline{g}_{\text{opt}} \leq 0, \tag{28}$$

$$\mathbb{E}[Y + Z^*] \geq T_{\min}. \tag{29}$$

By (28) and $\mathbb{E}[Y + Z^*] > 0$, we have

$$\frac{\mathbb{E}[q(Y, Z^*, Y')]}{\mathbb{E}[Y + Z^*]} \leq \overline{g}_{\text{opt}}.$$

Further, the inequality (29) suggests that the stationary randomized policy π^* is feasible for Problem (7). Hence,

$$\frac{\mathbb{E}[q(Y, Z^*, Y')]}{\mathbb{E}[Y + Z^*]} \geq \overline{g}_{\text{SR}}.$$

Therefore, $\overline{g}_{\text{SR}} \leq \overline{g}_{\text{opt}}$. This and (20) suggest that

$$\frac{\mathbb{E}[q(Y, Z^*, Y')]}{\mathbb{E}[Y + Z^*]} = \overline{g}_{\text{SR}} = \overline{g}_{\text{opt}}.$$

This completes the proof.

Appendix B
Proof of Theorem 2

Consider an arbitrarily chosen stationary randomized policy $\pi_1 \in \Pi_{\text{SR}}$ that is feasible for problem (7). We will show that there exists a feasible stationary deterministic policy that is no worse than policy π_1.

For any y, we can use the conditional probability $p(y, A)$ associated with policy π_1 to compute the conditional expectation $\mathbb{E}[Z|Y = y]$ by

$$\mathbb{E}[Z|Y = y] = \int_0^M z p(y, dz).$$

Since the conditional expectation $\mathbb{E}[Z|Y]$ is unique w.p.1 [10, Section 5.1], there is a deterministic function $z(\cdot)$ such that

$z(y) = \mathbb{E}[Z|Y = y]$ w.p.1. Consider the set $\Lambda \subset \Pi_{\text{SR}}$ of all stationary randomized policies that satisfy $\mathbb{E}[Z|Y = y] = z(y)$ w.p.1. Then, the stationary randomized policy π_1 is in Λ. It is also easy to show that the stationary deterministic policy $(Z_i = z(Y_i), i = 1, 2, \ldots)$ is also in Λ.

Using the iterated expectation, for any policy in Λ

$$\mathbb{E}[Y + Z] = \mathbb{E}\big[Y + \mathbb{E}[Z|Y]\big] = \mathbb{E}\big[Y + z(Y)\big]. \tag{30}$$

Because $\pi_1 \in \Lambda$ is feasible for problem (7), any policy in Λ is feasible for problem (7).

Since $q(y, z, y')$ is the integral of a non-decreasing function g, it is easy to show that the function $q(y, \cdot, y')$ is convex. For any policy $\pi \in \Lambda$, Jensen's inequality tells us that

$$\begin{aligned}
&\mathbb{E}[q(Y, Z, Y')|Y, Y'] \\
&\geq q(Y, \mathbb{E}[Z|Y, Y'], Y') \\
&= q(Y, \mathbb{E}[Z|Y], Y') \\
&= q(Y, z(Y), Y'), \ \text{(w.p.1)},
\end{aligned} \tag{31}$$

where (31) is due to the fact that Z is determined based on Y, but not Y'. Taking the expectation over (Y, Y'), yields

$$\mathbb{E}[q(Y, z(Y), Y')] \leq \mathbb{E}[q(Y, Z, Y')]$$

for any policy $\pi \in \Lambda$, where equality holds if $Z = z(Y)$. This and (30) suggest that the stationary deterministic policy $(Z_i = z(Y_i), i = 1, 2, \ldots)$ achieves the smallest objective value for problem (7) among all policies in Λ. In conclusion, for any feasible stationary randomized policy $\pi_1 \in \Pi_{\text{SR}}$, we can find a feasible stationary deterministic policy that is no worse than policy π_1. This completes the proof.

Appendix C
Proof of Lemma 1

We need the following lemma:

Lemma 4: If $l : \mathbb{R} \to \mathbb{R}$ is a convex function, then the functional $w : L^2(\mu_Y) \to \mathbb{R}$ defined by

$$w(z) = \int_0^\infty l(z(y)) d\mu_Y(y) \tag{32}$$

is also convex.

Proof: For any $\lambda \in [0, 1]$ and $z_1, z_2 \in L^2(\mu_Y)$, we have

$$\begin{aligned}
&w(\lambda z_1 + (1 - \lambda) z_2) \\
=& \int_0^\infty l(\lambda z_1(y) + (1 - \lambda) z_2(y)) d\mu_Y(y) \\
\leq& \int_0^\infty \left[\lambda l(z_1(y)) + (1 - \lambda) l(z_2(y))\right] d\mu_Y(y) \\
=& \lambda w(z_1) + (1 - \lambda) w(z_2).
\end{aligned} \tag{33}$$

By this, $w(z)$ is convex. ■

We now prove Lemma 1. Since $q(y, z, y')$ is the integral of a non-decreasing function g, it is easy to show that the function $q(y, \cdot, y')$ is non-negative and convex. Hence, the conditional expectation $\mathbb{E}\left[q(y, \cdot, Y')|Y = y\right]$ is also convex. We can obtain

$$\begin{aligned}
&\mathbb{E}\left[q(Y, z(Y), Y')\right] \\
=& \int_0^\infty \mathbb{E}\left[q(y, z(y), Y')|Y = y\right] d\mu_Y(y).
\end{aligned} \tag{34}$$

According to Lemma 4, $\mathbb{E}\left[q(Y, z(Y), Y')\right]$ is a convex functional of z and $\mathbb{E}[Y + z(Y)]$ is an affine functional of z. It is known that the ratio of a non-negative convex functional and positive affine functional is quasi-convex [28, p. 103]. Hence, $h(z)$ is quasi-convex, which completes the proof.

Appendix D
Proof of Theorem 3

We use the Lagrangian duality approach to solve Problem (10). The Lagrangian of Problem (10) is

$$\begin{aligned}
&L(z, \zeta, \gamma, \tau) \\
=& \int_0^\infty \mathbb{E}\left[q(y, z(y), Y')|Y = y\right] d\mu_Y(y) \\
&- c \int_0^\infty [y + z(y)] d\mu_Y(y) \\
&+ \zeta \left[T_{\min} - \int_0^\infty [y + z(y)] d\mu_Y(y)\right] \\
&- \int_0^\infty \gamma(y) z(y) d\mu_Y(y) + \int_0^\infty \tau(y)(z(y) - M) d\mu_Y(y) \\
=& \int_0^\infty \Big\{\mathbb{E}\left[q(y, z(y), Y')|Y = y\right] - (c + \zeta)[y + z(y)] \\
&- \gamma(y) z(y) + \tau(y)\Big[z(y) - M\Big]\Big\} d\mu_Y(y) + \zeta T_{\min}.
\end{aligned} \tag{35}$$

Since Problem (10) is feasible, all constraints are affine, the refined Slater's condition [28, Sec. 5.2.3] is satisfied. According to [29, Proposition 3.3.2] and [30, pp. 70-72], the Karush-Kuhn-Tucker (KKT) theorem remains valid for the Lebesgue space $L^2(\mu_Y)$. Hence, if a vector (z, ζ, γ, τ) satisfies the KKT conditions (36)-(42), it is an optimal solution to (10). The KKT conditions are given by

$$z = \min_{x \in L^2(\mu_Y)} L(x, \zeta, \gamma, \tau), \tag{36}$$

$$\zeta \geq 0, \int_0^\infty [y + z(y)] d\mu_Y(y) \geq T_{\min}, \tag{37}$$

$$\gamma(y) \geq 0, z(y) \geq 0, \forall\ y \geq 0, \tag{38}$$

$$\tau(y) \geq 0, z(y) \leq M, \forall\ y \geq 0, \tag{39}$$

$$\zeta \left[T_{\min} - \int_0^\infty [y + z(y)] d\mu_Y(y)\right] = 0, \tag{40}$$

$$\gamma(y) z(y) = 0, \forall\ y \geq 0, \tag{41}$$

$$\tau(y)(z(y) - M) = 0, \forall\ y \geq 0. \tag{42}$$

We now solve the KKT conditions (36)-(42) by using the calculus of variations. The one-sided Gâteaux derivative (similar to sub-gradient in finite dimensional space) of a functional h in the direction of $w \in L^2(\mu_Y)$ at $z \in L^2(\mu_Y)$ is defined as

$$\delta h(z; w) \triangleq \lim_{\epsilon \to 0^+} \frac{h(z + \epsilon w) - h(z)}{\epsilon}. \tag{43}$$

If h is a function on $\mathbb{R}$, then (43) becomes the common one-sided derivative. Let $l(z, y, \zeta, \gamma, \tau)$ denote the integrand in (35), and $r(z, y) = \mathbb{E}\left[q(y, z(y), Y')|Y = y\right]$. According to Lemma 4, the function $q(y, z, y')$ and functionals $r(z, y)$, $l(z, y, \zeta, \gamma, \tau)$, and $L(z, \zeta, \gamma, \tau)$ are all convex in z. Therefore, their one-sided Gâteaux derivatives with respect to z exist [31,

p. 709]. Since $g(x)$ is right-continuous, for any given (y, y'), the one-sided derivative $\delta q(y, z; w, y')$ of function $q(y, z, y')$ with respect to z is given by

$$\begin{aligned}
&\delta q(y, z; w, y') \\
&= \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\left\{q(y, z+\epsilon w, y') - q(y, z, y')\right\} \\
&= \begin{cases} \lim\limits_{x \to z^+} g(y+x+y')w, & \text{if } w \geq 0; \\ \lim\limits_{x \to z^-} g(y+x+y')w, & \text{if } w < 0. \end{cases}
\end{aligned}$$

Next, consider the one-sided Gaîeaux derivative $\delta r(z; w, y)$ of functional $r(z, y)$. Since the function $q(y, \cdot, y')$ is convex, the function $\epsilon \to [q(y, z+\epsilon w, y') - q(y, z, y')]/\epsilon$ is non-decreasing and bounded from above in a neighborhood of 0. By using the monotone convergence theorem [10, Theorem 1.5.6], we can interchange the limit and integral operators in $\delta r(z; w, y)$ such that

$$\begin{aligned}
&\delta r(z; w, y) \\
&= \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \mathbb{E}\left[q(y, z(y)+\epsilon w(y), Y') - q(y, z(y), Y') | Y = y\right] \\
&= \mathbb{E}\left[\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\left\{q(y, z(y)+\epsilon w(y), Y') - q(y, z(y), Y')\right\} \middle| Y = y\right] \\
&= \mathbb{E}\left[\lim_{x \to z(y)^+} g(y+x+Y')w(y)1_{\{w(y)>0\}}\right. \\
&\quad \left. + \lim_{x \to z(y)^-} g(y+x+Y')w(y)1_{\{w(y)<0\}} \middle| Y = y\right] \\
&= \lim_{x \to z(y)^+} \mathbb{E}\left[g(y+x+Y')w(y)1_{\{w(y)>0\}} \middle| Y = y\right] \\
&\quad + \lim_{x \to z(y)^-} \mathbb{E}\left[g(y+x+Y')w(y)1_{\{w(y)<0\}} \middle| Y = y\right], \quad (44)
\end{aligned}$$

where 1_E is the indicator function of event E. By using the monotone convergence theorem again, we have

$$\begin{aligned}
&\delta L(z; w, \zeta, \gamma, \tau) \\
&= \int_0^\infty \delta l(z; w, y, \zeta, \gamma, \tau) d\mu_Y(y) \\
&= \int_0^\infty \delta r(z; w, y) d\mu_Y(y) \\
&\quad + \int_0^\infty \left[-(c+\zeta) - \gamma(y) + \tau(y)\right] w(y) d\mu_Y(y). \quad (45)
\end{aligned}$$

According to [31, p. 710], z is an optimal solution to (36) if and only if

$$\delta L(z; w, \zeta, \gamma, \tau) \geq 0, \quad \forall\, w \in L^2(\mu_Y). \quad (46)$$

Since $w(\cdot)$ is an arbitrary function in $L^2(\mu_Y)$, considering positive functions $w(y) > 0$, we can obtain from (44)-(46) that for each $y \geq 0$, $z(y)$ must satisfy

$$\lim_{x \to z(y)^+} \mathbb{E}\left[g(y+x+Y') | Y = y\right] - (c+\zeta) - \gamma(y) + \tau(y) \geq 0. \quad (47)$$

Similarly, considering negative functions $w(y) < 0$, we can obtain that for each $y \geq 0$, $z(y)$ must satisfy

$$\lim_{x \to z(y)^-} \mathbb{E}\left[g(y+x+Y') | Y = y\right] - (c+\zeta) - \gamma(y) + \tau(y) \leq 0. \quad (48)$$

Because $g(\cdot)$ is non-decreasing, we can obtain from (47) and (48) that for each $y \geq 0$, $z(y)$ needs to satisfy

$$\mathbb{E}\left[g(y+x+Y') | Y = y\right] - (c+\zeta) - \gamma(y) + \tau(y) \geq 0 \quad (49)$$

for all $x > z(y)$, and

$$\mathbb{E}\left[g(y+x+Y') | Y = y\right] - (c+\zeta) - \gamma(y) + \tau(y) \leq 0 \quad (50)$$

for all $x < z(y)$.

We solve the optimal primal solution $z(\cdot)$ by considering the following three cases:

Case 1: $\gamma(y) = \tau(y) = 0$. The solutions to (49) and (50) may not be unique, if function g is non-strictly increasing. In particular, there may exist an interval $[a(y), b(y)]$ such that each $z(y) \in [a(y), b(y)]$ satisfies (49) and (50) for some y. In this case, we choose the largest possible solution of $z(y)$ to make sure that the constraint (9) is satisfied. The largest solution satisfying (47) and (48) is given by

$$z(y) = \sup\{x \in [0, M] : \mathbb{E}\left[g(y+x+Y') | Y = y\right] \leq c + \zeta\}, \quad \forall\, y \geq 0,$$

which is exactly (11).

Case 2: $\gamma(y) > 0$. By (41), we have $z(y) = 0$.

Case 3: $\tau(y) > 0$. By (42), we have $z(y) = M$.

In summary, the optimal primal solution $z(\cdot)$ is given by (11).

Next, we find the optimal dual variable ζ. By (37) and (40), the optimal ζ satisfies

$$\zeta = 0, \mathbb{E}\left[Y + z(Y)\right] \geq T_{\min} \quad (51)$$

or

$$\zeta > 0, \mathbb{E}\left[Y + z(Y)\right] = T_{\min}, \quad (52)$$

where $\mathbb{E}\left[Y + z(Y)\right]$ is determined by the optimal primal solution (11). Since $\mathbb{E}\left[Y + z(Y)\right]$ is non-decreasing in ζ, we can use bisection to search for the optimal ζ. By this, an optimal solution to (10) is obtained for any given c. Finally, according to Sections 4.2.5 and 11.4 of [28], the optimal c is solved by an outer-layer bisection search. Therefore, an optimal solution to Problem (8) is given by Algorithm 1. This completes the proof.

APPENDIX E
PROOF OF LEMMA 2

Let us rewrite the functional h_1 as

$$h_1(z) = \frac{\int_0^\infty [y + z(y)]^2 d\mu_Y(y)}{\int_0^\infty [y + z(y)] d\mu_Y(y)}.$$

We need to prove that the functional h_1 is convex when restricted to any line that intersects its domain. For any

$w \in L^2(\mu_Y)$, consider the function $u : \mathbb{R} \to \mathbb{R}$ defined as

$$u(\epsilon) = \frac{\int_0^\infty [z(y) + \epsilon w(y) + y]^2 d\mu_Y(y)}{\int_0^\infty [z(y) + \epsilon w(y) + y] d\mu_Y(y))}$$

with domain

$$\textbf{dom } u = \{\epsilon : z(y) + \epsilon w(y) \in [0, M], \ \forall y \geq 0, \epsilon \in \mathbb{R}\}.$$

Since the function $\epsilon \to [z(y)+\epsilon w(y)+y]^2$ is convex, the function $x \to \{[z(y)+(\epsilon+x)w(y)+y]^2 - [z(y)+\epsilon w(y)+y]^2\}/x$ is non-decreasing and bounded from above in a neighborhood of 0. By using monotone convergence theorem [10, Theorem 1.5.6], we can interchange the limit and integral operators such that

$$\frac{d}{d\epsilon}\int_0^\infty [z(y) + \epsilon w(y) + y]^2 d\mu_Y(y) = \int_0^\infty 2[z(y) + \epsilon w(y) + y]w(y) d\mu_Y(y).$$

Similarly,

$$\frac{d}{d\epsilon}\int_0^\infty [z(y) + \epsilon w(y) + y] d\mu_Y(y) = \int_0^\infty w(y) d\mu_Y(y).$$

By this, we have

$$\frac{du}{d\epsilon} = \frac{\int_0^\infty 2[z(y) + \epsilon w(y) + y]w(y) d\mu_Y(y)}{\int_0^\infty [z(y) + \epsilon w(y) + y] d\mu_Y(y)} - \frac{\int_0^\infty [z(y)+\epsilon w(y) + y]^2 d\mu_Y(y) \int_0^\infty w(y) d\mu_Y(y)}{\left[\int_0^\infty [z(y) + \epsilon w(y) + y] d\mu_Y(y)\right]^2}. \quad (53)$$

After some additional manipulations, we can obtain

$$\frac{d^2u}{d^2\epsilon} = \frac{2\left[\int_0^\infty [y+z(y)] d\mu_Y(y) \int_0^\infty w(y) d\mu_Y(y)\right]^2}{\left[\int_0^\infty [z(y) + \epsilon w(y) + y] d\mu_Y(y)\right]^3} \times \int_0^\infty \left[\frac{y+z(y)}{\int_0^\infty [y+z(y)] d\mu_Y(y)} - \frac{w(y)}{\int_0^\infty w(y) d\mu_Y(y)}\right]^2 d\mu_Y(y).$$

Since $z(y)+\epsilon w(y) \geq 0$ for all y on **dom** u, we have $\frac{d^2u}{d^2\epsilon} \geq 0$. Hence, the function u is convex for all $w \in L^2(\mu_Y)$. By this, the functional h_1 is convex, which completes the proof.

Appendix F
Proof of Theorem 4

The Lagrangian of Problem (13) is determined as

$$L_1(z, \zeta, \gamma, \tau) = \frac{\int_0^\infty [y + z(y)]^2 d\mu_Y(y)}{2\int_0^\infty [y + z(y)] d\mu_Y(y)} + \zeta\left[T_{\min} - \int_0^\infty [y + z(y)] d\mu_Y(y)\right] - \int_0^\infty \gamma(y)z(y) d\mu_Y(y) + \int_0^\infty \tau(y)(z(y) - M) d\mu_Y(y),$$

where $\zeta \in \mathbb{R}$, $\gamma, \tau \in L^2(\mu_Y)$ are dual variables. According to [29, Proposition 3.3.2] and [30, pp. 70-72], the KKT theorem remains valid for the Lebesgue space $L^2(\mu_Y)$. Hence, if a vector (z, ζ, γ, τ) satisfies the KKT conditions (54)-(60), it is an optimal solution to (13). The KKT conditions are given by:

$$z = \min_{x \in L^2(\mu_Y)} L_1(x, \zeta, \gamma, \tau), \quad (54)$$

$$\zeta \geq 0, \int_0^\infty [y + z(y)] d\mu_Y(y) \geq T_{\min}, \quad (55)$$

$$\gamma(y) \geq 0, z(y) \geq 0, \forall\ y \geq 0, \quad (56)$$

$$\tau(y) \geq 0, z(y) \leq M, \forall\ y \geq 0, \quad (57)$$

$$\zeta\left[T_{\min} - \int_0^\infty [y + z(y)] d\mu_Y(y)\right] = 0, \quad (58)$$

$$\gamma(y)z(y) = 0, \forall\ y \geq 0, \quad (59)$$

$$\tau(y)(z(y) - M) = 0, \forall\ y \geq 0. \quad (60)$$

We now solve the KKT conditions by using the calculus of variations. For any fixed (ζ, γ, τ), the Gâteaux derivative of the Lagrange L_1 in the direction of $w \in L^2(\mu_Y)$ at $z \in L^2(\mu_Y)$ is defined as

$$\delta L_1(z; w, \zeta, \gamma, \tau) \triangleq \lim_{\epsilon \to 0} \frac{L_1(z + \epsilon w, \zeta, \gamma, \tau) - L_1(z, \zeta, \gamma, \tau)}{\epsilon}.$$

Similar to the derivations of (53), we can obtain

$$\delta L_1(z; w, \zeta, \gamma, \tau) = \int_0^\infty \left[\frac{y + z(y)}{\int_0^\infty [y + z(y)] d\mu_Y(y)} - \frac{\int_0^\infty [y + z(y)]^2 d\mu_Y(y)}{2\left[\int_0^\infty [y + z(y)] d\mu_Y(y)\right]^2} - \zeta - \gamma(y) + \tau(y)\right] w(y) d\mu_Y(y), \quad \forall\ w \in L^2(\mu_Y).$$

Then, $z(\cdot)$ is an optimal solution to (54) if and only if [31, p. 710]

$$\delta L_1(z; w, \zeta, \gamma, \tau) \geq 0, \quad \forall\ w \in L^2(\mu_Y).$$

By $\delta L_1(z; w, \zeta, \gamma, \tau) = -\delta L_1(z; -w, \zeta, \gamma, \tau)$, we deduce

$$\delta L_1(z; w, \zeta, \gamma, \tau) = 0, \quad \forall\ w \in L^2(\mu_Y).$$

Since $w(\cdot)$ is arbitrary, we have

$$\frac{y + z(y)}{\int_0^\infty [y + z(y)] d\mu_Y(y)} - \frac{\int_0^\infty [y + z(y)]^2 d\mu_Y(y)}{2\left[\int_0^\infty [y + z(y)] d\mu_Y(y)\right]^2} - \zeta - \gamma(y) + \tau(y) = 0, \ \forall\ y \geq 0. \quad (61)$$

For notational simplicity, let us define

$$\beta \triangleq \zeta \int_0^\infty [y+z(y)] d\mu_Y(y) + \frac{\int_0^\infty [y + z(y)]^2 d\mu_Y(y)}{2\int_0^\infty [y+z(y)] d\mu_Y(y)}. \quad (62)$$

The optimal primal solution $z(\cdot)$ is obtained by considering the following three cases:

Case 1: If $\gamma(y) = \tau(y) = 0$, then by (61) and (62), we obtain $z(y) = \beta - y$. In this case, we require $\beta - y \in [0, M]$ by (56) and (57).

Case 2: If $\gamma(y) > 0$, then by (59), $z(y) = 0$.

Case 3: If $\tau(y) > 0$, then by (60), $z(y) = M$.

In summary, the optimal primal solution $z(\cdot)$ is given by (15).

The optimal dual variable β is obtained by considering two cases:

Case 1: $\zeta > 0$. Then, (58) and (62) imply that

$$\int_0^{\infty}[y+z(y)]d\mu_Y(y)=T_{\min}, \beta \geq \frac{\int_0^{\infty}[y+z(y)]^2 d\mu_Y(y)}{2\int_0^{\infty}[y+z(y)]d\mu_Y(y)}. \quad (63)$$

Case 2: $\zeta = 0$. Then, (55) and (62) imply that

$$\int_0^{\infty}[y+z(y)]d\mu_Y(y)\geq T_{\min}, \beta = \frac{\int_0^{\infty}[y+z(y)]^2 d\mu_Y(y)}{2\int_0^{\infty}[y+z(y)]d\mu_Y(y)}. \quad (64)$$

Combining (63) and (64), yields

$$\int_0^{\infty}[y+z(y)]d\mu_Y(y)=\max\left(T_{\min}, \frac{\int_0^{\infty}[z(y)+y]^2 d\mu_Y(y)}{2\beta}\right).$$

Then, (16) is obtained by subscribing (15) into this.

The solution in (15) and (16) satisfies the KKT conditions (54)-(60), and is thus an optimal solution to the convex optimization problem (13). This completes the proof.

Appendix G
Proof of Lemma 3

1). When the correlation coefficient between Y_i and Y_{i+1} is -1, $Y+Y'$ is equal a constant value with probability one. Choosing $z(y)=0$, $\zeta=0$, $c=g(Y+Y')$, $\gamma(y)=\tau(y)=0$, one can show that the KKT conditions (36)-(42) are satisfied.

2). If the Y_i's are equal to a constant value, $Y+Y'$ is equal a constant value. The remaining proof follows from part 1).

3). When $g(\cdot)$ is a constant function, all policies are optimal. This completes the proof.

References

[1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. INFOCOM*, Apr. 2012, pp. 2731–2735.

[2] ——, "Status updates through queues," in *Conf. on Info. Sciences and Systems*, Mar. 2012.

[3] R. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE Int. Symp. Inform. Theory*, Jul. 2012.

[4] C. Kam, S. Kompella, and A. Ephremides, "Age of information under random updates," in *Proc. IEEE Int. Symp. Inform. Theory*, 2013.

[5] ——, "Effect of message transmission diversity on status age," in *Proc. IEEE Int. Symp. Inform. Theory*, June 2014, pp. 2411–2415.

[6] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inform. Theory*, June 2014, pp. 1583–1587.

[7] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *Proc. IEEE Int. Symp. Inform. Theory*, 2015.

[8] R. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE Int. Symp. Inform. Theory*, 2015.

[9] T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. Info. Theory and Appl. Workshop*, Feb. 2015.

[10] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Univerisity Press, 2010.

[11] S. M. Ross, *Applies Probability Models with Optimization Applications*. San Francisco, CA: Holden-Day, 1970.

[12] H. Mine and S. Osaki, *Markovian Decision Processes*. New York: Elsevier, 1970.

[13] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA: Athena Scientific, 2005, vol. 1.

[14] D. Hayman and M. Sobel, *Stochastic models in Operations Research, Volume II: Stochastic Optimizations*. New York: McGraw-Hill, 1984.

[15] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[16] M. J. Neely, "Dynamic optimization and learning for renewal systems," *IEEE Trans. Auto. Control*, vol. 58, no. 1, pp. 32–46, Jan 2013.

[17] A. Federgruen, P. J. Schweitzer, and H. C. Tijms, "Denumerable undiscounted semi-Markov decision processes with unbounded rewards," *Mathematics of Operations Research*, vol. 8, no. 2, pp. 298–313, 1983.

[18] D. Klabjan and D. Adelman, "Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming," *SIAM J. Control Optim.*, vol. 44, no. 6, pp. 2104–2122, Jan. 2006.

[19] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint," *Journal of Mathematical Analysis and Applications*, vol. 112, no. 1, pp. 236 – 252, 1985.

[20] ——, "Time-average optimal constrained semi-Markov decision processes," *Adv. in Appl. Probab.*, vol. 18, no. 2, pp. 341–359, 1986.

[21] E. Feinberg, "Constrained semi-Markov decision processes with average rewards," *Zeitschrift für Operations Research*, vol. 39, no. 3, pp. 257–288, 1994.

[22] M. Baykal-Gürsoy and K. Gürsoy, "Semi-Markov decision processes: nonstandard criteria," *Probability in the Engineering and Informational Sciences*, vol. 21, pp. 635–657, Oct. 2007.

[23] E. A. Feinberg, "Constrained discounted semi-Markov decision processes," in *Markov Processes and Controlled Markov Chains*, Z. Hou, J. A. Filar, and A. Chen, Eds. Kluwer Academic Publishers, 2002.

[24] E. Altman, *Constrained Markov decision processes*. Boca Raton, FL, USA: Chapman & Hall/CRC, 1999.

[25] C. Derman, *Finite State Markovian Decision Processes*. Orlando, FL, USA: Academic Press, Inc., 1970.

[26] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.

[27] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill Book Co., 1987.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univerisity Press, 2004.

[29] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*. Norwell, MA: Kluwer Academic Publisher, 2000.

[30] G. G. Magaril-ll'yaev and V. M. Tikhomirov, *Convex Analysis: Theory and Applications*. American Mathematical Society, 2003.

[31] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.