Battle between Rate and Error in Minimizing Age of Information

Guidan Yao
Department of ECE
The Ohio State University
Columbus, OH, USA
yao.539@osu.edu

Ahmed M. Bedewy
Department of ECE
The Ohio State University
Columbus, OH, USA
bedewy.2@osu.edu

Ness B. Shroff
Department of ECE and CSE
The Ohio State University
Columbus, OH, USA
shroff.11@osu.edu

ABSTRACT

In this paper, we consider a status update system, in which update packets are sent to the destination via a wireless medium that allows for multiple rates, where a higher rate also naturally corresponds to a higher error probability. The data freshness is measured using age of information, which is defined as the age of the recent update at the destination. A packet that is transmitted with a higher rate, will encounter a shorter delay and a higher error probability. Thus, the choice of the transmission rate affects the age at the destination. In this paper, we design a low-complexity scheduler that selects between two different transmission rate and error probability pairs to be used at each transmission epoch. This problem can be cast as a Markov Decision Process. We show that there exists a threshold-type policy that is age-optimal. More importantly, we show that the objective function is quasi-convex or non-decreasing in the threshold, based on the system parameters values. This enables us to devise a *low-complexity algorithm* to minimize the age. These results reveal an interesting phenomenon: While choosing the rate with minimum mean delay is delay-optimal, this does not necessarily minimize the age.

CCS CONCEPTS

• **Networks** → **Network performance evaluation**; *Network performance analysis*;

KEYWORDS

Age of information, Markov decision process, Heterogenous transmission, Threshold-type policy

ACM Reference Format:
Guidan Yao, Ahmed M. Bedewy, and Ness B. Shroff. 2021. Battle between Rate and Error in Minimizing Age of Information. In *The Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '21), July 26–29, 2021, Shanghai, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3466772.3467041

1 INTRODUCTION

Age of information is a new metric that has attracted significant recent attention [3, 4, 10, 19]. This concept has been motivated by the rapid growth of real-time applications, e.g., health monitoring, automatic driving system, and agriculture automation, etc. For such applications, freshness of information updates is of utmost importance. However, traditional metric like delay cannot fully characterize the freshness of information updates. For example, if information is updated infrequently, then the updates are not fresh even though the delay is small. To this end, age of information, or simply the age, was proposed in [11] as a measure of the data freshness. Specifically, age of information is defined as the time elapsed since the generation of the most recently received status update.

There exist many works dealing with the age minimization problem. One class of works have focused on investigating optimal sampling and updating policy to minimize age of information. In [16], authors study the updating policy to minimize age in the presence of queuing delay. In [1, 7, 17, 20], sampling and updating polices are studied under energy constraint. In [1, 17], the authors assume that the channel is noiseless while in [20], authors assume that channel state is known a priori and updating cost is a function of channel state to ensure successful transmission. In [7], the authors consider transmission failure and investigate optimal sampling policy for age minimization under energy constraint. These works consider the effects of queueing delay, channel state, energy supply and minimize the age of information by controlling sampling and updating times, in which case they assume that there is only one transmission mode to transmit updates. However, in real systems, updates can be sent to a destination using heterogenous transmissions in terms of transmission delay and error probability. Two examples are provided as follows:

Error rate control: Error rate control scheme is managed at physical layer. In particular, the transmission rate is often adapted via modulation and coding scheme to meet a fixed target error rate [9]. It is known that choosing a lower target error rate corresponds to a lower transmission rate, and hence a longer transmission delay. On the other hand, a higher transmission rate (i.e., a shorter transmission delay) also corresponds to a higher transmission error probability of information delivery. Thus, there is a tradeoff between transmission delay and transmission error probability, both of which are affected by the target error rate.

Scheduling over channels in different frequencies: It is common that a device can access channels in different frequencies. For example, cellphones can access WiFi (high frequency) and LTE (low frequency). If updates are transmitted over such devices, then the age of information may experience different transmission properties based on the carrier frequency. In particular, it is known that it is hard for radio waves to distract obstacles that are in same or larger size than their wavelength. Thus, low-frequency radios (longer wavelength) are less vulnerable to blockage than high-frequency

MobiHoc '21, July 26–29, 2021, Shanghai, China

ACM ISBN 978-1-4503-8558-9/21/07…$15.00
https://doi.org/10.1145/3466772.3467041

radios, which implies that low frequency channels are more reliable than their high frequency counterparts. Of course, the higher frequency channels allow for higher rate (shorter delay) transmissions, resulting in a similar tradeoff between the transmission delay and transmission error probability.

The above examples clearly indicate that, transmission of updates can experience different transmission delays and error probabilities based on the choice of either target error rate or carrier frequency. In particular, a decrease in the transmission error probability will increase the chances of a successful update delivery (decrease age) while an increase in the transmission delay will increase the inter-delivery time (increase age). That is, the delay and error probability of a transmission mode affect the age in opposite direction. Thus, the key questions are: *when is it optimal to use the lower transmission rate with a lower error probability?*; *which variable plays a more important role in determining the optimal actions?*. To address these questions, we begin by investigating a status update system with two heterogenous transmissions and obtain the optimal transmission selection policy to minimize the average age. Studying the two-rate scenario provides us with some insights in the optimal policy for a more general multi-rate (multi-error probability) scenario, which is discussed in Section 5, and provides basis for our future work. Specifically, our contributions are outlined as follows:

- We investigate the optimal trade-off between transmission delay and error probability for minimizing the age. We show that there exists a stationary deterministic optimal transmission selection policy. Moreover, we show that the optimal transmission selection policy is of threshold-type in terms of the age (Theorem 4.1). In particular, we show that the optimal action is a non-increasing (non-decreasing) function of the age if the mean delay of the low rate transmission is smaller (larger) than that of the high rate transmission. This result was not anticipated: For example, in [8, 12], it was shown that the optimal delay policy chooses the server with minimum mean delay whenever it is available. With this, one may expect that using the transmission with higher mean delay would worsen the age performance. Surprisingly, however, we show that choosing the transmission with higher mean delay can sometimes improve the age performance.
- We derive the average cost as a function of the threshold with the aid of the state transition diagram. We then optimize the threshold to minimize the average cost function. In particular, although the optimization problem is non-convex, we are able to show that if the mean delay of the low rate transmission is smaller than that of the high rate transmission, the objective function is quasi-convex; otherwise, the optimal policy chooses higher rate transmission (Theorem 4.6). This enables us to devise a low-complexity algorithm to obtain the optimal policy.

The remainder of this paper is organized as follows. The system model is introduced in Section 2. In Section 3, we map the problem to an equivalent problem which can be regarded as an average cost MDP, and then formulate the MDP problem. In Section 4, we explore the structure of the optimal policy and properties of average cost function, and devise an efficient algorithm. In Section 5, we provide a disscusion on multi-rate scenario. In Section 6, we provide numerical results to verify our theoretical results.

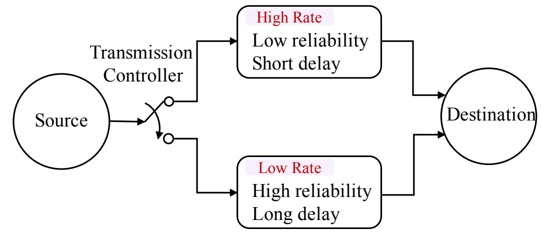


Figure 1: Model

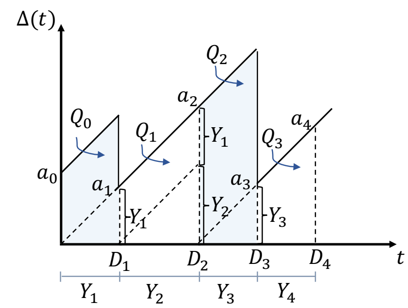


Figure 2: Example of Age Evolution

2 SYSTEM MODEL

We consider a status update system, in which update packets are sent to the destination via a wireless medium with varying transmission delay and error probability. The update packets are generated whenever the wireless medium becomes idle. We assume that there are two heterogenous transmissions available for updating, namely low rate and high rate transmissions. The high rate transmission offers a shorter transmission delay than low rate transmission; while low rate transmission offers more reliable transmission than high rate transmission. A decision maker chooses a transmission mode for each transmission opportunity. We denote the set of transmission modes as $\mathcal{U} \triangleq \{1, 2\}$, where 1 and 2 denote the low rate and high rate transmissions, respectively. We use $\mathcal{P} \triangleq \{p_j : 0 < p_j < 1, j \in \mathcal{U}\}$ and $\mathcal{D} \triangleq \{d_j : 0 < d_j < \infty, j \in \mathcal{U}\}$ to denote the set of transmission error probabilities and transmission delays, respectively. Transmission $j \in \mathcal{U}$ corresponds to transmission delay d_j and transmission error probability p_j. We assume that $d_1 > d_2$ and $p_1 < p_2$.

We use Y_i to denote the transmission delay of packet i, where $Y_i \in \mathcal{D}$. Let D_i denote the delivery time of packet i. Since updates are generated whenever the wireless medium becomes idle, D_i equals to the generation time of packet $i + 1$. Also, we have $D_i = \sum_{j=1}^{i} Y_j$. At any time t, the most recently received update packet is generated at time

$$U(t) = \max\{D_i : D_{i+1} \leq t\}. \quad (1)$$

Then, the *age of information*, or simply the *age* is defined as

$$\Delta(t) = t - U(t). \quad (2)$$

The age $\Delta(t)$ is a stochastic process that increases with t between update packets and is reset to a smaller value upon the successful delivery of a fresher packet. We suppose that the age $\Delta(t)$ is right-continuous. As shown in Fig. 2, packet 2 is sent at time D_1 and its delivery time is $D_2 = D_1 + Y_2$. Since this packet transmission fails, the age does not reset to a smaller value at D_2. Packet 3 transmission starts at D_2, which is successfully delivered at time D_3. Thus, the age increases linearly until it reaches to $\Delta(D_3^-) = Y_1 + Y_2 + Y_3$ before packet 3 is successfully sent, and then drops to $\Delta(D_3) = Y_3$ at D_3.

3 OPTIMIZATION PROBLEM

We use u_i to denote which transmission mode (low rate or high rate) is selected to transmit packet i, where $u_i \in \mathcal{U}$. In particular, if $u_i = 1$ (or $u_i = 2$), then packet i is transmitted using the low (or high) rate transmission, encounters transmission delay d_1 (or d_2), and is received successfully with probability $1 - p_1$ (or $1 - p_2$). A transmission selection policy π specifies a transmission selection decision for each stage[1]. For any policy π, we define the total average age as

$$\bar{\Delta}(\pi) = \limsup_{n\to\infty} \frac{\mathbb{E}[\int_0^{D_n} \Delta(t)dt]}{\mathbb{E}[D_n]}. \tag{3}$$

Our goal is to seek a transmission selection policy that solves the total average age minimization problem as follows:

$$\bar{\Delta}^* = \min_{\pi\in\Pi} \bar{\Delta}(\pi), \tag{4}$$

where $\bar{\Delta}^*$ denotes the optimal total average age. Let Π denote the set of all causal transmission selection policies, in which the policy $\pi \in \Pi$ depends on the history and current system state.

3.1 Equivalent Mapping of Problem (4)

We decompose the area under the curve $\Delta(t)$ into a sum of disjoint geometric parts as shown in Fig. 2. Observing the area in interval $[0, D_n]$, the area can be regarded as the concatenation of the areas Q_i. Then,

$$\int_0^{D_n} \Delta(t)dt = \sum_{i=0}^{n-1}[Q_i]. \tag{5}$$

Let a_i denote the age at time D_i, i.e., $a_i = \Delta(D_i)$. Then, Q_i can be expressed as

$$Q_i = a_i Y_{i+1} + \frac{1}{2}Y_{i+1}^2. \tag{6}$$

Recall that $D_n = \sum_{i=0}^{n-1} Y_{i+1}$. With Eq. (5) and Eq. (6), the total average age is expressed as

$$\limsup_{n\to\infty} \frac{\sum_0^{n-1} \mathbb{E}[a_i Y_{i+1} + \frac{1}{2}Y_{i+1}^2]}{\sum_{i=0}^{n-1} \mathbb{E}[Y_{i+1}]}. \tag{7}$$

With this, the optimal transmission selection problem for minimizing the total average age can be formulated as

$$\bar{\Delta}^* \triangleq \min_{\pi\in\Pi} \limsup_{n\to\infty} \frac{\sum_0^{n-1} \mathbb{E}[a_i Y_{i+1} + \frac{1}{2}Y_{i+1}^2]}{\sum_{i=0}^{n-1} \mathbb{E}[Y_{i+1}]}. \tag{8}$$

The problem is hard to solve in current form. Thus, we provide an equivalent mapping for it. A problem with parameter β is defined as follows:

$$p(\beta) \triangleq \min_{\pi\in\Pi} \limsup_{n\to\infty} \frac{1}{n}\sum_0^{n-1} \mathbb{E}[(a_i - \beta)Y_{i+1} + \frac{1}{2}Y_{i+1}^2]. \tag{9}$$

[1]Stage i corresponds to the duration from D_{i-1} to D_i.

Lemma 3.1. *The following statements are true:*
(i) $\bar{\Delta}^* \gtreqless \beta$ *if and only if* $p(\beta) \gtreqless 0$;
(ii) If $p(\beta) = 0$, *then the optimal transmission selection policies that solve* (8) *and* (9) *are identical.*

Proof. The proof is similar to that of Lemma 3.5 in [2]. The difference is that we use the boundedness of transmission delay while in [2], the boundedness of inter-sampling time is used. The detailed proof is provided in our technical report [18]. □

By Lemma 3.1, if $\beta = \bar{\Delta}^*$, then the optimal transmission selection policies that solve (8) and (9) are identical. With this, given β, we formulate the problem (9) as an infinite horizon average cost per stage MDP in Section 3.2 and show that the optimal policy for (9) is of threshold-type in Section 4.1. Since the value of β is arbitrary, we will be able to conclude that the optimal policy for (8) is of threshold-type. In addition, in Section 4.2, we are able to devise a low-complexity algorithm to obtain the optimal threshold.

3.2 The MDP problem of (9)

From [5], given β, problem (9) is equivalent to an average cost per stage MDP problem. The components of the MDP problem are described as follows:

- **States:** The system state at stage i is the age a_i. In this paper, we consider the state space $\mathcal{S} \triangleq \{a = ld_1 + vd_2 : l, v \in \{0, 1, \cdots\}\}$. If the initial state is outside $\mathcal{S}$, then eventually the state will enter $\mathcal{S}$ (with state d_1 or d_2); otherwise, a successful packet transmission never occurs. In fact, the maximal probability that no transmission succeeds after l stages is p_2^l, which decreases with number of stages l. After state enters $\mathcal{S}$, it will stay in $\mathcal{S}$ onwards (since transmission delay is either d_1 or d_2). Note that $\mathcal{S}$ is unbounded since successful packet transmission happens with certain probabilities.
- **Actions:** At delivery time D_{i-1}, the action that is chosen for stage i is $u_i \in \mathcal{U}$. The action u_i determines the transmission delay. For example, if $u_i = 1$, then the transmission delay at stage i is d_1.
- **Transition probabilities:** Given the current state a_i and action u_i at stage i, the transition probability to the state a_{i+1} at the stage $i + 1$ is defined as

$$P(a_{i+1} = a'|a_i = a, u_i = u) = \begin{cases} p_u & \text{if } a' = a + d_u, \\ 1 - p_u & \text{if } a' = d_u, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

- **Costs:** Given state a_i and action u_i at stage i, the cost at the stage is defined as

$$C(a_i, u_i) = (a_i - \beta)d_{u_i} + \frac{1}{2}d_{u_i}^2. \tag{11}$$

Given β, the average cost per stage under a transmission selection policy π is given by

$$J(\pi, \beta) \triangleq \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_\pi\left[\sum_{i=0}^{n-1} C(a_i, u_i)\right]. \tag{12}$$

Our objective is to find a transmission selection policy $\pi \in \Pi$ that minimizes the average cost per stage, which can be formulated as

Problem 1 (Average cost MDP)

$$\min_{\pi \in \Pi} J(\pi, \beta). \tag{13}$$

We say that a transmission selection policy π is *average-optimal* if it solves the problem in (13). A transmission selection policy is a sequence of decision rules, i.e., $\pi = (\zeta_1, \zeta_2, \cdots)$, where a decision rule ζ_i maps the history of states and actions, and the current state to an action. A transmission selection policy is called a *stationary deterministic* policy if $u_i = \zeta(a_i)$ for all $i \in \mathbb{N}^+$, where $\zeta : \mathcal{S} \to \mathcal{U}$ is a deterministic function. Stationary deterministic policies are the easiest to be implemented and evaluated. However, there may not exist a stationary deterministic policy that is average-optimal [5]. Next, we show that there exists a stationary deterministic transmission selection policy that is average-optimal. Moreover, we show that the optimal policy is of threshold-type.

4 STRUCTURE OF AVERAGE-OPTIMAL POLICY AND ALGORITHM DESIGN

In this section, we investigate the structure of the average-optimal policy and propose an efficient algorithm for the original problem (8).

4.1 Threshold Structure of Average-Optimal Policy

4.1.1 Threshold structure: The following theorem states that there exists a threshold-type stationary deterministic policy that is average-optimal. In particular, the problem is divided into two cases based on the relation between $d_1(1 - p_2)$ and $d_2(1 - p_1)$. Under these two cases, the threshold-type average-optimal policy shows opposite behaviors.

Theorem 4.1. *There exists a stationary deterministic average-optimal transmission selection policy that is of threshold-type. Specifically,*
(i) If $d_1(1 - p_2) \le d_2(1 - p_1)$, then the average-optimal policy is of the form $\pi^* = (\zeta^*, \zeta^*, \cdots)$, where

$$\zeta^*(a) = \begin{cases} 2 & \text{if } 0 \le a \le a_1^*, \\ 1 & \text{if } a_1^* < a, \end{cases} \tag{14}$$

where a_1^* denotes the age threshold.
(ii) If $d_1(1 - p_2) \ge d_2(1 - p_1)$, then the average-optimal policy is of the form $\pi^* = (\zeta^*, \zeta^*, \cdots)$, where

$$\zeta^*(a) = \begin{cases} 1 & \text{if } 0 \le a \le a_2^*, \\ 2 & \text{if } a_2^* < a, \end{cases} \tag{15}$$

where a_2^* denotes the age threshold.

Proof. Please see Section 4.1.2. □

Define the mean delay of transmission mode $j \in \mathcal{U}$ as

$$\bar{d}_j \triangleq \frac{d_j}{1 - p_j}. \tag{16}$$

By Theorem 4.1 (i), when the age exceeds a certain threshold, the optimal policy chooses the transmission with smaller mean delay. This result reveals an interesting phenomenon: While the transmission with minimum mean delay is the optimal decision for minimizing the average delay, this does not necessarily minimize the age. In particular, when the age is below a certain threshold, the average age is reduced by choosing a faster transmission that has a higher mean delay (i.e., a higher error probability). The reason is that if successful, the age remains low. If it fails, it provides an opportunity to generate a later packet that can be transmitted in a shorter period of time. In Section 4.2, based on Theorem 4.1 (ii), we will show that the averge-optimal policy under $d_1(1-p_2) \ge d_2(1-p_1)$ is to choose high rate transmission for each transmission opportunity. This is reasonable because both the delay and mean delay (including the impact of the error probability) of high rate transmission is shorter than that of low rate transmission.

4.1.2 Proof of Theorem 4.1. One way to investigate the average cost MDPs is to relate them to the discounted cost MDPs. To prove Theorem 4.1, we (i) address a discounted cost MDP, i.e., establish the existence of a stationary deterministic policy that solves the MDP and then study the structure of the optimal policy; and (ii) extend the results to the average cost MDP problem in (13).

Given an initial state a, the total expected α-discounted cost under a transmission selection policy $\pi \in \Pi$ is given by

$$V^\alpha(a; \pi) = \limsup_{n \to \infty} \mathbb{E}\left[\sum_{i=0}^{n-1} \alpha^i C(a_i, u_i)\right], \tag{17}$$

where $0 < \alpha < 1$ is the discount factor. Then, the optimization problem of minimizing the total expected α-discounted cost can be cast as

Problem 2 (Discounted cost MDP)

$$V^\alpha(a) \triangleq \min_{\pi \in \Pi} V^\alpha(a; \pi), \tag{18}$$

where $V^\alpha(a)$ denotes the optimal total expected α-discounted cost. A transmission selection policy is said to be *α-discounted cost optimal* if it solves the problem in (18). In Proposition 4.2, we show that there exists a stationary deterministic transmission selection policy which is α-discounted cost optimal and provide a way to explore the property of the optimal policy.

Proposition 4.2. *(a) The optimal total expected α-discounted cost V^α satisfies the following optimality equation:*

$$V^\alpha(a) = \min_{u \in \mathcal{U}} Q^\alpha(a, u), \tag{19}$$

where

$$Q^\alpha(a, u) = C(a, u) + \alpha p_u V^\alpha(a + d_u) + \alpha(1 - p_u)V^\alpha(d_u). \tag{20}$$

(b) The stationary deterministic policy determined by the right-hand-side of (19) *is α-discounted cost optimal.*

(c) Let $V_n^\alpha(a)$ be the cost-to-go function such that $V_0^\alpha(a) = \frac{d_1 - d_2}{\alpha(p_2 - p_1)} a$ and for $n \ge 0$

$$V_{n+1}^\alpha(a) = \min_{u \in \mathcal{U}} Q_{n+1}^\alpha(a, u), \tag{21}$$

where

$$Q_{n+1}^\alpha(a, u) = C(a, u) + \alpha p_u V_n^\alpha(a + d_u) + \alpha(1 - p_u)V_n^\alpha(d_u). \tag{22}$$

Then, we have that for each α, $V_n^\alpha(a) \to V^\alpha(a)$ as $n \to \infty$.

Proof. See our technical report [18]. □

Next, with the optimality equation (19) and value iteration (21), we show that the optimal policy is of threshold-type in Lemma 4.3.

LEMMA 4.3. *Given a discount factor α,*
(i) if $(1-\alpha p_2)d_1 \le (1-\alpha p_1)d_2$, then the α-discounted cost optimal policy is of threshold-type, i.e., the optimal action is a non-increasing function of the age.
(ii) if $(1-\alpha p_2)d_1 \ge (1-\alpha p_1)d_2$, then the α-discounted cost optimal policy is of threshold-type, i.e., the optimal action is a non-decreasing function of the age.

PROOF. Please see Appendix A □

This lemma proves that the α-discounted cost optimal policy is of threshold-type. Next, we extend the results to average cost MDP and show that there exists a stationary deterministic average-optimal policy which is of threshold-type. Based on the results in [15], we have the following lemma, which provides a candidate for average-optimal policy.

LEMMA 4.4. *(i) Let α_n be any sequence of discount factors converging to 1 with α_n-discounted cost optimal stationary deterministic policy π^{α_n}. There exists a subsequence γ_n and a stationary policy π^* that is a limit point of π^{γ_n}.*
(ii) If $d_1(1-p_2) \le d_2(1-p_1)$, π^* is of threshold-type in (14)*; if $d_1(1-p_2) \ge d_2(1-p_1)$, π^* is of threshold-type in* (15).

PROOF. See our technical report [18]. □

By [15], under certain conditions (A proof of these conditions verification is provided in our technical report [18]), π^* is average-optimal.

4.2 Algorithm Design

Recall that if $p(\beta) = 0$, then the optimal transmission selection policies that solve (8) and (9) are identical. Given β, the optimal policy that solves (9) is of threshold-type by Theorem 4.1 and then (9) can be re-expressed as

$$p(\beta) = \begin{cases} \min\limits_{\pi \in \Pi_1} J(\pi, \beta), & \text{if } d_1(1-p_2) < d_2(1-p_1) \quad (23) \\ \min\limits_{\pi \in \Pi_2} J(\pi, \beta), & \text{if } d_1(1-p_2) \ge d_2(1-p_1) \quad (24) \end{cases}$$

where Π_1 and Π_2 denote the sets of threshold-type policies in (14) and (15), respectively. Thus, the optimal policy that solves (8) can be obtained with two steps:

- **Step (i):** For each β, find the β-associated average-optimal policy π^*_β such that $p(\beta) = J(\pi^*_\beta, \beta)$.
- **Step (ii):** Find β^* such that $p(\beta^*) = 0$. This implies $\pi^*_{\beta^*}$ solves (8).

To narrow our searching range in (ii), in Lemma 4.5, we provide a lower bound $\beta_{\min}$ and an upper bound $\beta_{\max}$ of β^*. Then, for (i), we only need to pay attention to $p(\beta)$ for $\beta \in [\beta_{\min}, \beta_{\max}]$.

In particular, within the range of β, we show that J in (23) is *quasi-convex* in a threshold related variable, which enables us to devise a *low-complexity* algorithm based on golden section search. Moreover, we show that π^*_β that solves (24) always chooses $u = 2$, which allows us to get the optimal policy for (8) directly.

LEMMA 4.5. *The parameter β^* is lower bounded by $\beta_{min} \triangleq 1.5d_2$ and upper bounded by $\beta_{max} \triangleq \min\left\{(\frac{1}{1-p_1}+0.5)d_1, (\frac{1}{1-p_2}+0.5)d_2\right\}$.*

PROOF. See our technical report [18]. □

In the following content, we provide a theoretical analysis step by step for our algorithm design in Algorithm 1, which returns the optimal threshold and optimal average age for (8).

Algorithm 1: Threshold-based Age-Optimal Policy

```
given d_1, d_2, p_1, p_2, τ = (√5 − 1)/2, tolerance ε_1, ε_2, l = β_min, r = β_max;
while r − l > ε_1 do
    β = (r+l)/2;
    if d_1(1 − p_2) ≥ d_2(1 − p_1) then
        m = 0, k = 0, J* = J_2(0, 0, β);
    else
        k_max = ⌊d_1/d_2⌋, J* = f(1, 0, β);
        foreach k_1 ∈ {0, 1, · · · , k_max} do
            if ∂J_1(y, k_1, β)/∂y |_{y=1} < 0 then
                m = 0;
            else
                y_0 = 0, y_1 = 1, y_2 = y_1 − (y_1 − y_0)τ,
                  y_3 = y_0 + (y_1 − y_0)τ;
                while p_2 y_1 ≥ y_0 and y_1 − y_0 > ε_2 do
                    if J_1(y_2, k_1, β) > J_1(y_3, k_1, β) then
                        y_0 = y_2;
                    else
                        y_1 = y_3;
                    end
                    y_2 = y_1 − (y_1 − y_0)τ, y_3 = y_0 + (y_1 − y_0)τ;
                end
                t_1 = ⌊log_{p_2} y_0⌋, t_2 = ⌈log_{p_2} y_1⌉;
                m = arg min_{m∈{t_1,t_2}} J_1(p_2^m; k, β);
            end
            if J_1(p_2^m, k_1, β) ≤ J* then
                J* = J_1(p_2^m, k_1, β);
            end
        end
    end
    if J* ≥ 0 then
        l = β;
    else
        r = β;
    end
end
```

Step (i): Find the optimal policy π^*_β: Note that both of the threshold-type policies defined in (14) and (15) result in a Markov chain with a single positive recurrent class. Thus, given threshold a_1^* in (14) or a_2^* in (15), we can obtain the expression of average cost under the corresponding threshold-type policy with aid of state transition diagram. With this, we obtain some nice properties of average cost function in Theorem 4.6, which enables us to get a low-complexity algorithm. Before providing the result, we define the integer threshold which will be used in the theorem and algorithm.

Recall that age $a \in \mathcal{S}$ is expressed as the sum of multiple d_1's and d_2's. Note that under the threshold-type policy in (14), if $a \le a_1^*$, $\zeta(a) = 2$. This implies that if $a \le a_1^*$, a is in the form $a = d_j + ld_2, j \in \mathcal{U}, l \in \mathbb{N}$. Thus, it is sufficient to use the following integer threshold to represent the threshold-type policy in (14).

$$m_1 \triangleq \min\left\{l : d_1 + ld_2 > a_1^*, l \in \mathbb{N}\right\}, \tag{25}$$

$$n_1 \triangleq \min\left\{l : d_2 + ld_2 > a_1^*, l \in \mathbb{N}\right\}. \tag{26}$$

Table 1: Notations

A_1	$= p_2^{k_1}\big(d_2(1-p_1)-d_1(1-p_2)\big)\big(d_2(k_1+1)-d_1\big)$
B_1	$= -A_2 + p_2^{k_1}(1-p_2)(0.5d_1^2 - \beta d_1 + \frac{d_1^2}{1-p_1}) + (1-p_1)(\beta d_2 - 0.5d_2^2 - \frac{d_2^2}{1-p_2})$
C_1	$= (1-p_1)(0.5d_2^2 - \beta d_2 + \frac{1}{1-p_2})d_2^2$
D_1	$= \big(d_1(1-p_2)-d_2(1-p_1)\big)d_2 p_2^{k_1}$
A_2	$= p_1^{k_2}\big(d_1(1-p_2)-d_2(1-p_1)\big)\big(d_1(1-k_2)-d_2\big)$
B_2	$= -p_1^{k_2}(1-p_2)(0.5d_1^2 - \beta d_1 + \frac{d_1^2}{1-p_1}) + (1-p_1)(0.5d_2^2 - \beta d_2 + \frac{d_2^2}{1-p_2})$ $-\big(d_1(1-p_2)-d_2(1-p_1)\big)(d_1-d_2)$
C_2	$= (1-p_2)(0.5d_1^2 - \beta d_1 + \frac{d_1^2}{1-p_1})$
D_2	$= \big(d_2(1-p_1)-d_1(1-p_2)\big)d_1$

With this, the threshold policy in (14) is rewritten as

$$\zeta^*(a) = \begin{cases} 2 & \text{if } a < d_1 + m_1 d_2 \text{ and } a < d_2 + n_1 d_2, \\ 1 & \text{if } a \geq d_1 + m_1 d_2 \text{ or } a \geq d_2 + n_1 d_2. \end{cases} \tag{27}$$

Similarly, the policy in (15) can be rewritten as

$$\zeta^*(a) = \begin{cases} 1 & \text{if } a < d_1 + m_2 d_1 \text{ and } a < d_2 + n_2 d_1, \\ 2 & \text{if } a \geq d_1 + m_2 d_1 \text{ or } a \geq d_2 + n_2 d_1, \end{cases} \tag{28}$$

where

$$m_2 \triangleq \min\left\{l : d_1 + l d_1 > a_2^*, l \in \mathbb{N}\right\}, \tag{29}$$

$$n_2 \triangleq \min\left\{l : d_2 + l d_1 > a_2^*, l \in \mathbb{N}\right\}. \tag{30}$$

Based on the analysis, (23) and (24) can be re-expressed as

$$p(\beta) = \begin{cases} \min_{m_1,k_1} J_1(m_1,k_1,\beta), & \text{if } d_1(1-p_2)<d_2(1-p_1) \quad (31) \\ \min_{m_2,k_2} J_2(m_2,k_2,\beta), & \text{if } d_1(1-p_2)\geq d_2(1-p_1) \quad (32) \end{cases}$$

where $k_1 = n_1 - m_1$ and $k_2 = n_2 - m_2$. Also, we use $J_1(m_1,k_1,\beta)$ and $J_2(m_2,k_2,\beta)$ to denote the average cost under the policies in (27) and (28), respectively. We have $k_2 \in \mathcal{K}_2 \triangleq \{0,1\}$ and $k_1 \in \mathcal{K}_1 \triangleq \left\{0,\cdots,\lfloor\frac{d_1}{d_2}\rfloor\right\}$. In particular, according to the definition of m_2 and n_2, we have $(m_2+1)d_1 > a_2^* \geq d_2 + (n_2-1)d_1$ and $d_2 + n_2 d_1 > a_2^* \geq m_2 d_1$. Substitute $k_2 = n_2 - m_2$ into these two inequalities, we get $k_2 \in \mathcal{K}_2$. Similarly, we have $k_1 \in \mathcal{K}_1$.

In Theorem 4.6, we provide some nice properties for J_1 and J_2, which enables us to develop a low-complexity algorithm. To this end, we make a change of variable in Theorem 4.6 (i), i.e., m_1 is replaced with $\log_{p_2}(y)$ in J_1, where $y \in (0,1]$. Some notations used in this theorem are defined in Table 1.

Theorem 4.6. *Given $\beta \in [\beta_{min}, \beta_{max}]$,*
(i) if $d_1(1-p_2) < d_2(1-p_1)$, then the average cost is given by

$$J_1(y,k_1,\beta) = \frac{A_1 y^2 + B_1 y + C_1 + D_1 y \log_{p_2}(y)}{1 - p_1 + (-1 + p_1 + p_2^{k_1}(1-p_2))y}, \tag{33}$$

where $y \triangleq p_2^{m_1}$. Moreover, $J_1(y,k_1,\beta)$ is quasi-convex in y for $0 < y \leq 1$, given $k_1 \in \mathcal{K}_1$.

(ii) if $d_1(1-p_2) \geq d_2(1-p_1)$, then optimal average cost is given by

$$J_2(0,0,\beta) = \frac{A_2 + B_2 + C_2}{1 - p_1}. \tag{34}$$

Moreover, the average-optimal policy chooses $u = 2$ at every transmission opportunity.

Proof. Proof of part (i) is provided in Appendix B. For part (ii), the key idea is to show that for all m_2, $J_2(m_2,k_2,\beta)$ is non-decreasing in $k_2 \in \mathcal{K}_2$, and then $J_2(m_2,0,\beta)$ is non-decreasing in m_2. This implies that the optimal decision is $u = 2$ at each transmission opportunity. Due to the space limitation, detailed proof is provided in our technical report [18]. □

With the property in Theorem 4.6 (i), we are able to use golden section search [14] to find the optimal value of y under condition $d_1(1-p_2) < d_2(1-p_1)$. The details are provided in Algorithm 1 (Line 12-20). Note that $\log_{p_2}(\cdot)$ is one-to-one functions. Thus, after obtaining the optimal y that minimizes J_1, the corresponding optimal threshold m_1 can be easily obtained by comparing $J_1(p_2^{\lfloor\log_{p_2}(y)\rfloor}, k_1, \beta)$ and $J_1(p_2^{\lceil\log_{p_2}(y)\rceil}, k_1, \beta)$). The details are provided in Algorithm 1 (Line 21-22). Till now, we have solved $\min_{m_1} J_1(m_1,k_1,\beta)$ given k_1 and β. Note that $k_1 \in \mathcal{K}_1$ has finite and countable values. Then, we can easily solve (31) under condition $d_1(1-p_2) < d_2(1-p_1)$ by searching for the optimal k_1 in the finite set $\mathcal{K}_1$.

Note that Theorem 4.6 (ii) applies to all $\beta \in [\beta_{\min}, \beta_{\max}]$ including β^*. Thus, the optimal policy for (8) is to take $u = 2$ at every transmission opportunity, which is returned directly in Algorithm 1 (Line 4-5). Under this condition, the step (ii) is only used to find optimal average cost β^* for (8).

Moreover, in Line 9-10, we add a judgement sentence, i.e., if the condition $\frac{\partial J_1(y,k_1,\beta)}{\partial y}|_{y=1} < 0$ is satisfied, then we can directly obtain the optimal m_1 without running golden section method. This further reduces the algorithm complexity. The judgement is based on the fact that $\lim_{y\to 0} \frac{\partial J_1(y,k_1,\beta)}{\partial y} < 0$ (this is proved in the proof of Theorem 4.6 in Appendix B) and J_1 is quasi-convex. Thus, if $\frac{\partial J_1(y,k_1,\beta)}{\partial y}|_{y=1} < 0$, then J_1 is non-increasing in y.

Step (ii): Find β^*: By Lemma 3.1, if $p(\beta) > 0$, then $\beta < \beta^*$; if $p(\beta) < 0$, then $\beta > \beta^*$. Thus, we can use bisection method to search for β^*. The details are provided in Algorithm 1 (Line 2-3 and Line 29-33).

5 A DISCUSSION ON THE GENERAL MULTI-RATE TRANSMISSION SELECTION PROBLEM

For the multi-rate transmission selection (from more than two-rate) problem, we assume that there are $N \in \mathbb{N}^+$ transmission modes for selection such that transmission delays and error probabilities satisfy $d_j > d_{j+1}$ and $p_j < p_{j+1}$, for $j \in \{1,2,\cdots,N-1\}$, respectively. Since the transmission delays and transmission error probabilities affect the age in opposite direction, it is difficult to determine the optimal policy. Thanks to the results obtained for the two-rate transmission selection problem, we obtain some useful insights for the general multi-rate transmission selection.

Table 2: Optimal threshold versus delay

	d_1=1.5d_2	d_1=1.7d_2	d_1=1.9d_2	d_1=2.1d_2	d_1=2.3d_2
d_2=1	(0,0)	(0,0)	(1,2)	(3,4)	(15,16)
d_2=5	(0,1)	(0,1)	(1,2)	(3,4)	(15,16)
d_2=9	(0,1)	(0,1)	(1,2)	(3,4)	(15,16)

In particular, for the two-rate transmission selection problem, the optimal action is a non-increasing function of age under condition $\bar{d}_1 \leq \bar{d}_2$ and a non-decreasing function of age under condition $\bar{d}_1 \geq \bar{d}_2$, where $\bar{d}_1$ and $\bar{d}_2$ are mean delays of low and high rate transmissions, respectively, as defined in (16). With this, we can infer that for $N > 2$, the optimal policy will have following properties:

- (i) if $\bar{d}_1 \leq \bar{d}_2 \leq \cdots \leq \bar{d}_N$, then the optimal action will be a non-increasing function of the age,
- (ii) if $\bar{d}_1 \geq \bar{d}_2 \geq \cdots \geq \bar{d}_N$, then the optimal action will be a non-decreasing function of the age.

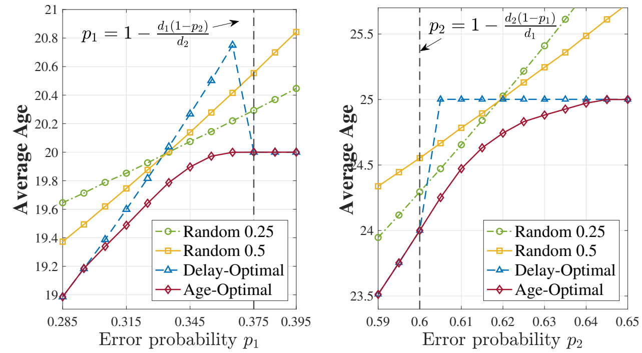
For the cases that are not covered in (i) and (ii), we need more investigation on the property of the optimal policy. In addition, if we can show that the optimal policy for the general multi-rate problem is of threshold-type, then machine learning algorithms can be used to determine the optimal threshold. For example, if we regard each threshold-type policy with a certain threshold as a bandit, then basic bandit algorithms like UCB can be exploited to find the optimal threshold [13].

6 NUMERICAL RESULTS

In this section, we present some numerical results to explore the performance of the threshold-based age-optimal policy and verify our theoretical results.

First, we consider an update system, in which $p_1 = 0.4$ and $p_2 = 0.75$. Table. 2 illustrates the relation between the optimal threshold versus the transmission delay d_2 and the delay ratio $\frac{d_1}{d_2}$ under condition $d_2(1-p_1) > d_1(1-p_2)$. The threshold (m_1, n_1) in the table is obtained by Algorithm 1. We observe that the threshold increases with either d_2 or the delay ratio $\frac{d_1}{d_2}$. Note that when the age is below the threshold, transmission mode 2 is selected. Thus, this observation implies that transmission mode 2 becomes more preferable either when d_2 increases with fixed $\frac{d_1}{d_2}$ or when $\frac{d_1}{d_2}$ increases with fixed d_2.

In Fig. 3, we consider an update system, in which the transmission delays are $d_1 = 10$ and $d_2 = 8$. We use "Delay-Optimal" to denote the optimal policy that minimizes the average updating delay by always choosing the transmission mode with minimum mean delay [8, 12]. Moreover, we use "Age-Optimal" to denote the optimal policy that is obtained from Algorithm 1. We use "Random p" to denote the policy that chooses $u = 1$ with probability p. We compare our threshold-based "Age-Optimal" policy with "Random p" policies and the "Delay-Optimal" policy, where $p \in \{0.25, 0.5\}$. Fig. 3a (3b) illustrates the total average age in (7) versus transmission error probability p_1 (p_2) given $p_2 = 0.5$ ($p_1 = 0.5$). The dashed line in the figure marks the point at which $d_1(1-p_2) = d_2(1-p_1)$. The left and right (right and left) of the line corresponds to conditions $d_1(1-p_2) < d_2(1-p_1)$ and $d_1(1-p_2) > d_2(1-p_1)$ in Fig. 3a (in Fig. 3b), respectively. As we can observe, the age-optimal policy



(a) Average age vs p_1 given $p_2 = 0.5$ **(b) Average age vs p_2 given $p_1 = 0.5$**

Figure 3: Total average age versus transmission error probability

outperforms other plotted policies. This agrees with Theorem 4.1. Moreover, the results confirm that the delay-optimal policy does not necessarily minimize the age. In particular, the gap between delay-optimal and age-optimal policy becomes larger as p_1 (p_2) approaches to the left side (right side) of the dashed line in Fig. 3a (Fig. 3b). The jump in the curve of the delay-optimal policy is incurred by the switch between two transmission modes. For example, in Fig. 3a, delay-optimal policy chooses $u = 1$ on the left side of the dashed line, while chooses $u = 2$ on the right side. This is because transmission mode 1 has smaller mean delay on the left side while transmission mode 2 has smaller mean delay on the right side.

7 CONCLUSION

In this paper, we studied the transmission selection problem for minimizing age of information in information update system with heterogenous transmissions. We assume that there are two different transmissions with varying delay and error probability. We showed that there exists a stationary deterministic optimal transmission selection policy which is of threshold-type in age (Theorem 4.1). This result reveals an interesting phenomenon: If the mean delay of the low rate transmission is smaller than that of high rate transmission, then the optimal action chooses the one with higher mean delay when age is smaller than a certain threshold. This is in contrary with the delay-optimal policy that always chooses the transmission with lower mean delay. In addition, we showed that if the mean delay of the low rate transmission is smaller than that of high rate transmission, the average cost is quasi-convex in a threshold related variable; otherwise, the optimal policy chooses $u = 2$ for each transmission opportunity (Theorem 4.6). This enabled us to design a low-complexity algorithm to obtain the optimal policy (Algorithm 1). For the future work, we plan to study the multi-rate scenario with more than two selections of heterogenous transmissions based on the insights discussed in Section 5.

ACKNOWLEDGMENTS

This work was funded in part through NSF grants: CNS-1955561, CNS-1901057, CNS- 2007231, CNS-1618520, and CNS-1409336, and an Office of Naval Research under Grant N00014-17-1-241.

REFERENCES

[1] Baran Tan Bacinoglu, Elif Tugce Ceran, and Elif Uysal-Biyikoglu. 2015. Age of information under energy replenishment constraints. In *2015 Information Theory and Applications Workshop (ITA)*. IEEE, 25–31.

[2] Ahmed M Bedewy, Yin Sun, Sastry Kompella, and Ness B Shroff. 2019. Age-optimal sampling and transmission scheduling in multi-source systems. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 121–130.

[3] Ahmed M Bedewy, Yin Sun, Rahul Singh, and Ness B Shroff. 2020. Optimizing information freshness using low-power status updates via sleep-wake scheduling. In *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*. 51–60.

[4] Ahmed M Bedewy, Yin Sun, Rahul Singh, and Ness B Shroff. 2021. Low-Power Status Updates via Sleep-Wake Scheduling. *arXiv preprint arXiv:2102.02846* (2021).

[5] Dimitri P Bertsekas. 2000. *Dynamic programming and optimal control: Vol. 1*. Athena scientific Belmont.

[6] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[7] Elif Tuğçe Ceran, Deniz Gündüz, and András György. 2019. Average age of information with hybrid ARQ under a resource constraint. *IEEE Transactions on Wireless Communications* 18, 3 (2019), 1900–1913.

[8] Francis De Véricourt and Yong-Pin Zhou. 2005. Managing response time in a call-routing problem with service failure. *Operations Research* 53, 6 (2005), 968–981.

[9] Xu Du, Yin Sun, Ness B Shroff, and Ashutosh Sabharwal. 2020. Balancing Queueing and Retransmission: Latency-Optimal Massive MIMO Design. *IEEE Transactions on Wireless Communications* 19, 4 (2020), 2293–2307.

[10] Igor Kadota, Abhishek Sinha, Elif Uysal-Biyikoglu, Rahul Singh, and Eytan Modiano. 2018. Scheduling policies for minimizing age of information in broadcast wireless networks. *IEEE/ACM Transactions on Networking* 26, 6 (2018), 2637–2650.

[11] Sanjit Kaul, Marco Gruteser, Vinuth Rai, and John Kenney. 2011. Minimizing age of information in vehicular networks. In *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. IEEE, 350–358.

[12] Erhun Ozkan and Jeffrey P Kharoufeh. 2014. Optimal control of a two-server queueing system with failures. *Probability in the Engineering and Informational Sciences* 28, 4 (2014), 489–527.

[13] KJ Prabuchandran, Tejas Bodas, and Theja Tulabandhula. 2016. Reinforcement learning algorithms for regret minimization in structured markov decision processes. *arXiv preprint arXiv:1608.04929* (2016).

[14] WH Press, SA Teukolsky, WT Vetterling, and BP Flannery. 2007. Section 10.2. Golden section search in one dimension. In *Numerical recipes: the art of scientific computing*. Cambridge University Press New York, 492–496.

[15] Linn I Sennott. 1989. Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs. *Operations Research* 37, 4 (1989), 626–633.

[16] Yin Sun, Elif Uysal-Biyikoglu, Roy D Yates, C Emre Koksal, and Ness B Shroff. 2017. Update or wait: How to keep your data fresh. *IEEE Transactions on Information Theory* 63, 11 (2017), 7492–7508.

[17] Xianwen Wu, Jing Yang, and Jingxian Wu. 2017. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Transactions on Green Communications and Networking* 2, 1 (2017), 193–204.

[18] Guidan Yao, Ahmed M. Bedewy, and Ness B. Shroff. 2020. Battle between Rate and Error in Minimizing Age of Information. arXiv:cs.IT/2012.09351

[19] Roy D Yates. 2015. Lazy is timely: Status updates by an energy harvesting source. In *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 3008–3012.

[20] Bo Zhou and Walid Saad. 2018. Optimal sampling and updating for minimizing age of information in the internet of things. In *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 1–6.

A PROOF OF LEMMA 4.3

(i) To show that the optimal action is a non-increasing function of age, we will show that if the optimal action is $u = 1$ at certain age, then for the age larger than this age, the optimal action is still $u = 1$. Let a_1 and a_2 be ages such that $a_1 \le a_2$. In particular, we will show that if $Q^\alpha(a_1, 1) \le Q^\alpha(a_1, 2)$, then $Q^\alpha(a_2, 1) \le Q^\alpha(a_2, 2)$. It suffices to show that $Q^\alpha(a, 1) - Q^\alpha(a, 2)$ decreases with age a, i.e.,

$$Q^\alpha(a_1, 1) - Q^\alpha(a_1, 2) \ge Q^\alpha(a_2, 1) - Q^\alpha(a_2, 2). \tag{35}$$

By Proposition 4.2, we only need to show that for $n \in \mathbb{N}$,

$$Q^\alpha_{n+1}(a_1, 1) - Q^\alpha_{n+1}(a_1, 2) \ge Q^\alpha_{n+1}(a_2, 1) - Q^\alpha_{n+1}(a_2, 2) \tag{36}$$

$$\begin{aligned}
\Leftrightarrow & C(a_1, 1) + \alpha p_1 V^\alpha_n(a_1 + d_1) + \alpha(1 - p_1)V^\alpha_n(d_1) \\
& - C(a_1, 2) - \alpha p_2 V^\alpha_n(a_1 + d_2) - \alpha(1 - p_2)V^\alpha_n(d_2) \\
& \ge C(a_2, 1) + \alpha p_1 V^\alpha_n(a_2 + d_1) + \alpha(1 - p_1)V^\alpha_n(d_1) \\
& - C(a_2, 2) - \alpha p_2 V^\alpha_n(a_2 + d_2) - \alpha(1 - p_2)V^\alpha_n(d_2)
\end{aligned} \tag{37}$$

$$\begin{aligned}
\Leftrightarrow & (a_1 - a_2)(d_1 - d_2) + \alpha p_1 V^\alpha_n(a_1 + d_1) - \alpha p_1 V^\alpha_n(a_2 + d_1) \\
& + \alpha p_2 V^\alpha_n(a_2 + d_2) - \alpha p_2 V^\alpha_n(a_1 + d_2) \ge 0.
\end{aligned} \tag{38}$$

We show (38) by induction. When $n = 0$, substitute $V^\alpha_0(a) = \frac{d_1 - d_2}{\alpha(p_2 - p_1)} a$ into the left-hand-side of (38) with n replaced with 0 and we have

$$\begin{aligned}
& (a_1 - a_2)(d_1 - d_2) + \alpha p_1 V^\alpha_0(a_1 + d_1) - \alpha p_1 V^\alpha_0(a_2 + d_1) \\
& + \alpha p_2 V^\alpha_0(a_2 + d_2) - \alpha p_2 V^\alpha_0(a_1 + d_2) = 0
\end{aligned} \tag{39}$$

Hence, (38) holds when $n = 0$. Suppose (38) holds for n, we will show that it holds for $n + 1$. Let u_1, u_2, u_3, u_4 be the optimal actions in state $a_1 + d_1, a_2 + d_1, a_2 + d_2, a_1 + d_2$, respectively. Specifically, $V^\alpha_{n+1}(a_1 + d_1) = Q^\alpha_{n+1}(a_1 + d_1, u_1)$, $V^\alpha_{n+1}(a_2 + d_1) = Q^\alpha_{n+1}(a_2 + d_1, u_2)$, $V^\alpha_{n+1}(a_2 + d_2) = Q^\alpha_{n+1}(a_2 + d_2, u_3)$ and $V^\alpha_{n+1}(a_1 + d_2) = Q^\alpha_{n+1}(a_1 + d_2, u_4)$. Then, the left-hand-side of (38) is

$$\begin{aligned}
& (a_1 - a_2)(d_1 - d_2) \\
& + \alpha p_1 Q^\alpha_{n+1}(a_1 + d_1, u_1) - \alpha p_1 Q^\alpha_{n+1}(a_2 + d_1, u_2) \\
& + \alpha p_2 Q^\alpha_{n+1}(a_2 + d_2, u_3) - \alpha p_2 Q^\alpha_{n+1}(a_1 + d_2, u_4) \\
= & (a_1 - a_2)(d_1 - d_2) \\
& + \underbrace{\alpha p_1 Q^\alpha_{n+1}(a_1 + d_1, u_1) - \alpha p_1 Q^\alpha_{n+1}(a_2 + d_1, u_1)}_{A} \\
& + \underbrace{\alpha p_1 Q^\alpha_{n+1}(a_2 + d_1, u_1) - \alpha p_1 Q^\alpha_{n+1}(a_2 + d_1, u_2)}_{\ge 0 \text{ (By optimality of action } u_2)} \\
& - \underbrace{\left(\alpha p_2 Q^\alpha_{n+1}(a_1 + d_2, u_3) - \alpha p_2 Q^\alpha_{n+1}(a_2 + d_2, u_3)\right)}_{B} \\
& + \underbrace{\alpha p_2 Q^\alpha_{n+1}(a_1 + d_2, u_3) - \alpha p_2 Q^\alpha_{n+1}(a_1 + d_2, u_4)}_{\ge 0 \text{ (By optimality of action } u_4)}.
\end{aligned} \tag{40}$$

By induction hypothesis, we have

$$\begin{aligned}
A \ge & \alpha p_1 \min_{u_1 \in \{1,2\}} \{(a_1 - a_2)d_{u_1} + \alpha p_{u_1} V^\alpha_n(a_1 + d_1 + d_{u_1}) \\
& - \alpha p_{u_1} V^\alpha_n(a_2 + d_1 + d_{u_1})\}
\end{aligned} \tag{41}$$

$$\begin{aligned}
= & \alpha p_1 \Big((a_1 - a_2)d_2 + \alpha p_2 V^\alpha_n(a_1 + d_1 + d_2) \\
& - \alpha p_2 V^\alpha_n(a_2 + d_1 + d_2)\Big).
\end{aligned} \tag{42}$$

Similarly,

$$\begin{aligned}
B \le & \alpha p_2 \max_{u_3 \in \{1,2\}} \{(a_1 - a_2)d_{u_3} + \alpha p_{u_3} V^\alpha_n(a_1 + d_2 + d_{u_3}) \\
& - \alpha p_{u_3} V^\alpha_n(a_2 + d_2 + d_{u_3})\}
\end{aligned} \tag{43}$$

$$\begin{aligned}
= & \alpha p_2 \Big((a_1 - a_2)d_1 + \alpha p_1 V^\alpha_n(a_1 + d_2 + d_1) \\
& - \alpha p_1 V^\alpha_n(a_2 + d_2 + d_1)\Big).
\end{aligned} \tag{44}$$

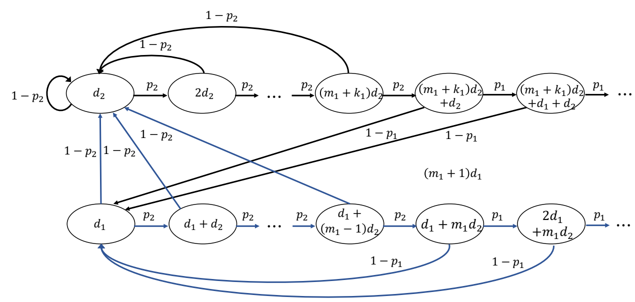


Figure 4: State transition diagram under policy in (14)

Hence, substitute (42) and (44) into (40) and we obtain

$$\begin{aligned}
&(a_1-a_2)(d_1-d_2)\\
&+\alpha p_1 Q^\alpha_{n+1}(a_1+d_1,u_1)-\alpha p_1 Q^\alpha_{n+1}(a_2+d_1,u_2)\\
&+\alpha p_2 Q^\alpha_{n+1}(a_2+d_2,u_3)-\alpha p_2 Q^\alpha_{n+1}(a_1+d_2,u_4)\\
\geq&(a_1-a_2)(d_1-d_2)+A-B && (45)\\
\geq&(a_1-a_2)\left((1-\alpha p_2)d_1-(1-\alpha p_1)d_2\right) && (46)\\
\geq&0, && (47)
\end{aligned}$$

where (47) holds by the condition $(1-\alpha p_2)d_1 \leq (1-\alpha p_1)d_2$.

(ii) Similar to part (i), it suffices to show that for $a_1 \leq a_2$,

$$Q^\alpha(a_1,1)-Q^\alpha(a_1,2) \leq Q^\alpha(a_2,1)-Q^\alpha(a_2,2). \tag{48}$$

The proof is similar to part (i).

B PROOF OF THEOREM 4.6 (I)

We first obtain expression of average cost in (33) with aid of state transition diagram. The state transition diagram under the policy in (14) is given in Fig. 4. Define the state steady probabilities x_l, x'_l, z_l and z'_l under policy in (14) as

$$x_l \triangleq \mathbb{P}(a=d_2+ld_2), \qquad 0\leq l\leq m_1+k_1, \tag{49}$$
$$x'_l \triangleq \mathbb{P}(a=d_2+(m_1+k_1)d_2+ld_1), \qquad l\geq 0, \tag{50}$$
$$z_l \triangleq \mathbb{P}(a=d_1+ld_2), \qquad 0\leq l\leq m_1, \tag{51}$$
$$z'_l \triangleq \mathbb{P}(a=d_1+m_1d_2+ld_1), \qquad l\geq 0. \tag{52}$$

Based on the state transition diagram, balance equations can be obtained as follows:

$$x_0=(1-p_2)\left(\sum_{l=0}^{m_1-1} z_l+\sum_{l=0}^{m_1+k_1-1} x_l\right), \tag{53}$$
$$p_2 x_l = x_{l+1}, \quad l\in\{0,1\cdots m_1+k_1-1\}, \tag{54}$$
$$p_2 z_l = z_{l+1}, \quad l\in\{0,1\cdots m_1-1\}, \tag{55}$$
$$p_1 x'_l = x'_{l+1}, \quad l\in\{0,1\cdots\}, \tag{56}$$
$$p_1 z_l = z'_{l+1}, \quad l\in\{0,1\cdots\}. \tag{57}$$

Solving the equations (53)-(57), we obtain the expressions of x_l, x'_l, z_l and z'_l in terms of x_0 as follows:

$$x_l = p_2^l x_0, \quad l\in\{0,1\cdots m_1+k_1\}, \tag{58}$$
$$z_l = \frac{p_2^{m_1+k_1}p_2^l}{1-p_2^{m_1}}x_0, \quad l\in\{0,1\cdots m_1\}, \tag{59}$$
$$x'_l = p_2^{m_1+k_1}p_1^l x_0, \quad l\in\{0,1\cdots\}, \tag{60}$$
$$z'_l = \frac{p_2^{2m_1+k_1}p_1^l}{1-p_2^{m_1}}x_0, \quad l\in\{0,1\cdots\}. \tag{61}$$

Substituting (58)-(61) into $\sum_{l=0}^{m_1+k_1} x_l+\sum_{l=0}^{m_1} z_l+\sum_{l=1}^{\infty} x'_l+\sum_{l=1}^{\infty} z'_l=1$, we obtain x_0 as

$$x_0=\frac{(1-p_1)(1-p_2)(1-p_2^{m_1})}{1-p_1+p_2^{m_1}(p_1-1)+p_2^{m_1+k_1}(1-p_2)}. \tag{62}$$

The average cost $J_1(m_1,k_1,\beta)$ is expressed as

$$\begin{aligned}
J_1(m_1,k_1,\beta)=&\sum_{l=0}^{m_1+k_1-1} C(d_2+ld_2,2)x_l+\sum_{l=0}^{m_1-1} C(d_1+ld_2,2)z_l\\
&+\sum_{l=0}^{\infty} C(d_2+(m_1+k_1)d_2+ld_1,1)x'_l\\
&+\sum_{l=0}^{\infty} C(d_1+m_1d_2+ld_1,1)z'_l.
\end{aligned} \tag{63}$$

where $C(\cdot,\cdot)$ is the cost function defined in (11). Substitute (58)-(61) and (62) into (63). After some algebraic manipulation and change of variable ($p_2^{m_1}$ is replaced by y), we obtain (33).

Next, we show that J_1 is quasi-convex. By definition of quasi-convex, it suffices to show that its first derivative $\frac{\partial J_1(y,k_1,\beta)}{\partial y}$ with respect to y satisfies at least one of the following conditions [6]:

- $\frac{\partial J_1(y,k_1,\beta)}{\partial y} \geq 0$
- $\frac{\partial J_1(y,k_1,\beta)}{\partial y} \leq 0$
- there exists a point $y_0\in(0,1]$ such that for $0<y\leq y_0$, $\frac{\partial J_1(y,k_1,\beta)}{\partial y}\leq 0$, and for $1\geq y\geq y_0$, $\frac{\partial J_1(y,k_1,\beta)}{\partial y}\geq 0$

After some algebraic manipulation, $\frac{\partial J_1(y,k_1,\beta)}{\partial y}$ is expressed as

$$\frac{\partial J_1(y,k_1,\beta)}{\partial y}=\frac{h(y,k_1,\beta)}{(1-p_1+ry)^2} \tag{64}$$

where $r=-1+p_1+p_2^{k_1}(1-p_2)$ and $h(y,k_1,\beta)$ is given by

$$\begin{aligned}
h(y,k_1,\beta)\triangleq&\left(A_1y^2+\frac{D_2y}{\ln p_2}-C_1\right)r\\
&+\left(B_1+2A_1y+\frac{D_1(1+\ln y)}{\ln p_2}\right)(1-p_2)
\end{aligned} \tag{65}$$

Note that the denominator of the (64) is positive. Thus, to show that at least one of the conditions above holds, it suffices to show that $h(y,k_1,\beta)$ satisfies at least one of the following conditions:

- B1: $h(y,k_1,\beta)\geq 0$;
- B2: $h(y,k_1,\beta)\leq 0$;
- B3: $\exists y_0\in(0,1]$ such that for $0<y\leq y_0$, $h(y,k_1,\beta)\leq 0$, and for $1\geq y\geq y_0$, $h(y,k_1,\beta)\geq 0$.

In fact, the first derivative of $h(y,k_1,\beta)$ with respect to y is

$$\begin{aligned}
&\frac{\partial h(y,k_1,\beta)}{\partial y}\\
=&\underbrace{\left(2(d_2(k_1+1)-d_1)y-\frac{d_2}{\ln p_2}\right)}_{G(y)}\underbrace{(ry+1-p_1)}_{H(y)}\frac{p_2^{k_1}W}{y}
\end{aligned} \tag{66}$$

where $W = d_2(1-p_1) - d_1(1-p_2)$. By condition $d_2(1-p_1) > d_1(1-p_2)$, $W > 0$. Note that $r \leq -1 + p_1 + (1 - p_2) = p_1 - p_2 < 0$ since $0 < p_1 < p_2 < 1$. Thus, $H(y) \geq r + 1 - p_1 = p_2^{k_1}(1 - p_2) > 0$, for $0 < y \leq 1$. Hence, $\frac{\partial h(y,k_1,\beta)}{\partial y}$ is positive (negative) if and only if $G(y)$ is positive (negative). Next, we will show our finial result by analyzing two different cases.

Case 1: If $d_2(k_1 + 1) - d_1 \geq 0$, then $G(y) \geq -\frac{d_2}{\ln p_2} > 0$, for $0 < y \leq 1$. In this case, $\frac{\partial h(y,k_1,\beta)}{\partial y} > 0$, for $0 < y \leq 1$. Thus, $h(y, k_1, \beta)$ is increasing in y, for $0 < y \leq 1$. Note that since $D_1 < 0$ (by condition $d_2(1 - p_1) > d_1(1 - p_2)$) and $p_2 < 1$, $\lim_{y\to 0} h(y, k_1, \beta) = -C_1 r + \left(B_1 + \frac{D_1}{\ln p_2} + \lim_{y\to 0} \frac{D_1 \ln y}{\ln p_2}\right)(1 - p_2) < 0$. Thus, if $h(1, k_1, \beta) \leq 0$, then B2 holds; if $h(1, k_1, \beta) > 0$, then B3 holds.

Case 2: If $d_2(k_1 + 1) - d_1 < 0$, then $G(y)$ is decreasing in y and $y_1 = \frac{d_2}{2(d_2(k_1+1)-d_1)\ln p_2} > 0$ is a turning point such that $G(y) > 0$ when $y < y_1$ and $G(y) < 0$ when $y > y_1$.

If $y_1 \geq 1$, then $G(y) \geq 0$ for $0 < y \leq 1$. Thus, $\frac{\partial h(y,k_1,\beta)}{\partial y} \geq 0$, which implies that $h(y, k_1, \beta)$ is increasing in y for $0 < y \leq 1$. Hence, B2 or B3 holds as explained in case 1.

If $y_1 < 1$, then $G(y) > 0$ when $y < y_1$ and $G(y) < 0$ when $1 > y > y_1$, which implies $h(y, k_1, \beta)$ first increases and then decreases for $0 < y \leq 1$. We claim that if $y_1 < 1$, then $h(1, k_1, \beta) \geq 0$ for $\beta \in [\beta_{\min}, \beta_{\max}]$ and $k_1 \in \mathcal{K}_1$. Recall that $\lim_{y\to 0} h(y, k_1, \beta) < 0$. With this, the claim implies that $h(y, k_1, \beta)$ starts with some negative value and increases to zero at some point y' and after $h(y, k_1, \beta)$ becomes positive, it will keep positive for $y' < y \leq 1$ (B3 holds). It remains to show that our claim holds.

Since $y_1 = \frac{d_2}{2(d_2(k_1+1)-d_1)\ln p_2} < 1$, we have

$$\frac{d_1}{d_2} > k_1 + 1 - \frac{1}{2\ln p_2} \tag{67}$$

After some algebraic manipulation and simplification, $h(1, k_1, \beta)$ is expressed as

$$\begin{aligned}
&h(1, k_1, \beta) \\
=&p_2^{2k_1} W \left(d_2(k_1 + 1) - d_1 - \frac{d_2}{\ln p_2}\right)(1 - p_2) \\
&+ p_2^{k_1}(1 - p_2)(1 - p_1)\left(\left(\frac{1}{1 - p_1} + 0.5\right)d_1^2 - \left(\frac{1}{1 - p_2} + 0.5\right)d_2^2\right) \\
&+ p_2^{k_1}(d_2 - d_1)(1 - p_1)(1 - p_2)\beta
\end{aligned} \tag{68}$$

Since $p_2^{k_1}(d_2 - d_1)(1 - p_1)(1 - p_2) < 0$, $h(1, k_1, \beta)$ is decreasing in β and thus $h(1, k_1, \beta) \geq h(1, k_1, \beta_{\max}) \geq h(1, k_1, (\frac{1}{1-p_1} + 0.5)d_1)$. It remains to show that $h(1, k_1, (\frac{1}{1-p_1} + 0.5)d_1) \geq 0$. It is equivalent to show that $\frac{h(1,k_1,(\frac{1}{1-p_1}+0.5)d_1)}{d_2^2} \geq 0$ since $d_2^2 \geq 0$. Let $z = \frac{d_1}{d_2}$, substitute this into $\frac{h(1,k_1,(\frac{1}{1-p_1}+0.5)d_1)}{d_2^2}$ and obtain a function of z denoted by $w(z)$. By (67), the first derivative of $w(z)$ satisfies

$$\begin{aligned}
w'(z) =& 2p_2^{2k_1}(1 - p_2)^2 z + p_2^{k_1}(1 - p_2)\left(\frac{1}{1 - p_1} + 0.5\right)(1 - p_1) \\
&+ p_2^{2k_1}(1 - p_2)\left(\frac{1 - p_2}{\ln p_2} - (1 - p_2)(k_1 + 1) - (1-p_1)\right)
\end{aligned} \tag{69}$$

$$\geq w'\left(k_1 + 1 - \frac{1}{2\ln p_2}\right) \tag{70}$$

$$\begin{aligned}
=&p_2^{k_1}(1 - p_2)\left(p_2^{k_1}(1 - p_2)(k_1 + 1) + \frac{1 - p_1}{2}\right) \\
&+ p_2^{k_1}(1-p_2)\left(1 - p_2^{k_1}(1-p_1)\right) > 0
\end{aligned} \tag{71}$$

The second inequality in (71) holds since $0 < p_1 < p_2 < 1$. Thus, $w(z)$ is increasing and we have

$$w(z) \geq w\left(k_1 + 1 - \frac{1}{2\ln p_2}\right) \tag{72}$$

$$\begin{aligned}
=&\underbrace{p_2^{k_1}(1 - p_2)}_{>0}\underbrace{\left(p_2^{k_1}(1 - p_2) + \ln p_2\right)}_{\leq 1-p_2+\ln p_2 \leq 0}\underbrace{\left(\frac{k_1}{2\ln p_2} - \frac{1}{4(\ln p_2)^2}\right)}_{<0} \\
&+ \frac{p_2^{k_1}}{2\ln p_2}\gamma(p_1, p_2, k_1)
\end{aligned} \tag{73}$$

where

$$\begin{aligned}
\gamma(p_1, p_2, k_1) =& (p_1 - p_2)(1 - p_2)p_2^{k_1} + 2(p_1 - p_2)\ln p_2 \\
&+ 2k_1(1 - 0.5p_2)(1 - p_2)\ln p_2 + (1 - p_2)(0.5p_1 - 1)
\end{aligned}$$

To show $w(z) \geq 0$, we only need to show $\gamma(p_1, p_2, k_1) \leq 0$. Actually,

$$\begin{aligned}
&\frac{\partial \gamma(p_1, p_2, k_1)}{\partial p_2} \\
=&\underbrace{(p_1 - p_2)(1 - p_2)p_2^{k_1}\ln p_2 + k_1(2 - p_1)\left(\frac{1}{p_2} - 1 - \ln p_2\right)}_{>0 \text{ by } 0<p_1<p_2<1} \\
&+ \underbrace{p_2^{k_1}(2p_2 - 1 - p_1) + \frac{2p_1}{p_2} - 1 - \frac{p_1}{2} - 2\ln p_2}_{\theta(p_1,p_2,k_1)} \geq 0
\end{aligned} \tag{74}$$

where the inequality holds since $\theta(p_1,p_2,k_1)$ decreases with p_2 and $\theta(p_1,1,k_1) = 0.5p_1 > 0$. Actually,

$$\frac{\partial\theta(p_1, p_2, k_1)}{\partial p_2} = (2p_2 - 1 - p_1)p_2^{k_1}\ln p_2 + 2p_2^{k_1} - 2\frac{p_1}{p_2^2} - \frac{2}{p_2} \tag{75}$$

$$< (p_2 - 1)p_2^{k_1}\ln p_2 + 2p_2^{k_1} - 2\frac{p_1}{p_2^2} - \frac{2}{p_2} \tag{76}$$

$$\leq -(p_2 - 1)p_2^{k_1}\frac{1}{2p_2} + 2p_2^{k_1} - 2\frac{p_1}{p_2^2} - \frac{2}{p_2} \tag{77}$$

$$\leq -(p_2 - 1)\frac{1}{2p_2} + 2 - 2\frac{p_1}{p_2^2} - \frac{2}{p_2} \tag{78}$$

$$< 0 \tag{79}$$

where (76) holds since $p_2 - p_1 > 0$ and $\ln p_2 < 0$; (77) holds by $s(x) \triangleq \frac{1}{x} + 2\ln x \geq 0$ for $0 < x \leq 1$; (78) holds since $p_2^{k_1} \leq 1$; (78) holds since $p_2 < 1$. In particular, the first derivative of $s(x)$ is $s'(x) = -\frac{1}{x^2} + \frac{2}{x}$. Thus, $s(x)$ decreases when $x \leq 0.5$ (since $s'(x) \leq 0$ when $x \leq 0.5$) and then increases when $x \geq 0.5$ (since $s'(x) \geq 0$ when $x \geq 0.5$). Thus, $s(x) \geq s(0.5) = 0.2213 > 0$. By (74), γ increases with p_2 and $\gamma(p_1, p_2, k_1) \leq \gamma(p_1, 1, k_1) = 0$. This completes our proof.