Age-Optimal Low-Power Status Update over Time-Correlated Fading Channel

Guidan Yao, *Student Member, IEEE,* Ahmed M. Bedewy and Ness B. Shroff, *Fellow, IEEE*

Abstract—In this paper, we consider transmission scheduling in a status update system, where updates are generated periodically and transmitted over a Gilbert-Elliott fading channel. The goal is to minimize the long-run average age of information (AoI) under a long-run average energy constraint. We consider two practical cases to obtain channel state information (CSI): (i) without channel sensing and (ii) with delayed channel sensing. For (i), CSI is revealed by the feedback (ACK/NACK) of a transmission, but when no transmission occurs, CSI is not revealed. Thus, we have to balance tradeoffs across energy, AoI, channel exploration, and channel exploitation. The problem is formulated as a constrained partially observable Markov decision process (POMDP). We show that the optimal policy is a randomized mixture of no more than two stationary deterministic policies each of which is of a threshold-type in the belief on the channel. For (ii), (delayed) CSI is available via channel sensing. Then, the tradeoff is only between the AoI and energy. The problem is formulated as a constrained MDP. The optimal policy is shown to have a similar structure as in (i) but with an AoI associated threshold. With these, we develop an optimal structure-aware algorithm for each case.

Index Terms—Age of information, partially observable MDP, MDP, threshold-type policy.

1 Introduction

For status update systems, where time-sensitive status updates of certain underlying physical process are sent to a remote destination, it is important that the destination receives fresh updates. The *age of information* (AoI) is a performance metric that is a good measure of the freshness of the data at the destination. In particular, AoI is defined as the time elapsed since the generation of the most recently received status update. Different from the long-established packet delay metric, AoI jointly captures packet delay and inter-delivery time.

The problem of minimizing the AoI in status update systems has attracted significant recent attention (e.g., [1], [2], [3], [4], [5], [6], [7], [8], [9]). Due to the fact that sensors in the status update system are usually battery-powered and thus have limited energy supply, the problem of minimizing the long-run average AoI has to take energy constraints into account. Moreover, communication over a wireless channel is subject to multiple impairments such as fading, path loss and interference, which may lead to status updating failure. Note that each failed retransmission consumes energy which is wasted. Thus, it is critical that we design intelligent transmission scheduling algorithms (e.g., to determine whether we should suspend transmission or retransmit) in order to increase channel utilization and prolong battery lifetime.

There have been a number of works that investigate AoI minimization problem under energy constraints with different settings of energy constraints, channel assumptions and service times [10], [11], [12], [13], [14], [15], [16], [17], [18]. Except that papers [13], [14], [16] consider average energy constraints, other papers consider energy harvesting transmitter. The paper [10] studies the online policies with stochastic service time. In [11], [12], [15], the authors assume that channel is noiseless and the service time is negligible, and propose offline or online status updating policies. Despite the noiseless channel assumption, the knowledge of the channel state is often assumed to be perfect so that successful transmission is guaranteed. In [13], the authors jointly design sampling and updating processes over a channel with perfect channel state information. The success of each transmission is guaranteed via using predefined transmission power which is a function of the channel state. However, in many practical scenarios, the channel state may not be known a priori. Thus, more recent works have also considered unreliable transmissions with imperfect knowledge of wireless channels. For example, in [16], the authors consider a block fading channel, where the channel is assumed to vary independently and identically over time slots. In [14], the authors consider an error-prone channel, where decoding error depends only on the number of retransmissions. In [17] and [18], authors consider a noisy channel with time-invariant success probability of delivery: unit-sized battery in [17] and infinite battery in [18].

However, these works neglect an important characteristics of the wireless fading channel: The *channel memory* or *time correlation* [19] when studying unreliable transmissions with imperfect knowledge of channel states. Indeed, the memory can be intelligently exploited to predict the channel state and thus to design efficient scheduling policies in the presence of transmission cost. A finite state Markov chain is an often used and appropriate model for fading channel [20]. A somewhat simplified but often-used abstraction is

- *This paper was presented in part at ISIT 2021*
- *Guidan Yao is with the Department of ECE, The Ohio State University, Columbus, OH 43210 USA. E-mail: yao.539@osu.edu*
- *A. M. Bedewy is with the Department of ECE, The Ohio State University, Columbus, OH 43210 USA. E-mail: bedewy.2@osu.edu*
- *N. B. Shroff is with the Department of ECE and the Department of CSE, The Ohio State University, Columbus, OH 43210 USA. E-mail: shroff.11@osu.edu*

a two-state Markovian model known as the Gilbert-Elliot channel [21]. The model assumes that the channel can be either in a good or bad state, and captures the essence of the fading process. In [22], the authors consider status updating in cognitive radio networks. The occupation of primary user's channel is modeled as a two-state Markov chain. Although a Markov chain is used to model occupation of primary channel, their threshold-type structural result is built on perfect knowledge of the channel state since update decisions are made based on perfect sensing results. In contrast, in our work, we do not assume that the channel state is known a priori at the time of making updating decisions.

Motivated by the time-correlation in a fading channel and the fact that sensors in practice are typically configured to generate status updates periodically [23], in this paper, we consider a status update system where the status update is generated periodically and transmitted over a Gilbert-Elliot channel. We do not assume that the channel state is known a priori and consider two practical cases to obtain the channel state information (CSI): (i) *(without channel sensing)* CSI is revealed by the ACK/NACK feedback of a transmission; (ii) *(with delayed channel sensing)* delayed CSI is always available via delayed channel sensing regardless of transmission decisions. With this context, we study the problem of how to minimize the long-run average AoI at the destination under a long-run average energy constraint at the source, which is motivated by the fact that sensors or IoT devices at the source usually have limited energy supply. The problem in case (i) is formulated as a constrained partially observable Markov decision process problem (POMDP) while in case (ii), it is formulated as a constrained Markov decision problem (MDP). It is known that in general POMDP is PSPACE hard to solve and MDP suffers from the curse of dimensionality. Thus, in this paper, we focus on providing theoretical guarantees such that we obtain an optimal policy with reduced complexity. To this end, we characterize the structure of the optimal policy in either case. Note that the problem in both cases involves long-run average cost with infinite state space and unbounded costs, which makes the analysis difficult. In particular, our key contributions include:

- For the case without channel sensing, we show that the optimal transmission scheduling policy is a randomized mixture of no more than two *stationary deterministic threshold-type* policies (Theorem 1 and Corollary 2). Note that although there are some works that deal with proving the optimality of threshold-type policies in POMDPs [24], [25], [26], [27], [28], the techniques in these papers cannot be applied to our problem. This is because, given hidden state and action, the one-stage cost in these papers is constant and bounded, while the one-stage cost in our paper depends on varying and unbounded AoI.
- We propose a finite-state approximation for our infinite-state (unbounded AoI and belief on channel state) belief MDP and show that the optimal policy for the approximate belief MDP converges to the original one (Theorem 2). Based on this, we propose an optimal efficient structure-aware transmission scheduling algorithm (Algorithm 1) for the approximate belief MDP.
- For the case with delayed channel sensing, we show that the optimal transmission scheduling policy is also a randomized mixture of no more than two stationary deterministic threshold-type policies. However, due to the simplification in the state, the threshold here is on AoI (Theorem 3). Moreover, we provide a relation between the thresholds associated with different channel states (Theorem 3). Based on the theoretical insights, we develop an efficient structure-aware algorithm (Algorithm 2).

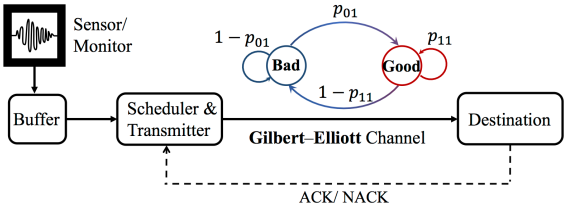


Fig. 1: System Model

The remainder of this paper is organized as follows. The system model is introduced in Section 2. For the case without channel sensing, we formulate the problem in Section 3, and in Section 4, we explore the structure of the optimal policy and propose a structure-aware algorithm. In Section 5, we investigate the case with delayed channel sensing. Section 7 contains numerical results.

2 System Model

We consider a status update system where status updates are generated periodically and transmitted to a remote destination over a time-correlated fading channel, as shown in Fig. 1. We consider a time-slotted system, where a time slot corresponds to the time duration of the packet transmission time and feedback period. Every K consecutive time slots form *a frame*. Updates are generated at the beginning of each frame. In any frame, if the generated status update is not delivered by the end of the frame, then it gets replaced by a new one in the next frame. Define $\mathcal{K}$ as the set of *relative slot index* within a frame, $\mathcal{K} \triangleq \{1, 2, \cdots, K\}$. Use $t \in \{1, 2, \cdots\}$ as an *absolute index* for the time slot count, which increments indefinitely with time. For any time slot t, the corresponding frame index $l_t \in \{1, 2, \cdots\}$ is determined by $l_t = \lceil \frac{t}{K} \rceil$ and relative slot index $k_t \in \mathcal{K}$ is determined by $k_t = ((t-1) \bmod K) + 1$, where $\lceil \cdot \rceil$ is the ceiling function.

2.1 Channel Model

The time-correlated fading channel for transmission is assumed to evolve as a two-state Gilbert-Elliot model [21]. Let h_t denote the channel state at time slot t. Then, $h_t = 1$ ($h_t = 0$) denotes that channel is in a "good" ("bad") state. In the "bad" state, the channel is assumed to be in a deep fade such that transmission fails with probability one; while in the "good" state, a transmission attempt is always successful. This assumption conforms with the signal-to-noise ratio (SNR) threshold model for reception where successful

decoding of a packet at the destination occurs if and only if the SNR exceeds a certain threshold value. The channel transition probabilities are given by $\mathbb{P}(h_{t+1}=1|h_t=1)=p_{11}$ and $\mathbb{P}(h_{t+1}=1|h_t=0)=p_{01}$. We assume that the channel transitions occur at the end of each time slot, and that p_{11} and p_{01} are known.

The presence of channel memory (time correlation) makes it possible to predict the channel state. In this paper, we assume that $p_{11} \geq p_{01}$ (positively correlated channel) (similar assumptions have been used in [26], [28]).

2.2 Transmission Scheduler and Channel State Information

At the beginning of each slot t, the scheduler takes a decision $u_t \in \mathcal{U} \triangleq \{0,1\}$, where $u_t = 1$ means transmitting or retransmitting the undelivered status update, and $u_t = 0$ denotes suspension of the transmission or retransmission. In each frame, if the generated update is delivered at the k_t-th slot of the frame, then we have $u_t = 0$ for the remaining slots in the frame. For simplicity, we use transmission to refer to both transmission and retransmission in the remaining content.

In this paper, we consider two practical cases to obtain CSI: (i) *(without channel sensing)* CSI is revealed via the feedback on transmission from the destination; (ii) *(with delayed channel sensing)* CSI of the last time slot is always available via delayed channel sensing regardless of transmission decisions. In particular, for case (i), if a transmission is attempted, then the scheduler receives an error-free ACK/NACK feedback from the destination specifying whether the status update was delivered or not before the end of the slot. We use Θ to denote the set of observations, $\Theta \triangleq \{0,1\}$. Let $\theta_t \in \Theta$ be the observation at time slot t. Then, $\theta_t = 1$ denotes a successful transmission. $\theta_t = 0$ occurs when the transmission occurs over the channel in the bad state or the transmission is suspended. Note that when a decision is made not to transmit updates, the scheduler will not obtain feedback revealing the CSI. Thus, the channel in this case is partially observable. In contrast, for case (ii), CSI of the last time slot is always available via delayed channel sensing regardless of transmission decisions.

2.3 Age of Information

Age of information (AoI) reflects the timeliness of the information at the destination. It is defined as the time elapsed since the generation of the most recently received update at the destination. Let Δ_t denote the AoI at the beginning of the time slot t. Let $U(t)$ denote the generation time of the last successfully received status update for time slot t. Then, Δ_t is given by $\Delta_t \triangleq t - U(t)$.

If a status update is not successfully delivered in a time slot, then the AoI increases by one, otherwise, the AoI drops to the time elapsed since the beginning of the frame (generation time of the newly delivered status update). Then, the value of Δ_{t+1} is updated as follows:

$$\Delta_{t+1} = \begin{cases} k_t & \text{if } u_t = 1, \theta_t = 1, \\ \Delta_t + 1 & \text{otherwise.} \end{cases} \tag{1}$$

Let $\mathcal{A}_k$ denote the set of all possible AoI values at the k-th slot of a frame. By (1), $\mathcal{A}_k = \{\Delta : \Delta = mK + (k)_-, m \in \{0,1,2,\cdots\}\}$, where $(k)_- \triangleq ((K + k - 2) \bmod K) + 1$ denotes the relative slot index before k. An example of the AoI evolution with $K = 4$ is illustrated in Fig. 2.

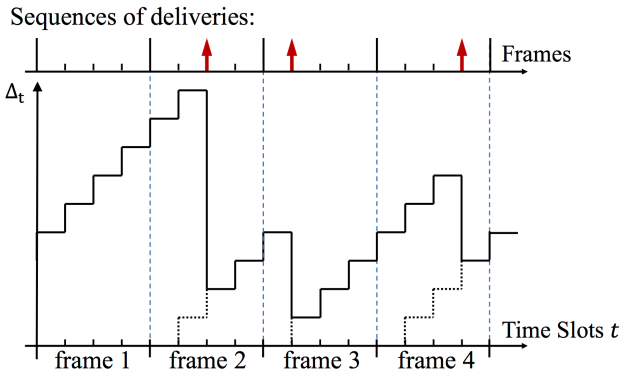


Fig. 2: On the top, a sample sequence of deliveries during four frames. Each frame consists of 4 time slots. The upward arrows represent the times of deliveries. On the bottom, the associated evolution of AoI.

2.4 Optimization Problem

A transmission scheduling policy $\pi = \{d_1, d_2, \cdots\}$ specifies the decision rules for each time slot, where a decision rule d_t is a function that maps the past actions, past and current AoI, relative slot index of a frame and channel states to actions. The transmitter consumes energy for each packet transmission. In addition, energy is consumed for channel sensing in the case with delayed channel sensing. However, for channel sensing, a few pilot symbols will be enough. For example, IEEE 802.11a uses only 4 pilot symbols for channel sensing [29], [30]. That is, the energy cost of channel sensing is often much smaller than that of packet transmission and is a constant value. Thus, it will not change the outcome of the optimization problem. Thus, we only consider the transmission energy in optimization problems. We assume that each transmission consumes the same energy which is normalized as one unit energy. Note that if there is no energy constraint at the source, then exploiting every time slot in transmitting the undelivered update is optimal. This is because suspending the transmission of an undelivered status update does not contribute to decreasing the AoI and also wastes an opportunity to learn the channel state in the case without channel sensing. However, repeated transmission attempts could result in excessive energy consumption, and could be impractical for sensors or IoT devices that are usually energy constrained. Accordingly, we employ a long-run average energy consumption constraint at the source. In particular, our objective in this paper is to design a transmission scheduling policy π that minimizes the following long-run average AoI

$$\bar{A}(\pi) \triangleq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\Big[\sum_{t=1}^{T} \Delta_t | \Delta_1, k_1, h_1\Big], \tag{2}$$

while the long-run average energy consumption $\bar{E}(\pi)$ does not exceed $E_{\max} \in (0,1]$, i.e.

$$\bar{E}(\pi) \triangleq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\Big[\sum_{t=1}^{T} u_t | \Delta_1, k_1, h_1\Big] \leq E_{\max}, \tag{3}$$

where $\mathbb{E}_\pi$ denotes expectation under policy π. Observe that $E_{\max} = 1$ means that we have enough energy to support a transmission in every time slot.

In case (i), although a failed transmission does not decrease AoI, it provides CSI at the cost of energy. Thus, the transmission scheduler has to balance tradeoffs across energy, AoI, channel exploration, and channel exploitation. In case (ii), delayed CSI is always available regardless of transmission decisions. Thus, the tradeoff is only between the AoI and energy.

3 Constrained POMDP Formulation and Lagrangian relaxation without Channel Sensing

3.1 Constrained POMDP Formulation

At the beginning of each time slot, the scheduler chooses an action u. Given that the state of the underlying Markov channel is i, the user observes $\theta(i,u) \in \{0,1\}$, which indicates the state of the current channel. Specifically, an ACK will be received if and only if the status update is transmitted over a "good" channel, i.e. $\theta(1,1) = 1$. Otherwise, for $(i,u) \neq (1,1)$, $\theta(i,u) = 0$. Upon receipt of the feedback/observation, the AoI changes accordingly at the end of this slot. The sequence of operations in each slot is illustrated in Fig. 3. Note that when transmission is suspended, the channel state is not directly observable. Together with the average energy constraint, the problem we consider in the paper turns out to be a constrained partially observable Markov decision problem (POMDP).

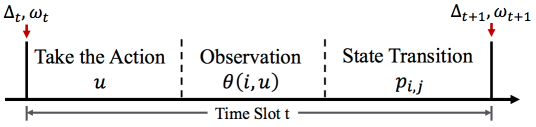


Fig. 3: Sequence of operations in a slot

It has been shown in [31] that for any slot t, a belief state ω_t is a *sufficient statistic* to describe the knowledge of underlying channel state and thus can be used for making optimal decisions at time slot t.

Definition 1. *The belief state ω_t is the conditional probability (given observation and action history) that the channel is in a good state at the beginning of the time slot t.*

Thus, adding the belief to the system state, the constrained POMDP can be written as constrained belief MDP [32]. We describe the components of the framework as follows:

States: The system state consists of completely observable states and the belief state, i.e., the system state at slot t is defined by a 3-tuple $\mathbf{s}_t=(\Delta_t, k_t, \omega_t)$, where $\Delta_t \in \mathcal{A}_{k_t}$ is the AoI state that evolves as (1); $k_t \in \mathcal{K}$ is the relative slot index in the frame l_t that evolves as $k_{t+1}=(k_t)_+$, where $(y)_+ \triangleq (y \mod K)+1$; ω_t is the belief state whose evolution is defined in the following paragraph.

Belief Update: Given u_t and θ_t, the belief state in time slot $t+1$ is updated by $\omega_{t+1} = \Lambda(\omega_t, u_t, \theta_t)$, where $\Lambda(\omega_t, u_t, \theta_t)$ is given by

$$\omega_{t+1} = \Lambda(\omega_t, u_t, \theta_t) = \begin{cases} p_{11} & \text{if } u_t = 1, \theta_t = 1, \\ p_{01} & \text{if } u_t = 1, \theta_t = 0, \\ \mathcal{T}(\omega_t) & \text{if } u_t = 0, \end{cases} \quad (4)$$

where $\mathcal{T}(\omega_t) = \omega_t p_{11} + (1-\omega_t)p_{01}$ denotes the one-step belief update. Observe that, if $u_t = 0$, then the scheduler will not learn the channel state and the belief is updated only according to the Markov chain. If $u_t = 1$, the observation θ_t after the transmission provides the true channel state before the state transition, which occurs at the end of the time slot (see Fig. 3).

Let $\mathcal{T}^m(\omega_t) \triangleq \mathbb{P}(h_{t+m} = 1|\omega_t)$ denote m-step belief update when the channel is unobserved for m consecutive slots, where $m \in \{0, 1, \cdots\}$ and $\mathcal{T}^0(\omega) = \omega$. Note that by (4), after a transmission ($u_t = 1$), ω_{t+1} is either p_{01} or p_{11}. The belief state ω is, hereafter, updated by $\mathcal{T}$ upon each suspension until next transmission attempt. Thus, the belief state ω is in the form of $\mathcal{T}^m(p_{01})$ or $\mathcal{T}^m(p_{11})$, where $m \geq 0$. Moreover, an increase in AoI by one results from either a failed transmission or suspension. Thus, given AoI state Δ_t, the maximum suspension time after last transmission is no longer than $\Delta_t - 1$. By this, given AoI state Δ, the belief state belongs to the following set $\Omega_\Delta \triangleq \{\omega : \omega = \mathcal{T}^m(p_{01})$ or $\mathcal{T}^m(p_{11}), 0 \leq m < \Delta\}$. As a result, the state space is given by $\mathcal{S} \triangleq \{(\Delta, k, \omega) : k \in \mathcal{K}, \Delta \in \mathcal{A}_k, \omega \in \Omega_\Delta\}$.

Actions: Action set is $\mathcal{U} = \{0, 1\}$ defined in Section 2.2.

Transition probabilities: Given the current state $\mathbf{s}_t = (\Delta_t, k_t, \omega_t)$ and action u_t at time slot t, the transition probability to the state $\mathbf{s}_{t+1} = (\Delta_{t+1}, k_{t+1}, \omega_{t+1})$ at the next time slot $t+1$, which is denoted by $P_{\mathbf{s}_t \mathbf{s}_{t+1}}(u_t)$, is defined as

$$\begin{aligned} P_{\mathbf{s}_t \mathbf{s}_{t+1}}(u_t) &\triangleq \mathbb{P}(\mathbf{s}_{t+1}|\mathbf{s}_t, u_t) \\ &= \sum_{\theta_t \in \Theta} \mathbb{P}(\theta_t|\mathbf{s}_t, u_t)\mathbb{P}(\mathbf{s}_{t+1}|\mathbf{s}_t, u_t, \theta_t), \end{aligned} \quad (5)$$

where

$$\mathbb{P}(\theta_t|\mathbf{s}_t, u_t) = \begin{cases} \omega_t & \text{if } u_t = 1, \theta_t = 1, \\ 1-\omega_t & \text{if } u_t = 1, \theta_t = 0, \\ 1 & \text{if } u_t = 0, \theta_t = 0, \\ 0 & \text{otherwise}, \end{cases} \quad (6)$$

$$\begin{aligned} &\mathbb{P}(\mathbf{s}_{t+1}|\mathbf{s}_t, u_t, \theta_t) \\ &= \begin{cases} 1 \text{ if } \mathbf{s}_{t+1} = (k_t, (k_t)_+, \Lambda(\omega_t, u_t, \theta_t)), u_t = 1, \theta_t = 1, \\ 1 \text{ if } \mathbf{s}_{t+1} = (\Delta_t + 1, (k_t)_+, \Lambda(\omega_t, u_t, \theta_t)), \theta_t = 0, \\ 0 \text{ otherwise.} \end{cases} \end{aligned} \quad (7)$$

Costs: Given a state $\mathbf{s}_t = (\Delta_t, k_t, \omega_t)$ and an action choice u_t at time slot t, the cost of one slot is the AoI at the beginning of this slot, i.e., we have

$$C_\Delta(\mathbf{s}_t, u_t) = \Delta_t. \quad (8)$$

Moreover, the energy consumption of one slot is

$$C_E(\mathbf{s}_t, u_t) = u_t. \quad (9)$$

For any policy π, we assume that the resulted Markov chain is a unichain (same assumptions are also made in [13], [33]). The transmission scheduling problem can be formulated as a constrained belief MDP:

Problem 1 (Constrained average-AoI belief MDP):

$$\bar{A}^\star \triangleq \min_{\pi} \quad \bar{A}(\pi) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\big[\sum_{t=1}^{T} C_\Delta(\mathbf{s}_t, u_t)\big] \tag{10}$$

$$\text{s.t.} \quad \bar{E}(\pi) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\big[\sum_{t=1}^{T} C_E(\mathbf{s}_t, u_t)\big] \le E_{\max}.$$

We use $\bar{A}^\star$ to denote the optimal average AoI, which is the solution to the problem (10).

A policy is *stationary* if the decision rule is independent of time, i.e., $d_t = d$, for all t. Moreover, a policy is *randomized* if $d_t : \mathcal{S} \to \mathcal{P}(\mathcal{U})$ specifies a probability distribution on the set of actions. The policy is *deterministic* if $d_t : \mathcal{S} \to \mathcal{U}$ chooses an action with certainty. We show in Section 4 that there exists a stationary policy which is a randomized mixture of no more than two deterministic policies that achieves $\bar{A}^\star$.

3.2 Lagrange Formulation of the Constrained POMDP

To obtain the optimal transmission scheduling policy, we reformulate the constrained average-AoI belief MDP in (10) as a parameterized unconstrained average cost belief MDP. Given Lagrange multiplier λ, the instantaneous Lagrangian cost at time slot t is defined by

$$C(\mathbf{s}_t, u_t; \lambda) = C_\Delta(\mathbf{s}_t, u_t) + \lambda C_E(\mathbf{s}_t, u_t). \tag{11}$$

Then, the average Lagrangian cost under policy π is given by

$$\bar{L}(\pi; \lambda) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\big[\sum_{t=1}^{T} C(\mathbf{s}_t, u_t; \lambda)\big]. \tag{12}$$

Then, we have an unconstrained average cost belief MDP, which aims at minimizing the above average Lagrangian cost:

Problem 2 (Unconstrained average cost belief MDP):

$$\bar{L}^\star(\lambda) \triangleq \min_{\pi} \ \bar{L}(\pi; \lambda), \tag{13}$$

where $\bar{L}^\star(\lambda)$ is the optimal average Lagrangian cost with regard to λ. A policy is said to be *average cost optimal* if it minimizes the average Lagrangian cost.

The relation between the optimal solutions of the problems (10) and (13) is provided in the following corollary.

Corollary 1. *The optimal average AoI of problem* (10) *and the optimal average Lagrangian cost of problem* (13) *satisfy*

$$\bar{A}^\star = \sup_{\lambda\ge 0} \bar{L}^\star(\lambda) - \lambda E_{max}. \tag{14}$$

Proof. By Theorem 12.7 in [34], we only need to check the following condition: for all $r \in \mathbb{R}$, the set $G(r) \triangleq \{\mathbf{s} \in \mathcal{S} : \inf_u C_\Delta(\mathbf{s}, u) < r\}$ is finite. Given r, for any $\mathbf{s}' = (\Delta', k', \omega') \in G(r)$, $\Delta' = \inf_u C_\Delta(\mathbf{s}', u) < r$. With fixed finite Δ', $\Omega_{\Delta'}$ is finite. Thus, $G(r)$ is finite. □

4 Structure Based Algorithm Design

In this section, we investigate the structure of the optimal policy for the constrained average-AoI belief MDP in (10), develop a finite approximation for the infinite belief MDP and propose an optimal algorithm using the structure.

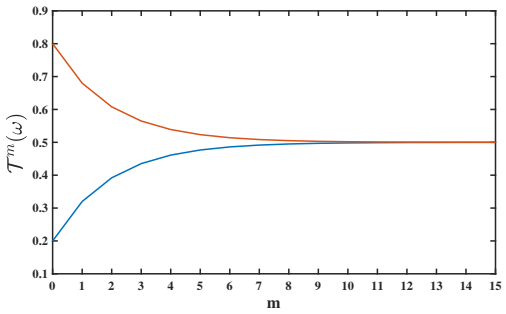


Fig. 4: The m-step belief update with $p_{01} = 0.2$ and $p_{11} = 0.8$

4.1 Structure of Constrained Average-AoI Optimal Policy

4.1.1 Main results

To explore the structure, we first show that there exists a stationary deterministic *threshold-type* scheduling policy that solves the unconstrained average cost belief MDP in (13).

Theorem 1. *Given λ, there exists a stationary deterministic unconstrained average cost optimal policy that is of threshold-type in belief. Specifically,* (13) *can be minimized by a policy of the form $\pi^\star_\lambda = (d^\star_\lambda, d^\star_\lambda, \cdots)$, where*

$$d^\star_\lambda(\Delta, k, \omega) = \begin{cases} 0 & \text{if } 0 \le \omega < \omega^\star(\Delta, k; \lambda), \\ 1 & \text{if } \omega^\star(\Delta, k; \lambda) \le \omega, \end{cases} \tag{15}$$

where $\omega^\star(\Delta, k; \lambda)$ denotes the threshold given pair of AoI and relative slot index (Δ, k) and Lagrange multiplier λ.

Proof. Please see Section 4.1.2. □

Remark:

- Note that the techniques in papers dealing with the optimality of the threshold-type policies in POMDP [24], [25], [26], [27], [28] cannot be applied to our problem. This is because, given hidden state and action, the one-stage cost in these papers is *constant and bounded*, while the one-stage cost in our paper depends on *varying and unbounded* AoI. To deal with this, we jointly prove some properties of the value functions and the optimality of threshold-type policies (see Lemma 1).
- In Fig. 4, we provide an example of m-step belief update. In particular, the $\mathcal{T}^m(p_{01})$ increases with m while $\mathcal{T}^m(p_{11})$ decreases with m, and the two curves will converge to a same value. Denote the value as ω_0. With this, we can conclude the smallest belief threshold must be smaller than ω_0. Otherwise, the transmitter stops updating after a failed transmission (belief changes to p_{01} and never satisfies threshold condition afterwards), which leads to the infinite average AoI.
- If $K = 1$, then the optimal policy will behave as follows:
  Suppose the initial state is $(1, p_{11})$, we start with (a).
  (a) Upon a successful transmission, the state changes to $(1, p_{11})$. Then, the transmitter suspends transmission until the waiting time $t_s \ge 0$ satisfies $\mathcal{T}^{t_s}(p_{11}) \ge$

$\omega^*(1+t_s)$. If the transmission is successful, go to (a); otherwise, go to (b).
(b) Upon a failed transmission, the state changes to (Δ, p_{01}), where $\Delta \geq 2$. Then, the the transmitter suspends transmission until the waiting time $t_f \geq 0$ satisfies $\mathcal{T}^{t_f}(p_{01}) \geq \omega^*(\Delta + t_f)$. If the transmission is successful, go to (a); otherwise, go to (b).
Note that waiting time t_s is constant while t_f is AoI dependent.

Next, we show that the optimal policy for the original problem (10) is a mixture of no more than two stationary deterministic threshold-type policies.

Corollary 2. *There exists a stationary randomized policy $\pi^\star$ that is the optimal solution to the constrained average-AoI belief MDP in* (10)*, where $\pi^\star$ is a randomized mixture of threshold-type policies as follows:*

$$\pi^\star = q\pi^\star_{\lambda_1} + (1-q)\pi^\star_{\lambda_2}, \tag{16}$$

where $q \in [0,1]$ is a randomization factor, and $\pi^\star_{\lambda_1}$ and $\pi^\star_{\lambda_2}$ are the optimal threshold-type policies (15) *for some Lagrange multipliers λ_1 and λ_2, respectively.*

Proof. Note that a stationary policy that transmits at the beginning of every $\left\lceil \frac{1}{KE_{\max}} \right\rceil$ frames satisfies energy constraint, where $\lceil \cdot \rceil$ is the ceiling function. Thus, the problem (10) is feasible. Together with our unichain assumption, the result follows from Theorem 4.4 in [34]. □

The method to determine λ_1, λ_2 and q will be discussed in Section 4.2.2.

4.1.2 Proof of Theorem 1

We prove Theorem 1 in two steps: (i) address an unconstrained discounted cost belief MDP; (ii) relate it to the unconstrained average cost belief MDP. In particular, we show that the optimal policy for the unconstrained discounted cost belief MDP is of threshold-type in ω, which implies that the optimal policy for the unconstrained average cost belief MDP is of threshold-type in ω

Given an initial state $\mathbf{s}$, the total expected discounted Lagrangian cost under policy π is given by

$$L^\beta_{\mathbf{s}}(\pi;\lambda) = \limsup_{T\to\infty} \mathbb{E}_\pi\big[\sum_{t=1}^{T} \beta^{t-1} C(\mathbf{s}_t, u_t; \lambda) | \mathbf{s}\big], \tag{17}$$

where $\beta \in (0,1)$ is a discount factor. The optimization problem of minimizing the total expected discounted Lagrangian cost can be cast as

Problem 3 (Unconstrained discounted cost belief MDP):

$$V^\beta(\mathbf{s}) \triangleq \min_{\pi}\ L^\beta_{\mathbf{s}}(\pi;\lambda), \tag{18}$$

where $V^\beta(\mathbf{s})$ denotes the optimal total expected β-discounted Lagrangian cost (for convenience, we omit λ in notation $V^\beta(\mathbf{s})$).

A policy is said to be *β-discounted cost optimal* if it minimizes the total expected β-discounted Lagrangian cost. In Proposition 1, we introduce the optimality equation of $V^\beta(\mathbf{s})$.

Proposition 1. *(a) The optimal total expected β-discounted Lagrangian cost $V^\beta(\Delta,k,\omega)$ satisfies the optimality equation as follows:*

$$V^\beta(\Delta,k,\omega) = \min_{u\in\{0,1\}} Q^\beta(\Delta,k,\omega;u), \tag{19}$$

where

$$Q^\beta(\Delta,k,\omega;0) = \Delta + \beta V^\beta\left(\Delta+1, (k)_+, \mathcal{T}(\omega)\right); \tag{20}$$

$$Q^\beta(\Delta,k,\omega;1) = \Delta + \lambda + \beta\Big(\omega V^\beta\left(k,(k)_+,p_{11}\right) + (1-\omega)V^\beta\left(\Delta+1,(k)_+,p_{01}\right)\Big). \tag{21}$$

(b) A stationary deterministic policy determined by the right-hand-side of (19) *is β-discounted cost optimal.*
(c) Let $V^\beta_n(\mathbf{s})$ be the cost-to-go function such that $V^\beta_0(\mathbf{s}) = 0$, for all $\mathbf{s} \in \mathcal{S}$ and for $n \geq 0$,

$$V^\beta_{n+1}(\Delta,k,\omega) = \min_{u\in\{0,1\}} Q^\beta_{n+1}(\Delta,k,\omega;u), \tag{22}$$

where

$$Q^\beta_{n+1}(\Delta,k,\omega;0) = \Delta + \beta V^\beta_n\left(\Delta+1, (k)_+, \mathcal{T}(\omega)\right); \tag{23}$$

$$Q^\beta_{n+1}(\Delta,k,\omega;1) = \Delta + \lambda + \beta\Big(\omega V^\beta_n\left(k,(k)_+,p_{11}\right) + (1-\omega)V^\beta_n\left(\Delta+1,(k)_+,p_{01}\right)\Big). \tag{24}$$

Then, we have $V^\beta_n(\mathbf{s}) \to V^\beta(\mathbf{s})$ as $n\to\infty$, for every $\mathbf{s}$, β.

Proof. According to [35], it suffices to show that there exists a stationary deterministic policy f such that for all $\beta, \mathbf{s}$, we have $L^\beta_{\mathbf{s}}(f;\lambda) < \infty$. Let f be a policy that chooses $u=0$ for every time slot. For any initial state $\mathbf{s}_1 = (\Delta,t,\omega)$ under this policy, we have

$$\begin{aligned} L^\beta_{\mathbf{s}_1}(f;\lambda) &= \limsup_{T\to\infty} \mathbb{E}_f\big[\sum_{t=1}^{T} \beta^{t-1} C(\mathbf{s}_t, 0; \lambda) | \mathbf{s}_1\big] \\ &= \sum_{n=0}^{\infty} \beta^n(\Delta+n) \\ &= \frac{\Delta}{1-\beta} + \frac{\beta}{(1-\beta)^2} < \infty. \end{aligned}$$

□

Using (c) in Proposition 1, we show properties of V^β in Lemma 1.

Lemma 1. *If $p_{11} \geq p_{01}$, then the value function V^β has the following properties:*
(a) $V^\beta(\Delta,k,\omega)$ is non-decreasing with regard to age Δ.
(b) $V^\beta(\Delta,k,\omega)$ is non-increasing with regard to belief ω.
(c) For beliefs x,y,z,ω that satisfy $z = \omega x + (1-\omega)y$ and $x \geq y$, we have

$$(1-\omega)\lambda + \omega V^\beta(\Delta,k,x) + (1-\omega)V^\beta(\Delta,k,y) \geq V^\beta(\Delta,k,z). \tag{25}$$

(d) The optimal policy corresponding to V^β is of a threshold-type in ω, i.e. given Δ, k, there exists a threshold $\omega^\star_\beta(\Delta,k;\lambda)$ such that it is optimal to transmit only when $\omega \geq \omega^\star_\beta(\Delta,k;\lambda)$.

Proof. Please see Appendix A. □

By (d) in Lemma 1, the β-discounted cost optimal policies are of threshold-type in belief. By [35], under certain

conditions (A proof of these conditions verification is provided in Appendix B), average cost optimal policy can be viewed as a limit of a sequence of β-discounted cost optimal policies as $\beta \to 1$. Thus, the average cost optimal policies are of threshold-type in belief.

4.2 Structure-Aware Algorithm Design

It is known that MDP suffers from the curse of dimensionality. Thus, in this section, we utilize the structure obtained in the last section to design an algorithm with reduced complexity. In particular, we exploit Corollary 2 to design a structure-aware algorithm for (10) in two steps: We first design a structure-aware algorithm for (13), and then construct a way to determine parameters λ_1, λ_2 and q.

4.2.1 Structure-Aware Algorithm for the approximate unconstrained average cost belief MDP

In practice, classic value iteration cannot work if the state space is infinite, since an infinite number of Q-functions associated with the infinite state space need to be updated for each iteration. To deal with this, we first propose a finite-state approximation for infinite-state belief MDP in (13) and show the convergence of our approximate belief MDPs to the original one. Based on this, we propose an optimal structure-aware algorithm for the approximate belief MDP.

Let N be an upper bound for the AoI and the number of Markov transitions from p_{01} or p_{11}. Since $\mathcal{T}^i(p_{01}) \leq \mathcal{T}^{i+1}(p_{01})$ and $\mathcal{T}^i(p_{11}) \geq \mathcal{T}^{i+1}(p_{11})$ for $i \in \mathbb{N}$, we have that with bound N, the state space of the approximate belief MDP is given by $\mathcal{S}^N \triangleq \{(\Delta, k, \omega) \in \mathcal{S} : k \in \mathcal{K}, \Delta \in \mathcal{A}_k, \Delta \leq N, p_{01} \leq \omega \leq \mathcal{T}^N(p_{01}) \text{ or } \mathcal{T}^N(p_{11}) \leq \omega \leq p_{11}\}$. Without loss of generality, we assume $N > K$.

Given the state $(\Delta_t, k_t, \omega_t) \in \mathcal{S}^N$, the state $\mathbf{s}_{t+1} = (\Delta_{t+1}, k_{t+1}, \omega_{t+1}) \in \mathcal{S}^N$ is updated as follows:

$$\mathbf{s}_{t+1}=\begin{cases}(k_t,(k_t)_+,p_{11}) & \text{if } u_t=1,\theta_t=1,\\ (\phi(\Delta_t+1),(k_t)_+,p_{01}) & \text{if } u_t=1,\theta_t=0,\\ (\phi(\Delta_t+1),(k_t)_+,\varphi(\mathcal{T}(\omega_t))) & \text{if } u_t=0,\end{cases} \tag{26}$$

where $\phi(x) = \min\{x, N\}$, and $\varphi(y)$ is given by[1]

$$\varphi(y) = \begin{cases}\mathcal{T}^N(p_{11}) & \text{if } \mathcal{T}^N(p_{01}) < y < \mathcal{T}^N(p_{11}),\\ y & \text{otherwise.}\end{cases} \tag{27}$$

Given action u, the transition probability from $\mathbf{s}$ to $\mathbf{s}'$ on state space $\mathcal{S}^N$, denoted by $P^N_{\mathbf{ss}'}(u)$, is expressed as

$$P^N_{\mathbf{ss}'}(u) = P_{\mathbf{ss}'}(u) + \sum_{\mathbf{r}\in\mathcal{S}-\mathcal{S}^N} P_{\mathbf{sr}}(u)\mathbb{1}_{\{\nu(\mathbf{r})=\mathbf{s}'\}}, \tag{28}$$

where $P_{\mathbf{ss}'}(u)$ and $P_{\mathbf{sr}}(u)$ are the transition probabilities on $\mathcal{S}$ defined in (5), $\mathbb{1}_{\{\cdot\}}$ is the indicator function, and approximation operation to state is

$$\nu\left((z1, z2, z3)\right) \triangleq (\phi(z1), z2, \varphi(z3)). \tag{29}$$

In general, a sequence of approximate MDPs may not converge to the original MDP [36]. In Theorem 2, we show the convergence of our approximate MDPs to the original MDP.

Theorem 2. *Let $\bar{L}^{N\star}(\lambda)$ be the minimum average Lagrangian cost for the approximate MDP with regard to bound N and Lagrange multiplier λ. Then, $\bar{L}^{N\star}(\lambda) \to \bar{L}^{\star}(\lambda)$ as $N \to \infty$.*

Proof. Please see Appendix C. □

The Relative Value Iteration (RVI) algorithm can be utilized to obtain an optimal stationary deterministic policy for the approximate MDP. In particular, RVI starts with $V^N_0(\mathbf{s}) = 0$, $\forall \mathbf{s} \in \mathcal{S}^N$ and updates $V^N_{n+1}(\mathbf{s})$ by minimizing the RHS of equation (30) in the $(n+1)$-th iteration, $n \in \{0, 1, 2, \cdots\}$.

$$V^N_{n+1}(\mathbf{s}) = \min_u \Big\{C(\mathbf{s}, u; \lambda) + \sum_{\mathbf{s}'\in\mathcal{S}^N} P^N_{\mathbf{ss}'}(u)h^N_n(\mathbf{s}') - h^N_n(\mathbf{0})\Big\}, \tag{30}$$

where $\mathbf{0}$ is the reference state and $h^N_n(\mathbf{s}) = V^N_n(\mathbf{s}) - V^N_n(\mathbf{0})$. Note that similar to the proof in Section 4.1, it can be shown that the optimal policy for the approximate MDP is still of threshold-type. Thus, we utilize the threshold property in RVI algorithm and propose a threshold-type RVI to reduce the complexity in Algorithm 1 (Line 4-24). Specifically, in each iteration n, we need to update the optimal action u^* for all states by minimizing the right-hand-side of (30). Since the size of the state space $\mathcal{S}^N$ is $O(N \cdot K \cdot (2N+2))$, the computation complexity of updating actions for all states in each iteration is $O(N \cdot K \cdot (2N+2))$ without using threshold property. But with the threshold property, if certain state satisfies the threshold condition (Line 11), then the optimal action for the state is determined *immediately* without doing the minimization operation (Line 12), which reduces the algorithm complexity greatly. The reduction degree is affected by the optimal thresholds and is hard to be quantified. To give a flavor of the reduction, we provide a simulation result here. In our simulation with parameters $\epsilon = 0.001$, $N = 500$, $K = 4$, $p_{11} = 0.6$ and $p_{01} = 0.2$, the times of minimization operation without the threshold property (71854272 times) is around 31 times the one with the threshold property (2220468 times) till the termination of loop in Line 4-24. Another benefit of introducing the threshold property is to reduce the memory for storing optimal policies. Without the threshold property, we have to store actions for all states, which consumes $O(N \cdot K \cdot (2N+2))$. But with the threshold property, we only need to store thresholds for each pair of the AoI and the relative slot index, which consumes $O(N \cdot K)$.

4.2.2 Lagrange Multiplier Estimation

By Lemma 3.4 of [37], for $\lambda_1 < \lambda_2$, we have $\bar{A}(\pi^\star_{\lambda_1}) \leq \bar{A}(\pi^\star_{\lambda_2})$ and $\bar{E}(\pi^\star_{\lambda_1}) \geq \bar{E}(\pi^\star_{\lambda_2})$. Thus, the optimal Lagrangian multiplier $\lambda^\star$ is defined as $\lambda^\star \triangleq \inf\{\lambda > 0 : \bar{E}(\pi^\star_\lambda) \leq E_{\max}\}$. If there exists $\lambda^\star$ such that $\bar{E}(\pi^\star_{\lambda^\star}) = E_{\max}$, then the constrained average-AoI optimal policy is a stationary deterministic policy where q in Corollary 2 is either 0 or 1. Otherwise, the optimal policy $\pi^\star$ chooses policy $\pi^\star_{\lambda^{\star}-}$ with probability

1. We upper bound the belief state by $\mathcal{T}^N(p_{11})$. This ensures that the optimal policy for the approximate unconstrained belief MDP is of threshold-type.

Algorithm 1: Structure-Aware Scheduling without channel sensing

1 Given tolerance $\epsilon > 0, \epsilon_\lambda > 0$, $\lambda^{\star-}$, $\lambda^{\star+}$, N ;
2 **while** $|\lambda^{\star+} - \lambda^{\star-}| > \epsilon_\lambda$ **do**
3 $\lambda = (\lambda^{\star+} + \lambda^{\star-})/2$;
4 $V^N(\mathbf{s}) = 0, h^N(\mathbf{s}) = 0, h^N_{prev}(\mathbf{s}) = \infty$, for all $\mathbf{s} \in \mathcal{S}^N$;
5 **while** $\max_{\mathbf{s}\in\mathcal{S}^N} |h^N(\mathbf{s}) - h^N_{prev}(\mathbf{s})| > \epsilon$ **do**
6 $\omega^\star(\Delta, k; \lambda) = \infty$ for all $\mathbf{s} = (\Delta, k, \omega) \in \mathcal{S}^N$;
7 **foreach** $\mathbf{s} = (\Delta, k, \omega) \in \mathcal{S}^N$ **do**
8 **if** $\Delta < K$ **then**
9 $u^\star = 0$;
10 **else**
11 **if** $\omega \geq \omega^\star(\Delta, k; \lambda)$ **then**
12 $u^\star = 1$;
13 **else**
14 $u^\star = \arg\min_{u\in\{0,1\}}\{C(\mathbf{s}, u; \lambda) + \sum_{\mathbf{s}'\in\mathcal{S}^N} P^N_{\mathbf{ss}'}(u) h^N(\mathbf{s}')\}$;
15 **if** $u^\star = 1$ **then**
16 $\omega^\star(\Delta, k; \lambda) = \omega$;
17 **end**
18 **end**
19 **end**
20 $V^N(\mathbf{s}) = C(\mathbf{s}, u^\star; \lambda) + \sum_{\mathbf{s}'\in\mathcal{S}^N} P^N_{\mathbf{ss}'}(u^\star) h^N(\mathbf{s}') - h^N(\mathbf{0})$;
21 $h^N_{prev}(\mathbf{s}) = h^N(\mathbf{s})$;
22 $h^N(\mathbf{s}) = V^N(\mathbf{s}) - V^N(\mathbf{0})$;
23 **end**
24 **end**
25 Compute the average energy cost $\bar{E}(\lambda)$;
26 **if** $\bar{E}(\lambda) > E_{max}$ **then**
27 $\lambda^{\star-} = \lambda$;
28 **else**
29 $\lambda^{\star+} = \lambda$;
30 **end**
31 **end**

q and policy $\pi^\star_{\lambda^{\star+}}$ with probability $1-q$. The randomization factor q can be computed by

$$q = \frac{E_{\max} - \bar{E}(\pi^\star_{\lambda^{\star+}})}{\bar{E}(\pi^\star_{\lambda^{\star-}}) - \bar{E}(\pi^\star_{\lambda^{\star+}})}. \tag{31}$$

The bisection method is used to compute $\lambda^{\star-}$, $\lambda^{\star+}$ and thus q (Line 2-3 and Line 26-30 in Algorithm 1). The algorithm starts with $\lambda^{\star-} = 0$ and sufficiently large $\lambda^{\star+}$.

5 Scheduling with Delayed Channel Sensing

With delayed channel sensing, the CSI of the last time slot is always available at the beginning of each slot. Thus, the problem in this case can be formulated as a constrained MDP. The state space reduces to $\mathcal{S} \triangleq \{(\Delta, k, g) : k \in \mathcal{K}, \Delta \in \mathcal{A}_k, g \in \{0, 1\}\}$, where g denotes the CSI of the last time slot. Given $\mathbf{s}_t = (\Delta_t, k_t, g_t)$ and u_t at time slot t, the transition probability to $\mathbf{s}_{t+1} = (\Delta_{t+1}, k_{t+1}, g_{t+1})$ is written as follows:

$$P_{\mathbf{s}_t\mathbf{s}_{t+1}}(u_t) = \begin{cases} p_{g_t 1} & \text{if } u_t = 1, \mathbf{s}_{t+1} = (k_t, (k_t)_+, 1), \\ 1 - p_{g_t 1} & \text{if } u_t = 1, \mathbf{s}_{t+1} = (\Delta_t + 1, (k_t)_+, 0), \\ p_{g_t g_{t+1}} & \text{if } u_t = 0, \mathbf{s}_{t+1} = (\Delta_t + 1, (k_t)_+, g_{t+1}). \end{cases} \tag{32}$$

Following Section 3.2 and Section 4, the optimal transmission scheduling policy in this case is also a randomized mixture of no more than two stationary deterministic policies, each of which is optimal for an unconstrained average cost MDP. But thanks to the simplification in state, we can show that the optimal policy for the unconstrained average cost MDP in this case is of threshold-type in AoI in Theorem 3.

Theorem 3. *Given Lagrange multiplier λ, there exists a stationary unconstrained average cost optimal policy that is deterministic and of threshold-type in AoI. Specifically, the policy is in the form $\pi^\star_\lambda = (d^\star_\lambda, d^\star_\lambda, \cdots)$, where*

$$d^\star_\lambda(\Delta, k, g) = \begin{cases} 0 & \text{if } 0 \leq \Delta < \Delta^\star(k, g; \lambda), \\ 1 & \text{if } \Delta^\star(k, g; \lambda) \leq \Delta, \end{cases} \tag{33}$$

and

$$\Delta^\star(k, 1; \lambda) \leq \Delta^\star(k, 0; \lambda), \tag{34}$$

where $\Delta^\star(k, g; \lambda)$ denotes the threshold given pair of relative slot index and delayed CSI (k, g) and Lagrange multiplier λ.

Different from Theorem 1 which provides a threshold structure in the belief ω, Theorem 3 obtains that (i) the average cost optimal policy is of threshold-type in AoI, and (ii) threshold when $g = 1$ is no larger than the threshold when $g = 0$. Indeed, (ii) is used in algorithm to further reduce algorithm complexity. In particular, similar to Section 4.2.1, we bound AoI with N and propose a threshold-type algorithm in Algorithm 2 to minimize unconstrained average cost. Different from corresponding part in Algorithm 1, $\Delta^\star(k, 1; \lambda)$ is updated along with each threshold updating (Line 15) to keep the threshold relation in (34). This further reduces algorithm complexity.

Algorithm 2: Threshold-type scheduling for unconstrained average cost MDP with delayed channel sensing

1 Given tolerance $\epsilon > 0$, Lagrange multiplier λ and bound N ;
2 $V^N(\mathbf{s}) = 0, h^N(\mathbf{s}) = 0, h^N_{prev}(\mathbf{s}) = \infty$, for all $\mathbf{s} \in \mathcal{S}^N$;
3 **while** $\max_{\mathbf{s}\in\mathcal{S}^N} |h^N(\mathbf{s}) - h^N_{prev}(\mathbf{s})| > \epsilon$ **do**
4 $\Delta^\star(k, g; \lambda) = \infty$ for all $\mathbf{s} = (\Delta, k, g) \in \mathcal{S}^N$;
5 **foreach** $\mathbf{s} = (\Delta, k, g) \in \mathcal{S}^N$ **do**
6 **if** $\Delta < K$ **then**
7 $u^\star = 0$;
8 **else**
9 **if** $\Delta \geq \Delta^\star(k, g; \lambda)$ **then**
10 $u^\star = 1$;
11 **else**
12 $u^\star = \arg\min_{u\in\{0,1\}}\{C(\mathbf{s}, u; \lambda) + \sum_{\mathbf{s}'\in\mathcal{S}^N} P^N_{\mathbf{ss}'}(u) h^N(\mathbf{s}')\}$;
13 **if** $u^\star = 1$ **then**
14 $\Delta^\star(k, g; \lambda) = \Delta$;
15 $\Delta^\star(k, 1; \lambda) = \min\{\Delta, \Delta^\star(k, 1; \lambda)\}$;
16 **end**
17 **end**
18 **end**
19 $V^N(\mathbf{s}) = C(\mathbf{s}, u^\star; \lambda) + \sum_{\mathbf{s}'\in\mathcal{S}^N} P^N_{\mathbf{ss}'}(u^\star) h^N(\mathbf{s}') - h^N(\mathbf{0})$;
20 $h^N_{prev}(\mathbf{s}) = h^N(\mathbf{s})$;
21 $h^N(\mathbf{s}) = V^N(\mathbf{s}) - V^N(\mathbf{0})$;
22 **end**
23 **end**

The proof idea of Theorem 3 is similar to Theorem 1. We relate average cost MDPs to discounted cost MDPs. Next, we explore the structure of discounted cost optimal policies.

The optimality equation in (19) is modified as follows:

$$V^{\beta}(\Delta,k,g)=\Delta+\beta\min\Big\{\sum_{g'\in\{0,1\}}p_{gg'}V^{\beta}(\Delta+1,(k)_{+},g'),$$
$$\lambda+p_{g1}V^{\beta}\left(k,(k)_{+},1\right)+p_{g0}V^{\beta}\left(\Delta+1,(k)_{+},0\right)\Big\}. \quad (35)$$

First, we prove the monotonicity of value function V^{β} in AoI in Lemma 2.

Lemma 2. *The function $V^{\beta}(\Delta,k,g)$ is non-decreasing with regard to AoI Δ.*

Proof. Please see Appendix D. □

With this, we characterize the structure of optimal policy for the unconstrained discounted cost MDP in Lemma 3.

Lemma 3. *Given λ and β, the optimal policy that minimizes the β-discounted Lagrangian cost is of threshold-type in AoI Δ, i.e. given k,g, there exists a threshold $\Delta^{\star}_{\beta}(k,g;\lambda)$ such that it is optimal to transmit only when $\Delta\geq\Delta^{\star}_{\beta}(k,g;\lambda)$. In addition, $\Delta^{\star}_{\beta}(k,1;\lambda)\leq\Delta^{\star}_{\beta}(k,0;\lambda)$.*

Proof. Please see Appendix E. □

Similar to the proof of Theorem 1, we can extend the result to the unconstrained average cost MDP as in Theorem 3.

6 A SPECIAL CASE WITH $K=1$ AND $p_{01}=p_{11}$

In this section, we analyze optimal policies for a special case with $K=1$ and $p_{01}=p_{11}$. Let $p=p_{01}$. Note that if $p_{01}=p_{11}$, then $\mathcal{T}^{m}(p_{01})=\mathcal{T}^{m}(p_{11})=p$ for $m\geq 0$. Then, for the case without channel sensing, the belief state $\omega=p$ at each time slot. On the other side, for the case with channel sensing, the transition probability $p_{g1}=p$ for $g\in\{0,1\}$. Together with $K=1$, the optimality equations in both cases (Eq. (19) and (35)) reduce to the same equation as follows:

$$V^{\beta}(\Delta)=\Delta+\beta\min\Big\{V^{\beta}(\Delta+1),$$
$$\lambda+pV^{\beta}(1)+(1-p)V^{\beta}(\Delta+1)\Big\}. \quad (36)$$

With similar proof of Theorem 3, the optimal policies for the unconstrained MDPs in both cases when $K=1$ and $p_{01}=p_{11}$ features a threshold policy as in Theorem 4.

Theorem 4. *Given Lagrange multiplier λ, there exists a stationary unconstrained average cost optimal policy that is deterministic and of threshold-type in AoI. Specifically, the policy is in the form $\pi^{\star}_{\lambda}=(d^{\star}_{\lambda},d^{\star}_{\lambda},\cdots)$, where*

$$d^{\star}_{\lambda}(\Delta)=\begin{cases}0 & \text{if } 0\leq\Delta<\Delta^{\star}_{\lambda},\\ 1 & \text{if } \Delta^{\star}_{\lambda}\leq\Delta.\end{cases} \quad (37)$$

The policy above can be viewed as a special version of the (33) for the case with channel sensing. If we let $\omega^{\star}(\Delta;\lambda)=1$ for $\Delta<\Delta^{\star}_{\lambda}$ and $\omega^{\star}(\Delta;\lambda)=0$ for $\Delta\geq\Delta^{\star}_{\lambda}$, then the policy can be viewed as a special version of the policy (15) for the case without channel sensing.

For the original problem in both cases, the optimal solution is a randomized mixture of threshold-type policies as in Corollary 2. Thanks to the simplification, the optimal can be easily presented in Fig. 5. The dashed curves correspond to the optimal threshold-type policies in (37) with $\lambda^{\star}_{1}$ and $\lambda^{\star}_{2}$, where $\lambda^{\star}_{1}\leq\lambda^{\star}_{2}$. The solid curve represents the optimal randomized policy for the original problem. For $\Delta^{\star}_{\lambda^{\star}_{1}}\leq\Delta<\Delta^{\star}_{\lambda^{\star}_{2}}$, the probability of transmitting is the randomization factor q.

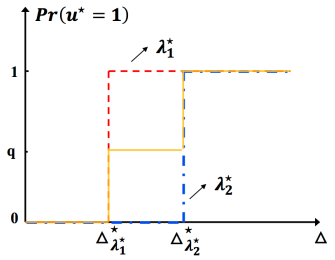


Fig. 5: Optimal policy for the constrained MDPs in both cases with $K=1$ and $p_{01}=p_{11}$. The dashed curves correspond to the optimal policies for the unconstrained MDPs with $\lambda^{\star}_{1}$ and $\lambda^{\star}_{2}$. The solid curve represents the optimal randomized policy for the constrained MDP.

7 NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the proposed algorithms.

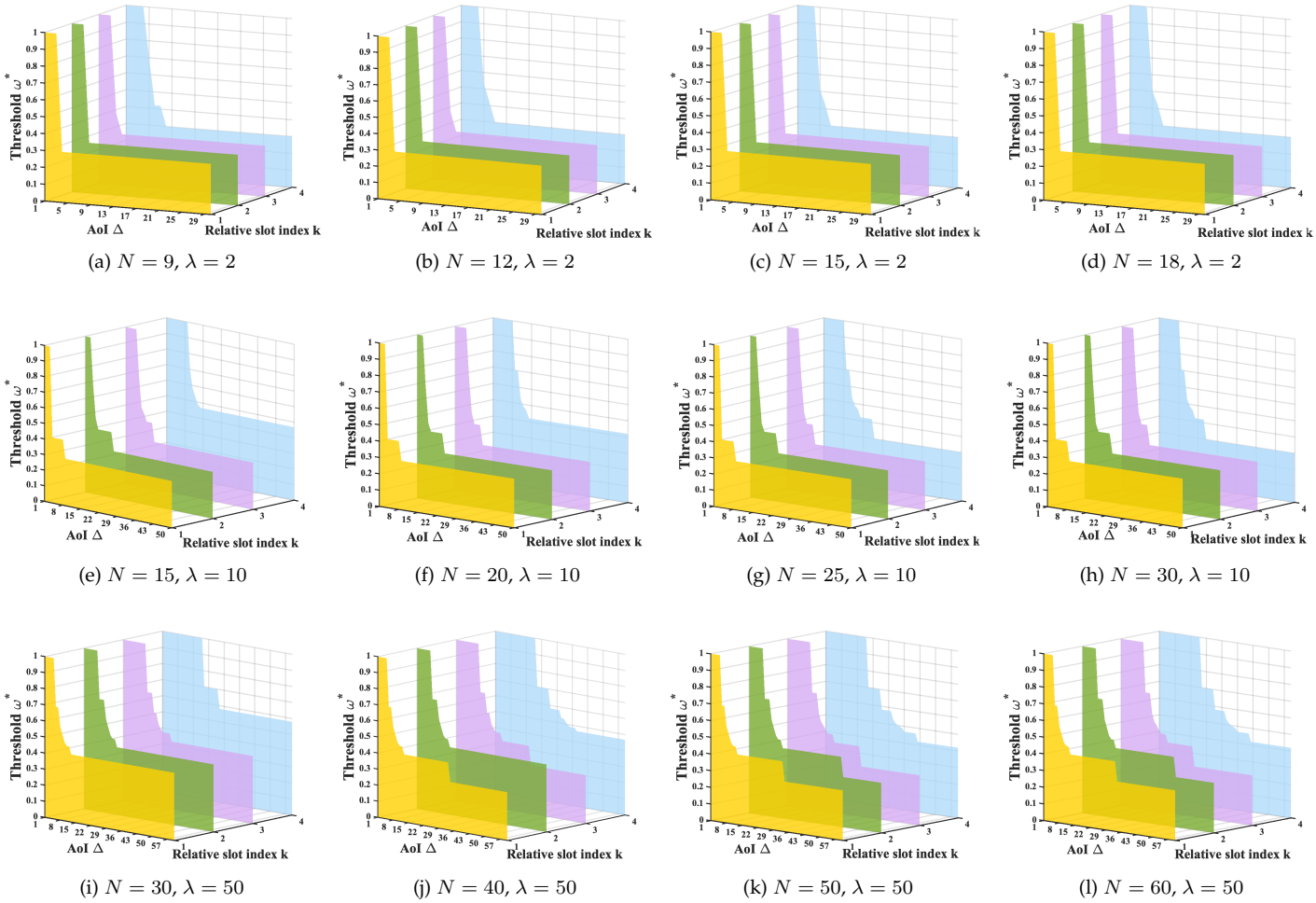
7.1 Policy estimation for the unconstrained infinite-state MDP

Recall that in our setting, the state space is infinite due to the unreliable channel. Thus, value iteration cannot be used directly to obtain the optimal policy for the unconstrained MDP in either case. To handle this, we proposed finite-state approximations with upper bound N in each case, and then proposed structure-aware policies in Algorithm 1 and Algorithm 2 to obtain the optimal policies for the approximate unconstrained MDPs in the two cases, respectively.

For the case without channel sensing, Line 5 - Line 24 in Algorithm 1 are used to obtain a structure-aware optimal policy for the unconstrained approximate MDP given bound N and Lagrangian multiplier λ. The optimal policy features a threshold-type in belief state, and the optimal threshold ω^{*} is a function of the AoI Δ and the relative slot index k. Since we use N to bound the AoI, the resulting policy only has threshold information for the AoI below $N+1$. To apply this policy to the infinite-state MDP, we set $\omega^{*}(\Delta,k)=\omega^{*}(N,k)$, for $\forall\Delta>N$.

In Fig. 6, we investigate how N and λ affect the resulting policy for the unconstrained infinite-state MDP. In particular, each figure in Fig. 6 presents the optimal threshold as a function of Δ and k. Note that given k, the optimal threshold corresponding to the AoI outside the AoI range in the figure is the same as the one corresponding to the largest AoI in the figure. In simulation, we set $p_{11}=0.7$, $p_{01}=0.3$, and $K=4$. We can observe that with fixed λ, i.e. Fig. 6a-6d, the policy converges as N increases. In addition, by viewing results from the first row to the last row, we can find that as λ increases, policies converges on larger N.

Further, we use the average Lagrangian cost to present the performance of the resulting policy and the number of times that the value function $V^{N}(\cdot)$ is updated till the termination of the structure-aware RVI in Algorithm 1 to

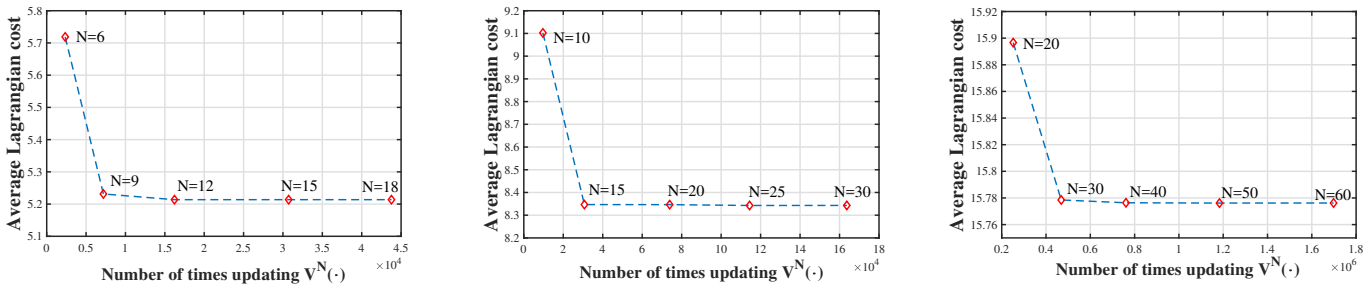


(a) $N = 9, \lambda = 2$ (b) $N = 12, \lambda = 2$ (c) $N = 15, \lambda = 2$ (d) $N = 18, \lambda = 2$

(e) $N = 15, \lambda = 10$ (f) $N = 20, \lambda = 10$ (g) $N = 25, \lambda = 10$ (h) $N = 30, \lambda = 10$

(i) $N = 30, \lambda = 50$ (j) $N = 40, \lambda = 50$ (k) $N = 50, \lambda = 50$ (l) $N = 60, \lambda = 50$

Fig. 6: Policy estimation with different values of N and λ (without channel sensing)



(a) $\lambda = 2$ (b) $\lambda = 10$ (c) $\lambda = 50$

Fig. 7: Average Lagrangian cost vs updating times with different values of N (without channel sensing)

present the complexity. In Fig. 7, we study the average Lagrangian cost and the number of updating times by varying N, which implies the trade-off between complexity and accuracy based on the choice of N. We set $\epsilon = 0.01$ in simulation. We observe that (i) as N increases, the average Lagrangian cost decreases more slowly while the number of times for updating $V^N(\cdot)$ keeps increasing, and (ii) the average Lagrangian cost remains the same after N exceeds a certain value. It provides numerical support to our proof of convergence for the policy.

Similarly, we conduct simulations for the case with channel sensing. Compared to the case without channel sensing, the difference is that the threshold in this case is in AoI and the threshold AoI is a function of the relative slot index k and the CSI of the last time slot g. Since the threshold can be ∞, we use a table to denote the estimated policy for the unconstrained infinite-state MDP as shown in Table 1-Table 9. The three consecutive tables in a row represent the estimated policy versus N and each row corresponds to a certain λ. Fig. 8 studies the average Lagrangian cost and the number of times for updating value functions till the termination of Algorithm 2 by varying values of N. The trade-off is similar as the case without channel sensing. But in this case, it takes smaller number of times to converge given N.

TABLE 1: Estimated Policy given $\lambda = 2$ and $N = 5$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	3	5	5	5
	1	1	2	4	5

TABLE 2: Estimated Policy given $\lambda = 2$ and $N = 10$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	3	4	5	7
	1	1	2	3	5

TABLE 3: Estimated Policy given $\lambda = 2$ and $15 \leq N \leq 1000$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	2	4	5	7
	1	1	2	3	5

TABLE 4: Estimated Policy given $\lambda = 10$ and $N = 20$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	9	11	13	20
	1	3	4	6	9

TABLE 5: Estimated Policy given $\lambda = 10$ and $N = 25$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	9	11	12	19
	1	3	4	6	9

TABLE 6: Estimated Policy given $\lambda = 10$ and $30 \leq N \leq 1000$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	9	11	12	18
	1	3	4	6	9

TABLE 7: Estimated Policy given $\lambda = 50$ and $N = 35$ (with channel sensing)

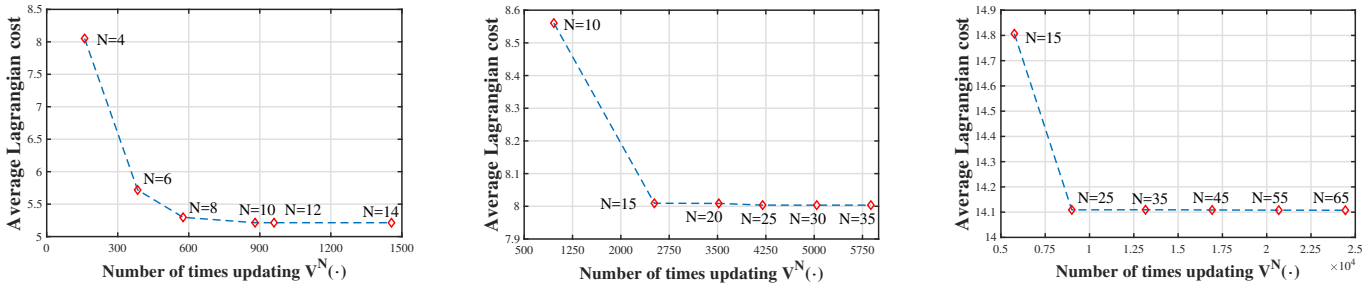
Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	∞	∞	∞	∞
	1	8	10	13	21

TABLE 8: Estimated Policy given $\lambda = 50$ and $N = 50$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	39	42	44	∞
	1	8	10	13	21

TABLE 9: Estimated Policy given $\lambda = 50$ and $65 \leq N \leq 1000$ (with channel sensing)

Threshold Δ^*		Relative slot index k			
		1	2	3	4
CSI of last slot g	0	39	42	44	61
	1	8	10	13	21



(a) $\lambda = 2$ (b) $\lambda = 10$ (c) $\lambda = 50$

Fig. 8: Average Lagrangian cost vs updating times with different values of N (with channel sensing)

This is mainly because that the number of combinations of (k, g) is $2K$ while the number of combinations of (Δ, k) is NK.

In the remaining sections, we assume $N = 1000$ and obtain all simulation results over 10^5 time slots.

7.2 Average AoI Performance

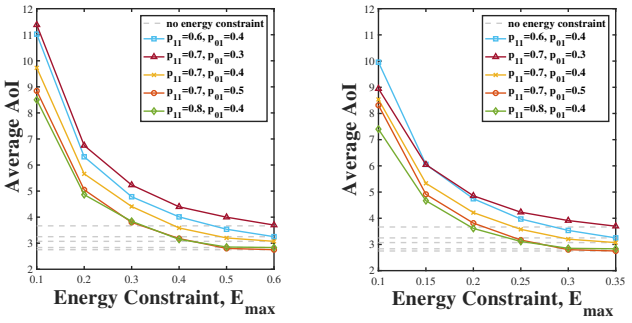
Fig. 9 plots the AoI-energy tradeoff with different fading characteristics (different p_{11} and p_{01}) for the two cases that we consider in this paper. In this simulation, we set $K = 3$. The optimal average AoI with no energy constraint is plotted as a gray dashed line accordingly. When comparing Fig. 9a with Fig. 9b, it is easy to observe that for fixed energy constraint and pair of p_{11} and p_{01}, the average AoI with delayed channel sensing is no larger than that without channel sensing.

Moreover, the curves in Fig. 9a and Fig. 9b exhibit the same trend as follows. For each pair of p_{11} and p_{01}, average AoI decreases with energy constraint. Note that it is prohibited to transmit delivered status update. Thus, even if there is no energy constraint, obtaining the optimal average AoI does not necessarily imply transmitting at every time slot. This explains why the average AoI achieved by our proposed policies approaches the gray line even when $E_{\max} \neq 1$. In addition, we can observe that for certain energy constraint, the average AoI decreases with either p_{11} or p_{01}. This is due to the fact that increase in either p_{11} or p_{01} results in the increase of steady state probability that channel is in good state.

Fig. 10 studies the average AoI performance vs frame length with different fading characteristics in the two cases. We set the energy constraint $E_{\max} = 0.3$. Compare Fig. 10a with Fig. 10b, we have same observations in Fig. 9 that average AoI is smaller in the case with delayed channel sensing. Except this, average AoI increases with frame length. This is reasonable since increase of frame length means less frequent generation of status updates. We can also find that the increasing rate is changing. This is actually the outcome of both frame length and energy constraints.

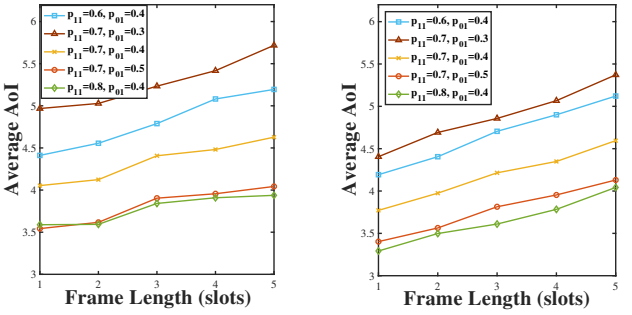
7.3 Comparison with greedy policy

Let e_t denote total energy consumption before slot t. Then, $\bar{e}_t \triangleq e_t/(t-1)$ denotes the average energy consumed before



(a) Without channel sensing (b) With delayed channel sensing

Fig. 9: AoI-energy tradeoff with different transition probabilities



(a) Without channel sensing (b) With delayed channel sensing

Fig. 10: Average AoI vs frame length with different transition probabilities

slot t. We compare the proposed transmission scheduling policies with a greedy policy that transmits when $\bar{e}_t < E_{\max}$ and $\Delta_t \geq K$. We set $K = 3$, $p_{11} = 0.7$, $p_{01} = 0.3$, in which case the optimal AoI with no energy constraint is achieved with 0.6167 units energy on average. Thus, the comparison is conducted with energy constraint ranging from 0.1 to 0.6. In Fig. 11, it is easy to observe that the proposed transmission scheduling policy outperforms the greedy policy in both cases. The gap between the greedy policy and scheduling policy in either case narrows as the energy constraint is loosened.

7.4 Comparison with fixed-threshold policy

Recall that the proposed transmission scheduling policies feature multi-thresholds and the randomization factor q. We may wonder whether a fixed-threshold policy can be used instead in practice to simplify manipulation. In this section, we compare the average AoI performance of the proposed policies with an optimized fixed-threshold policy. In particular, the optimized fixed-threshold policy refers to the one that has the smallest average AoI with energy constraint satisfied among all fixed-threshold policies. For the case without channel sensing, the fixed threshold ω^* is chosen from $\{\omega : p_{01} \leq \omega \leq \mathcal{T}^N(p_{01})$ or $\mathcal{T}^N(p_{11}) \leq \omega \leq p_{11}\}$. For the case with channel sensing, the fixed threshold Δ^* is chosen from $\{1, \cdots, N\}$.

As a metric for the comparison, we define average AoI increase $\hat{A}$ as

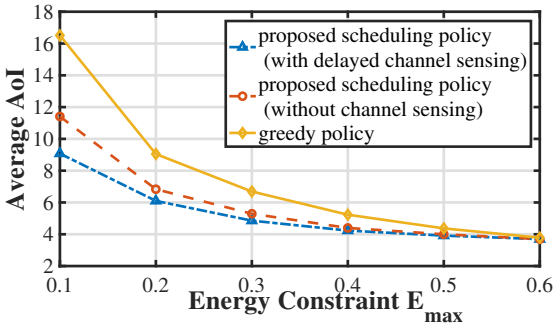


Fig. 11: Comparison with greedy policy

$$\hat{A} = \frac{\bar{A}^*_{\text{fixed}} - \bar{A}^*_{\text{multi}}}{\bar{A}^*_{\text{multi}}} \tag{38}$$

where $\bar{A}^*_{\text{multi}}$, $\bar{A}^*_{\text{fixed}}$ denote the average AoI obtained with our proposed policy and the optimized fixed-threshold policy, respectively.

In Table 10, we investigated average AoI increase $\hat{A}$ in scenarios with different combinations of channel conditions ($(p_{11}, p_{01}) \in \{(0.6, 0.4), (0.7, 0.3), (0.7, 0.4), (0.7, 0.5), (0.8, 0.4)\}$), energy constraints ($E_{\max} \in \{0.2, 0.5, 0.8\}$) and frame lengths ($K \in \{1, 2, 3\}$). We provide 19 scenarios in total. Recall that it is prohibited to retransmit delivered status update. Thus, even if there is no energy constraint, obtaining the optimal average AoI does not necessarily imply transmitting at every time slot. In particular, for all the scenarios with $K = 1$, the optimal policy without energy constraint is to transmit at every time slot since there is an update at every time slot. For scenarios with $p_{11} = 0.7$ and $p_{01} = 0.3$, the optimal AoI with no energy constraint is achieved with 0.75 and 0.6167 units energy on average when $K = 2$ and $K = 3$, respectively. Thus, we do not provide $E_{\max} = 0.8$ in scenarios with $K = 2$ or $K = 3$.

Scenarios No. 1-15 have the same frame length $K = 1$, in which each group of the three consecutive scenarios (numbered $3l - 2$, $3l - 1$, $3l$, for $l \in \mathbb{N}^+$) have the same channel conditions but with different energy constraints. Scenarios No. 16-17 (No. 18-19) form a group that has the same frame length $K = 2$ ($K = 3$) and channel conditions but with different energy constraints. We can observe in each group that for both cases, $\hat{A}$ mainly decreases with $E_{\max}$. This is because that as $E_{\max}$ becomes larger, the optimal multiple thresholds tend to get closer, which narrows the performance gap between our policies and the optimized fixed-threshold policy. In few situations like No. 11 and No. 12 in the case with channel sensing, the $\hat{A}$ increases with $E_{\max}$. Such kind of situations can be explained as follows. Notice that the threshold of the optimized fixed-threshold policy should be a threshold bounded by the smallest and largest values of optimal thresholds of our policies. Usually, choosing a threshold in middle should be better than using the smallest value as a threshold. Thus, when the optimized fixed-threshold policy has to choose the smallest value as its optimal threshold, $\hat{A}$ may increase. For example, in scenario No. 11, the threshold of the optimized fixed-threshold policy is 3, in the middle of optimal thresh-

TABLE 10: Average AoI increase with different frame lengths, energy constraints and channel conditions in two cases

Scenarios	Frame length K	Channel conditions		Energy constraint $E_{\max}$	Average AoI increase $\hat{A}$ (%)	
		p_{11}	p_{01}		Without channel sensing	With channel sensing
1	1	0.6	0.4	0.2	170.085	13.276
2	1	0.6	0.4	0.5	89.840	12.054
3	1	0.6	0.4	0.8	32.133	5.418
4	1	0.7	0.3	0.2	165.703	39.473
5	1	0.7	0.3	0.5	58.757	12.452
6	1	0.7	0.3	0.8	19.701	5.323
7	1	0.7	0.4	0.2	186.876	27.733
8	1	0.7	0.4	0.5	63.551	18.282
9	1	0.7	0.4	0.8	20.539	7.209
10	1	0.7	0.5	0.2	208.362	9.129
11	1	0.7	0.5	0.5	108.793	4.908
12	1	0.7	0.5	0.8	172.684	9.521
13	1	0.8	0.4	0.2	267.096	31.843
14	1	0.8	0.4	0.5	100.203	9.026
15	1	0.8	0.4	0.8	9.947	9.851
16	2	0.7	0.3	0.2	120.124	35.271
17	2	0.7	0.3	0.5	3.114	11.208
18	3	0.7	0.3	0.2	171.136	42.235
19	3	0.7	0.3	0.5	13.452	28.367

olds $\Delta^*(1,1) = 2, \Delta^*(1,0) = 4$ of our policy. However, in scenario No. 12, the threshold of the optimized fixed-threshold policy is 2, the smallest one of optimal thresholds $\Delta^*(1,1) = 2, \Delta^*(1,0) = 3$ of our policy.

Every three scenarios in No. 1-No.15 differ in channel conditions. We can observe that channel conditions do have impact on $\hat{A}$ but the trend is indeterminate. By comparing scenarios No. 4, No. 16 and No. 18, and scenarios No.5, No. 17 and No. 19, we can observe that the frame length also affects $\hat{A}$ but the trend is indeterminate.

In addition, the average AoI increase in the case without channel sensing is much larger than the case with channel sensing. With the results, we can say that it is necessary to use our policy when energy constraint is not close to the energy required to achieve the optimal AoI without energy constraint, especially for the case without channel sensing. When energy constraint is loose, we need to judge $\hat{A}$ based on the channel conditions and frame length to see whether the fixed-threshold policy can be used with only a small sacrifice in the performance.

8 Conclusion

We studied scheduling transmission of periodically generated updates over a Gilbert-Elliott fading channel in two cases. For the case without channel sensing, the problem is a constrained POMDP and is rewritten as a constrained belief MDP by introducing belief state. We show that the optimal policy for the constrained belief MDP is a randomization of no more than two stationary deterministic policies, each of which is of a threshold-type in the belief on the channel. For the case with delayed channel sensing, we show that the optimal policy has a similar structure as the one in the former case but with AoI associated threshold. In addition, we show that the AoI threshold has monotonic behavior in the delayed channel state in this case. The structure is utilized in either case to reduce algorithm complexity.

Acknowledgments

This work was funded in part through NSF grants: CNS-1955535, CNS-2106932, CNS-2106933, CNS-1901057, CNS-2007231, CNS-1618520, and CNS-1409336, and an Office of Naval Research under Grant N00014-17-1-241.

References

[1] Igor Kadota, Elif Uysal-Biyikoglu, Rahul Singh, and Eytan Modiano. Minimizing the age of information in broadcast wireless networks. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 844–851. IEEE, 2016.

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

[3] Ahmed M Bedewy, Yin Sun, Rahul Singh, and Ness B Shroff. 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*, pages 51–60, 2020.

[4] Deli Qiao and M Cenk Gursoy. Age minimization for status update systems with packet based transmissions over fading channels. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2019.

[5] Ahmed M Bedewy, Yin Sun, Sastry Kompella, and Ness B Shroff. Optimal sampling and scheduling for timely status updates in multi-source networks. *arXiv preprint arXiv:2001.09863*, 2020.

[6] Parisa Rafiee, Peng Zou, Omur Ozel, and Suresh Subramaniam. Maintaining information freshness in power-efficient status update systems. *arXiv preprint arXiv:2003.13577*, 2020.

[7] Ahmed M Bedewy, Yin Sun, and Ness B Shroff. The age of information in multihop networks. *IEEE/ACM Transactions on Networking*, 27(3):1248–1257, 2019.

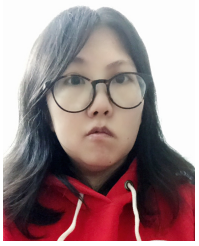
[8] Ahmed M Bedewy, Yin Sun, and Ness B Shroff. Minimizing the age of information through queues. *IEEE Transactions on Information Theory*, 65(8):5215–5232, 2019.

[9] Guidan Yao, Ahmed M. Bedewy, and Ness B. Shroff. Battle between rate and error in minimizing age of information. *arXiv preprint arXiv:2012.09351*, 2020.

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

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

[12] Xianwen Wu, Jing Yang, and Jingxian Wu. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Transactions on Green Communications and Networking*, 2(1):193–204, 2017.
[13] Bo Zhou and Walid Saad. Joint status sampling and updating for minimizing age of information in the internet of things. *IEEE Transactions on Communications*, 67(11):7468–7482, 2019.
[14] Elif Tuğçe Ceran, Deniz Gündüz, and András György. Average age of information with hybrid arq under a resource constraint. *IEEE Transactions on Wireless Communications*, 18(3):1900–1913, 2019.
[15] Ahmed Arafa, Jing Yang, Sennur Ulukus, and H Vincent Poor. Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies. *IEEE Transactions on Information Theory*, 66(1):534–556, 2019.
[16] Haitao Huang, Deli Qiao, and M Cenk Gursoy. Age-energy tradeoff in fading channels with packet-based transmissions. *arXiv preprint arXiv:2005.05610*, 2020.
[17] Ahmed Arafa, Jing Yang, Sennur Ulukus, and H Vincent Poor. Online timely status updates with erasures for energy harvesting sensors. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 966–972. IEEE, 2018.
[18] Songtao Feng and Jing Yang. Age of information minimization for an energy harvesting source with updating erasures: Without and with feedback. *IEEE Transactions on Communications*, 2021.
[19] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.
[20] Qinqing Zhang and Saleem A Kassam. Finite-state markov model for rayleigh fading channels. *IEEE Transactions on communications*, 47(11):1688–1692, 1999.
[21] Edgar N Gilbert. Capacity of a burst-noise channel. *Bell system technical journal*, 39(5):1253–1265, 1960.
[22] Shiyang Leng and Aylin Yener. Age of information minimization for an energy harvesting cognitive radio. *IEEE Transactions on Cognitive Communications and Networking*, 5(2):427–439, 2019.
[23] Jaya Prakash Champati, Hussein Al-Zubaidy, and James Gross. Statistical guarantee optimization for aoi in single-hop and two-hop systems with periodic arrivals. *arXiv preprint arXiv:1910.09949*, 2019.
[24] William S Lovejoy. Some monotonicity results for partially observed markov decision processes. *Operations Research*, 35(5):736–743, 1987.
[25] S Christian Albright. Structural results for partially observable markov decision processes. *Operations Research*, 27(5):1041–1053, 1979.
[26] Amine Laourine and Lang Tong. Betting on gilbert-elliot channels. *IEEE Transactions on Wireless communications*, 9(2):723–733, 2010.
[27] Keqin Liu and Qing Zhao. Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access. *IEEE Transactions on Information Theory*, 56(11):5547–5567, 2010.
[28] Mehdi Salehi Heydar Abad, Ozgur Ercetin, and Deniz Gündüz. Channel sensing and communication over a time-correlated channel with an energy harvesting transmitter. *IEEE Transactions on Green Communications and Networking*, 2(1):114–126, 2017.
[29] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.
[30] Yiyang Pei, Ying-Chang Liang, Kah Chan Teh, and Kwok Hung Li. Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order. *IEEE Journal on Selected Areas in Communications*, 29(8):1648–1659, 2011.
[31] Richard D Smallwood and Edward J Sondik. The optimal control of partially observable markov processes over a finite horizon. *Operations research*, 21(5):1071–1088, 1973.
[32] Yoshikazu Sawaragi and Tsuneo Yoshikawa. Discrete-time markovian decision processes with incomplete state observation. *The Annals of Mathematical Statistics*, 41(1):78–86, 1970.
[33] Dejan V Djonin and Vikram Krishnamurthy. Mimo transmission control in fading channels- a constrained markov decision process formulation with monotone randomized policies. *IEEE Transactions on Signal processing*, 55(10):5069–5083, 2007.
[34] Eitan Altman. *Constrained Markov decision processes*, volume 7. CRC Press, 1999.
[35] Linn I Sennott. Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs. *Operations Research*, 37(4):626–633, 1989.
[36] Linn I Sennott. *Stochastic dynamic programming and the control of queueing systems*, volume 504. John Wiley & Sons, 2009.
[37] Linn I Sennott. Constrained average cost markov decision chains. *Probability in the Engineering and Informational Sciences*, 7(1):69–83, 1993.
[38] Yu-Pin Hsu, Eytan Modiano, and Lingjie Duan. Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals: The no-buffer case. *arXiv preprint arXiv:1712.07419*, 2017.
[39] Linn I Sennott. On computing average cost optimal policies with application to routing to parallel queues. *Mathematical methods of operations research*, 45(1):45–62, 1997.



Guidan Yao received the B.E. and M.E. degrees in electrical engineering from Tianjin University, Tianjin, China, in 2013 and 2016 respectively, She is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering from The Ohio State University, OH, USA. Her research interests include wireless communication, resource allocation, information freshness, Markov decision process, and scheduling algorithms. She received China national scholarship for being one of the top 0.2% undergraduate students in China, for the year 2011 and 2012.

Ahmed M. Bedewy received the B.S. and M.S. degrees in electrical and electronics engineering from Alexandria University, Alexandria, Egypt, in 2011 and 2015, respectively, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, OH, USA, in 2021. His research interests include wireless communication, cognitive radios, resource allocation, communication networks, information freshness, optimization, and scheduling algorithms. He received the Awarded Certificate of Merit and First Class Honors for being one of the top ten undergraduate students, for the period 20062008, and the First, for the period 20082011, in electrical and electronics engineering. His article received the runner-up for the Best Paper Award of ACM MobiHoc 2020.

Ness B. Shroff (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Columbia University in 1994. Thereafter, he joined Purdue University as an Assistant Professor with the School of Electrical and Computer Engineering. At Purdue, he became a Full Professor of ECE and the Director of the University-Wide Center on wireless systems and applications in 2004. In 2007, he joined The Ohio State University, where he currently holds the Ohio Eminent Scholar Endowed Chair in networking and communications with the Department of Electronics and Communication Engineering and the Department of Computer Science and Engineering. He also holds or has held positions as a Visiting (chaired) Professor with Tsinghua University, Beijing, China, Shanghai Jiao Tong University, Shanghai, China, and the IIT Bombay, Mumbai, India. He has received numerous best paper awards for his research and is listed in the Thomson Reuters on The Worlds Most Influential Scientific Minds, and is noted as a Highly Cited Researcher by Thomson Reuters. He also received the IEEE INFOCOM Achievement Award for seminal contributions to scheduling and resource allocation in wireless networks. He serves as the Steering Committee Chair for ACM Mobihoc and the Editor-in-Chief of the IEEE/ACM TRANSACTIONS ON NETWORKING.

APPENDIX A
PROOF OF LEMMA 1

Without loss of generality, we extend space of belief state to $[0,1]$ and show that (a)-(d) hold for $\omega \in [0,1]$. By Proposition 1, $V_n^\beta(\mathbf{s}) \to V^\beta(\mathbf{s})$ as $n \to \infty$. Thus, we show that $V_n^\beta(\mathbf{s})$ satisfies (a)-(d) for $n \geq 0$ via induction. Note that $V_0^\beta(\mathbf{s}) = 0$ satisfies (a)-(d).

Suppose that (a)-(d) hold for n. We (1) show that (d) holds for $n+1$ based on the assumption that (a)-(c) hold for n, and (2) show that (a)-(c) holds for $n+1$ based on the result that (d) hold for $n+1$ shown in step (1) and the assumption that (a)-(c) hold for n.

Step (1): We show that (d) holds for $n+1$. Recall that $V_{n+1}^\beta(\mathbf{s}) = \min\{Q_{n+1}^\beta(\mathbf{s};1), Q_{n+1}^\beta(\mathbf{s};0)\}$. Thus, we can obtain the threshold property by examining the Q functions $Q_{n+1}^\beta(\mathbf{s};0)$ and $Q_{n+1}^\beta(\mathbf{s};1)$ given in (23) and (24). By the expression in (24), $Q_{n+1}^\beta(\Delta,k,\omega;1)$ is linear in ω. Besides, the value function $V_n^\beta(\Delta,k,\omega;1)$ in our case is a piecewise linear and concave function with respect to the belief state for all n, which can be shown via induction similar to [31]. Thus, $Q_{n+1}^\beta(\mathbf{s};0)$ is concave by (24). Moreover, by definition, we have $Q_{n+1}^\beta(\Delta,k,0;1) \geq Q_{n+1}^\beta(\Delta,k,0;0)$. Based on the relation between values of $Q_{n+1}^\beta(\Delta,k,1;1)$ and $Q_{n+1}^\beta(\Delta,k,1;0)$, there are two possible cases for curves of $Q_{n+1}^\beta(\Delta,k,\omega;1)$ and $Q_{n+1}^\beta(\Delta,k,\omega;0)$ as shown in Fig. 12.

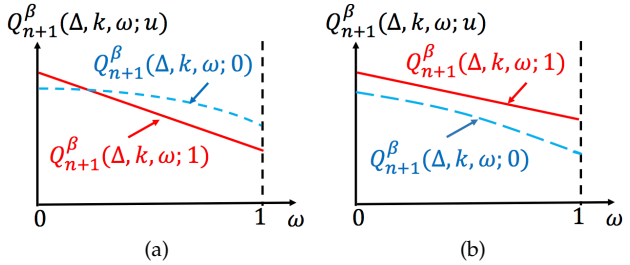


Fig. 12: Values of $Q_{n+1}^\beta(\Delta,k,\omega;u)$

Case 1: $Q_{n+1}^\beta(\Delta,k,1;1) < Q_{n+1}^\beta(\Delta,k,1;0)$ as in Fig. 12a. Due to the concavity of $Q_{n+1}^\beta(\Delta,k,\omega;0)$ and linearity of $Q_{n+1}^\beta(\Delta,k,\omega;1)$ in ω, there must be one unique intersection (corresponds to threshold).

Case 2: $Q_{n+1}^\beta(\Delta,k,1;1) \geq Q_{n+1}^\beta(\Delta,k,1;0)$ (see Fig. 12b):. In the case, we will show that it is always optimal to suspend for any ω given Δ and k, i.e. $Q_{n+1}^\beta(\Delta,k,\omega;1) \geq Q_{n+1}^\beta(\Delta,k,\omega;0)$ for every ω. In particular, by $Q_{n+1}^\beta(\Delta,k,1;1) \geq Q_{n+1}^\beta(\Delta,k,1;0)$, and definitions (23) and (24), we have $\lambda + \beta V_n^\beta(k,(k)_+,p_{11}) - \beta V_n^\beta(\Delta+1,(k)_+,p_{11}) \geq 0$. Moreover, by induction hypothesis, (c) holds for n. Thus, we have

$$\begin{aligned}
&Q_{n+1}^\beta(\Delta,k,\omega;1) - Q_{n+1}^\beta(\Delta,k,\omega;0)\\
=&\omega\Big(\lambda + \beta V_n^\beta(k,(k)_+,p_{11}) - \beta V_n^\beta(\Delta+1,(k)_+,p_{11})\Big)\\
&+\beta\Big(\omega V_n^\beta(\Delta+1,(k)_+,p_{11}) - V_n^\beta(\Delta+1,(k)_+,\mathcal{T}(\omega))\\
&(1-\omega)V_n^\beta(\Delta+1,(k)_+,p_{01})\Big) + (1-\omega)\lambda \qquad (39)\\
\geq&0 \qquad (40)
\end{aligned}$$

Step (2): We show that (a)-(c) hold for $n+1$. First, we consider property (a). It suffices to show that if $\Delta' > \Delta$, then $V_{n+1}^\beta(\Delta',k,\omega) \geq V_{n+1}^\beta(\Delta,k,\omega)$. Since $V_{n+1}^\beta(\mathbf{s}) = \min\{Q_{n+1}^\beta(\mathbf{s};1), Q_{n+1}^\beta(\mathbf{s};0)\}$, we only need to show that for any u that applies to state (Δ',k,ω), there exists an action u' such that $Q_{n+1}^\beta(\Delta',k,\omega;u) \geq Q_{n+1}^\beta(\Delta,k,\omega;u')$

If $u=0$, then we have

$$\begin{aligned}
&Q_{n+1}^\beta(\Delta',k,\omega;0)\\
=&\Delta' + \beta V_n^\beta(\Delta'+1,(k)_+,\mathcal{T}(\omega)) \qquad (41)\\
\geq&\Delta + \beta V_n^\beta(\Delta+1,(k)_+,\mathcal{T}(\omega)) \qquad (42)\\
=&Q_{n+1}^\beta(\Delta,k,\omega;0) \qquad (43)
\end{aligned}$$

The inequality (42) holds since property (a) holds for n by induction hypothesis.

If $u=1$, according to values of Δ, we have two cases to consider specified as follows. If $\Delta < K$, then $\Delta = k-1$ and it implies that the receiver has received the latest status update generated at the beginning of the frame. In the case, the action chosen for state (Δ,k,ω) is to suspend. Recall that $\Delta' = mK + k - 1$ at the k-th slot of certain frame, where $m > 0$. For the case, we have

$$\begin{aligned}
&Q_{n+1}^\beta(\Delta+mK,k,\omega;1)\\
=&\Delta + mK + \lambda + \beta\Big(\omega V_n^\beta(k,(k)_+,p_{11})\\
&+(1-\omega)V_n^\beta(\Delta+K+1,(k)_+,p_{01})\Big) \qquad (44)\\
\geq&\Delta + \lambda + \beta\Big(\omega V_n^\beta(k,(k)_+,p_{11})\\
&+(1-\omega)V_n^\beta(k,(k)_+,p_{01})\Big) \qquad (45)\\
\geq&\Delta + \beta V_n^\beta(k,(k)_+,\mathcal{T}(\omega)) \qquad (46)\\
=&Q_{n+1}^\beta(\Delta,k,\omega;0) \qquad (47)
\end{aligned}$$

The inequality (45) holds since property (a) holds for n by induction hypothesis. The inequality (46) holds since property (c) holds for n by induction hypothesis.

If $\Delta \geq K$, then we have

$$\begin{aligned}
&Q_{n+1}^\beta(\Delta',k,\omega;1)\\
=&\Delta' + \lambda + \beta\Big(\omega V_n^\beta(k,(k)_+,p_{11})\\
&+(1-\omega)V_n^\beta(\Delta'+1,(k)_+,p_{01})\Big) \qquad (48)\\
\geq&\Delta + \lambda + \beta\Big(\omega V_n^\beta(k,(k)_+,p_{11})\\
&+(1-\omega)V_n^\beta(\Delta+1,(k)_+,p_{01})\Big) \qquad (49)\\
=&Q_{n+1}^\beta(\Delta,k,\omega;1) \qquad (50)
\end{aligned}$$

The inequality (49) holds since property (a) holds for n by induction hypothesis.

Second, we consider property (b). It suffices to show that if $\omega' \leq \omega$, then $V_{n+1}^\beta(\Delta,t,\omega') \geq V_{n+1}^\beta(\Delta,t,\omega)$ given V_n^β has properties (a)-(c). The general idea to show this is same to that in proving property (a).

Since $p_{11} \geq p_{01}$, $\mathcal{T}(\omega) = (p_{11} - p_{01})\omega + p_{01}$ is non-decreasing in ω and $\mathcal{T}(\omega') \leq \mathcal{T}(\omega)$. Then, we have

$$\begin{aligned} & Q^{\beta}_{n+1}(\Delta, k, \omega'; 0) \\ =& \Delta + \beta V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(\omega')) && (51) \\ \geq& \Delta + \beta V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(\omega)) && (52) \\ =& Q^{\beta}_{n+1}(\Delta, k, \omega; 0) && (53) \end{aligned}$$

The inequality (52) holds since property (b) holds for n by induction hypothesis.

Recall that for the $(k)_{+}$-th slot of certain frame, the smallest age is k. Then, $V^{\beta}_{n}(k, (k)_{+}, p_{11}) - V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01}) \leq 0$ since properties (a) and (b) in Lemma 1 hold for n by induction hypothesis. Hence, we have

$$\begin{aligned} & Q^{\beta}_{n+1}(\Delta, k, \omega'; u=1) \\ =& \Delta + \lambda + \beta\Big(V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01}) \\ & + \omega'\left(V^{\beta}_{n}(k, (k)_{+}, p_{11}) - V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\right)\Big) && (54) \\ \geq& \Delta + \lambda + \beta\Big(V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01}) \\ & + \omega(V^{\beta}_{n}(k, (k)_{+}, p_{11}) - V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01}))\Big) && (55) \\ =& Q^{\beta}_{n+1}(\Delta, k, \omega; u=1) && (56) \end{aligned}$$

The inequality (55) holds since $\omega' \leq \omega$ and $V^{\beta}_{n}(k, (k)_{+}, p_{11}) - V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01}) \leq 0$.

Finally, we consider property (c). Note that $x \geq y$ and $z = \omega x + (1-\omega)y$. For the left-hand-side of Eq. (25), there are three possible combinations of actions for state (Δ, k, x) and (Δ, k, y), i.e. suspending for both states, transmitting for both states and suspending for latter state but transmitting for former state. Note that $x \geq y$ implies that if the optimal action for state (Δ, k, y) is to update, then the optimal action for state (Δ, k, x) is also to update since the optimal policy for $n+1$-th iteration is of threshold type.

For the case of suspending for both states, we have

$$\begin{aligned} & (1-\omega)\lambda + \omega Q^{\beta}_{n+1}(\Delta, k, x; 0) + (1-\omega) Q^{\beta}_{n+1}(\Delta, k, y; 0) \\ =& (1-\omega)\lambda + \omega\left(\Delta + \beta V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(x))\right) \\ & + (1-\omega)\left(\Delta + \beta V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(y))\right) && (57) \\ \geq& \Delta + \beta V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(z)) && (58) \\ =& Q^{\beta}_{n+1}(\Delta, k, z; 0) && (59) \\ \geq& V^{\beta}_{n+1}(\Delta, k, z) && (60) \end{aligned}$$

The inequality (58) holds since property (c) holds for n by induction hypothesis. The inequality (60) holds by (22).

For the case of transmitting for both states, we have

$$\begin{aligned} & (1-\omega)\lambda + \omega Q^{\beta}_{n+1}(\Delta, k, x; 1) + (1-\omega) Q^{\beta}_{n+1}(\Delta, k, y; 1) \\ =& (1-\omega)\lambda + \Delta + \lambda + \beta\Big(z V^{\beta}_{n}(k, (k)_{+}, p_{11}) \\ & + (1-z) V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\Big) && (61) \\ =& (1-\omega)\lambda + Q^{\beta}_{n+1}(\Delta, k, z; 1) && (62) \\ \geq& Q^{\beta}_{n+1}(\Delta, k, z; 1) && (63) \\ \geq& V^{\beta}_{n+1}(\Delta, k, z) && (64) \end{aligned}$$

The first equality is by (24) plus some basic calculation. The second equality is by (24). The inequality (64) holds by (22).

For the case of transmitting for state (Δ, k, x) but suspending for (Δ, k, y), we have

$$\begin{aligned} & (1-\omega)\lambda + \omega Q^{\beta}_{n+1}(\Delta, k, x; 1) + (1-\omega) Q^{\beta}_{n+1}(\Delta, k, y; 0) \\ =& \lambda + \Delta + \beta\omega\Big(x V^{\beta}_{n}(k, (k)_{+}, p_{11}) \\ & + (1-x) V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\Big) \\ & + \beta(1-\omega) V^{\beta}_{n}(\Delta+1, (k)_{+}, \mathcal{T}(y)) && (65) \\ \geq& \lambda + \Delta + \beta\omega\Big(x V^{\beta}_{n}(k, (k)_{+}, p_{11}) \\ & + (1-x) V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\Big) \\ & + \beta(1-\omega)\Big(y V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{11}) \\ & + (1-y) V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\Big) && (66) \\ \geq& \lambda + \Delta + \beta\Big(z V^{\beta}_{n}(k, (k)_{+}, p_{11}) \\ & + (1-z) V^{\beta}_{n}(\Delta+1, (k)_{+}, p_{01})\Big) && (67) \\ =& Q^{\beta}_{n+1}(\Delta, k, z; 1) && (68) \\ \geq& V^{\beta}_{n+1}(\Delta, k, z) && (69) \end{aligned}$$

The equality (65) is by (23) and (24). The inequality (66) holds since the value function is a piecewise linear and concave function with respect to the belief state, which can be verified with theory developed in [31]. The inequality (67) holds since property (a) holds for n by induction hypothesis with some basic calculation. The inequality (69) holds by (22).

APPENDIX B
PROOF FOR VERIFICATION OF CONDITIONS IN [35]

The conditions are listed below:

- A1: $V^{\beta}(\mathbf{s})$ defined in (18) is finite $\forall \mathbf{s}, \beta$.
- A2: $\exists I \geq 0$ s.t. $-I \leq h^{\beta}(\mathbf{s}) \triangleq V^{\beta}(\mathbf{s}) - V^{\beta}(\mathbf{0})$, $\forall \mathbf{s}, \beta$.
- A3: $\exists M(\mathbf{s}) \geq 0$ s.t. $h^{\beta}(\mathbf{s}) \leq M(\mathbf{s})$, $\forall \mathbf{s}, \beta$. Moreover, for each $\mathbf{s}$, $\exists u(\mathbf{s})$ s.t. $\sum_{\mathbf{s}' \in \mathcal{S}} \mathbb{P}(\mathbf{s}'|\mathbf{s}, u(\mathbf{s})) M(\mathbf{s}') < \infty$.
- A4: $\sum_{\mathbf{s}' \in \mathcal{S}} \mathbb{P}(\mathbf{s}'|\mathbf{s}, u) M(\mathbf{s}') < \infty$ $\forall \mathbf{s}, u$.

In Proposition 1, we showed that a policy f that chooses $u = 0$ at every time slot satisfies $L^{\beta}_{\mathbf{s}}(f; \lambda) < \infty$. By (18), we have $V^{\beta}(\mathbf{s}) \leq L^{\beta}_{\mathbf{s}}(f; \lambda)$, which implies A1. Moreover, we have V^{β} increasing in Δ and decreasing in ω by Lemma 1. Hence, by setting $I = V^{\beta}(\mathbf{0}) - \min_{k \in \mathcal{K}} V^{\beta}((k)_{-}, k, p_{11}) \geq 0$, where $\mathbf{0} = (K, 1, p_{11})$ is the reference state, we proves A2.

Let δ be the policy that transmits at each time slot. Similar to proof of Lemma 6 in [38], The AoI can be regarded as a stable AoI queue. In particular, average arrival rate is one since age increases by 1 at each time slot, and average service rate is infinite since the channel is in a good state with positive probability and can serve infinite number of age packets when it is in a good state. In the case, the age queue is stable. Hence, states that occur after delivery are recurrent. This implies that $\mathbf{0}$ is recurrent. Actually, the probability of not entering state $\mathbf{0}$ after l frames is no more than b^{lK}, where b is steady state probability that channel is in a bad state. Hence, under policy δ the expected cost of the

first passage from state $\mathbf{s}$ to $\mathbf{0}$, denoted by $c_{\mathbf{s},\mathbf{0}}(\delta)$, is finite. Let δ' be a mix policy where δ is used until entering state $\mathbf{0}$ and the discounted Lagrange cost optimal policy δ_β is used afterwards. Suppose T is the first time slot when system enters $\mathbf{0}$. Then, we have

$$\begin{aligned}
&V^\beta(\mathbf{s})\\
\leq&\mathbb{E}_{\delta'}[\sum_{t=1}^{T-1}\beta^{t-1}C(\mathbf{s}_t,u_t)|\mathbf{s}]+\mathbb{E}_{\delta'}[\sum_{t=T}^{\infty}\beta^{t-1}C(\mathbf{s}_t,u_t)|\mathbf{0}] &(70)\\
\leq&c_{\mathbf{s},\mathbf{0}}(\delta)+\mathbb{E}_{\delta_\beta}(\beta^{(T-1)})V^\beta(\mathbf{0}) &(71)\\
\leq&c_{\mathbf{s},\mathbf{0}}(\delta)+V^\beta(\mathbf{0}). &(72)
\end{aligned}$$

Hence, by setting $M(\mathbf{0})=0$ and $M(\mathbf{s})=c_{\mathbf{s},\mathbf{0}}(\delta)$ for $\mathbf{s}\neq\mathbf{0}$, we proves A3. After transition from $\mathbf{s}$ under any action, there will be at most two possible states. Since for all $\mathbf{s}$, $M(\mathbf{s})<\infty$, the sum of at most two $M(\cdot)$ is also finite. Hence, A4 holds.

APPENDIX C
PROOF OF THEOREM 2

Let $V^{\beta,N}$ be the optimal β-discounted Lagrangian cost for the approximate MDP with bound N and $h^{\beta,N}(\mathbf{s})=V^{\beta,N}(\mathbf{s})-V^{\beta,N}(\mathbf{0})$. By [39], it suffices to verify the following conditions B1-B2.

- B1: $\exists\ I\geq 0$, $M(\cdot)\geq 0$ on $\mathcal{S}$ s.t. $-I\leq h^{\beta,N}(\mathbf{s})\leq M(\mathbf{s})$ for $\mathbf{s}\in\mathcal{S}^N$, where $\beta\in(0,1)$ and $N=K+1,K+2,\cdots$.
- B2: $\limsup_{N\to\infty}\bar{L}^{N\star}(\lambda)\leq\bar{L}^\star(\lambda)$.

Consider policy π that updates at each time slot with equal probability. Let $c_{\mathbf{s},\mathbf{0}}(\pi)$ and $c^N_{\mathbf{s},\mathbf{0}}(\pi)$ be the expected cost of the first passage from a state $\mathbf{s}$ to $\mathbf{0}$ by applying π to original and approximate MDP, respectively. Similar to the proof in Appendix B, we have $I=V^{\beta,N}(\mathbf{0})-\min_{k\in\mathcal{K}}V^{\beta,N}((k)_-,k,p_{11})$, $c_{\mathbf{s},\mathbf{0}}(\pi)<\infty$ and $h^{\beta,N}(\mathbf{s})\leq c^N_{\mathbf{s},\mathbf{0}}(\pi)$. Next, we show that $c^N_{\mathbf{s},\mathbf{0}}(\pi)\leq c_{\mathbf{s},\mathbf{0}}(\pi)$. Then, $M(\mathbf{s})=c_{\mathbf{s},\mathbf{0}}(\pi)$. By the proof of Corollary 4.3 in [39], it suffices to show that

$$\sum_{\mathbf{s}'\in\mathcal{S}^N}P^N_{\mathbf{s}\mathbf{s}'}(u)c_{\mathbf{s}',\mathbf{0}}(\pi)\leq\sum_{\mathbf{s}'\in\mathcal{S}}P_{\mathbf{s}\mathbf{s}'}(u)c_{\mathbf{s}',\mathbf{0}}(\pi) \tag{73}$$

Recall that ν is approximation operation to the state defined in (29). Then, we have

$$\begin{aligned}
&\sum_{\mathbf{s}'\in\mathcal{S}^N}P^N_{\mathbf{s}\mathbf{s}'}(u)c_{\mathbf{s}',\mathbf{0}}(\pi)\\
=&\sum_{\mathbf{s}'\in\mathcal{S}^N}\Big(P_{\mathbf{s}\mathbf{s}'}(u)+\sum_{\mathbf{r}\in\mathcal{S}-\mathcal{S}^N}P_{\mathbf{s}\mathbf{r}}(u)\mathbb{1}_{\{\nu(\mathbf{r})=\mathbf{s}'\}}\Big)c_{\mathbf{s}',\mathbf{0}}(\pi) &(74)\\
\leq&\sum_{\mathbf{s}'\in\mathcal{S}^N}P_{\mathbf{s}\mathbf{s}'}(u)c_{\mathbf{s}',\mathbf{0}}(\pi)+\sum_{\mathbf{r}\in\mathcal{S}-\mathcal{S}^N}P_{\mathbf{s}\mathbf{r}}(u)c_{\mathbf{r},\mathbf{0}}(\pi) &(75)\\
=&\sum_{\mathbf{s}'\in\mathcal{S}}P_{\mathbf{s}\mathbf{s}'}(u)c_{\mathbf{s}',\mathbf{0}}(\pi) &(76)
\end{aligned}$$

The inequality (75) holds since policy π does not depend on states and thus $c_{(\Delta,k,\omega),\mathbf{0}}(\pi)\leq c_{(\Delta',k,\omega'),\mathbf{0}}(\pi)$ for $\Delta\leq\Delta'$.

For B2, we need to show that $V^{\beta,N}(\mathbf{s})\leq V^\beta(\mathbf{s})$ for all N. By this, we will have for all N, $\bar{L}^{N\star}(\lambda)=\lim_{\beta\to1}(1-\beta)V^{\beta,N}(\mathbf{s})\leq\lim_{\beta\to1}(1-\beta)V^\beta(\mathbf{s})=\bar{L}^\star(\lambda)$, which completes our proof. Next, we use induction to prove this inequality $V^{\beta,N}(\mathbf{s})\leq V^\beta(\mathbf{s})$. The inequality holds obviously when $n=0$. Suppose $V^{\beta,N}_n(\mathbf{s})\leq V^\beta_n(\mathbf{s})$, then

$$\begin{aligned}
&V^{\beta,N}_{n+1}(\mathbf{s})\\
=&\min_u\{C(\mathbf{s},u;\lambda)+\beta\sum_{\mathbf{s}'\in\mathcal{S}^N}P^N_{\mathbf{s}\mathbf{s}'}(u)V^{\beta,N}_n(\mathbf{s}')\} &(77)\\
\leq&\min_u\{C(\mathbf{s},u;\lambda)+\beta\sum_{\mathbf{s}'\in\mathcal{S}^N}P^N_{\mathbf{s}\mathbf{s}'}(u)V^{\beta}_n(\mathbf{s}')\} &(78)\\
\leq&\min_u\{C(\mathbf{s},u;\lambda)+\beta\sum_{\mathbf{s}'\in\mathcal{S}}P_{\mathbf{s}\mathbf{s}'}(u)V^{\beta}_n(\mathbf{s}')\} &(79)\\
=&V^\beta_{n+1}(\mathbf{s}) &(80)
\end{aligned}$$

The inequality (78) is due to the induction hypothesis. The inequality (79) is due to $\sum_{\mathbf{s}'\in\mathcal{S}^N}P^N_{\mathbf{s}\mathbf{s}'}(u)V^\beta_n(\mathbf{s}')\leq\sum_{\mathbf{s}'\in\mathcal{S}}P_{\mathbf{s}\mathbf{s}'}(u)V^\beta_n(\mathbf{s}')$, which can be shown similar to (73).

APPENDIX D
PROOF OF LEMMA 2

Let $V^\beta_n(\mathbf{s})$ be the cost-to-go function such that $V^\beta_0(\mathbf{s})=0$ for all $\mathbf{s}\in\mathcal{S}$ and for $n\geq0$,

$$V^\beta_{n+1}(\Delta,k,g)=\min_{u\in\{0,1\}}Q^\beta_{n+1}(\Delta,k,g;u) \tag{81}$$

where

$$\begin{aligned}
Q^\beta_{n+1}(\Delta,k,g;0)=&\Delta+\beta\sum_{g'\in\{0,1\}}p_{gg'}V^\beta_n\left(\Delta+1,(k)_+,g'\right) &(82)\\
Q^\beta_{n+1}(\Delta,k,g;1)=&\Delta+\lambda+\beta\Big(p_{g1}V^\beta_n\left(k,(k)_+,1\right)\\
&+p_{g0}V^\beta_n\left(\Delta+1,(k)_+,0\right)\Big) &(83)
\end{aligned}$$

With similar argument in proof of Proposition 1, we can obtain that $V^\beta_n(\mathbf{s})\to V^\beta(\mathbf{s})$ as $n\to\infty$, for every $\mathbf{s}$, β. Hence, we only need to show that for all n, the function $V^\beta_n(\Delta,k,g)$ is non-decreasing in AoI. Next, we show the result using induction. Note that zero function (i.e., $V^\beta_0(\mathbf{s})=0$) satisfies the property. In other words, for $n=0$, the property holds. Suppose that the property holds for n. It remains to show that the property holds for $n+1$. Suppose $\Delta'>\Delta$, we will show $V^\beta_{n+1}(\Delta',k,g)\geq V^\beta_{n+1}(\Delta,k,g)$. Since $V^\beta_{n+1}(\mathbf{s})=\min\{Q^\beta_{n+1}(\mathbf{s};1),Q^\beta_{n+1}(\mathbf{s};0)\}$, it suffices to show that for each u that applies to state (Δ',k,g), there exists an action u' such that $Q^\beta_{n+1}(\Delta',k,g;u)\geq Q^\beta_{n+1}(\Delta,k,g;u')$.

If $u=0$, then we have

$$\begin{aligned}
&Q^\beta_{n+1}(\Delta',k,g;0)\\
=&\Delta'+\beta\sum_{g'\in\{0,1\}}p_{gg'}V^\beta_n(\Delta'+1,(k)_+,g') &(84)\\
\geq&\Delta+\beta\sum_{g'\in\{0,1\}}p_{gg'}V^\beta_n(\Delta+1,(k)_+,g') &(85)\\
=&Q^\beta_{n+1}(\Delta,k,g;0) &(86)
\end{aligned}$$

The inequality (85) holds by our induction hypothesis.

If $u=1$, then we have two cases to consider based on the values of Δ. At the k-th slot of a time frame, if $\Delta<K$, then $\Delta=k-1$ and it implies that the receiver has received the latest status update generated at the beginning of the frame. In the case, the action is to suspend. Recall that $\Delta'=$

$mK+k-1$ at the k-th slot of certain frame, where $m>0$. For the case, we have

$$\begin{aligned}
&Q_{n+1}^{\beta}(mK+k-1,k,g;1)\\
=&mK+k-1+\lambda+\beta\Big(p_{g1}V_n^{\beta}(k,(k)_+,1)\\
&+p_{g0}V_n^{\beta}(mK+k,(k)_+,0)\Big) &(87)\\
\geq&k-1+\beta\left(p_{g1}V_n^{\beta}(k,(k)_+,1)+p_{g0}V_n^{\beta}(k,(k)_+,0)\right) &(88)\\
=&Q_{n+1}^{\beta}(k-1,k,g;0) &(89)\\
=&Q_{n+1}^{\beta}(\Delta,k,g;0) &(90)
\end{aligned}$$

The inequality (88) holds by induction hypothesis.

If $\Delta\geq K$, then we have

$$\begin{aligned}
&Q_{n+1}^{\beta}(\Delta',k,g;1)\\
=&\Delta'+\lambda+\beta\Big(p_{g1}V_n^{\beta}(k,(k)_+,1)\\
&+p_{g0}V_n^{\beta}(\Delta'+1,(k)_+,0)\Big) &(91)\\
\geq&\Delta+\lambda+\beta\Big(p_{g1}V_n^{\beta}(k,(k)_+,1)\\
&+p_{g0}V_n^{\beta}(\Delta+1,(k)_+,0)\Big) &(92)\\
=&Q_{n+1}^{\beta}(\Delta,k,g;1) &(93)
\end{aligned}$$

The inequality (92) holds by induction hypothesis.

APPENDIX E
PROOF OF LEMMA 3

Without loss of generality, we assume that at state (Δ,k,g) it is optimal to attempt a transmit. That is, $Q^{\beta}(\Delta,k,g;1)\leq Q^{\beta}(\Delta,k,g;0)$. Then, for any $\Delta'>\Delta$,

$$\begin{aligned}
&Q^{\beta}(\Delta',k,g;1)-Q^{\beta}(\Delta',k,g;0)\\
=&\lambda+\beta p_{g1}(V^{\beta}\left(k,(k)_+,1\right)-V^{\beta}\left(\Delta'+1,(k)_+,1\right)) &(94)\\
\leq&\lambda+\beta p_{g1}(V^{\beta}\left(k,(k)_+,1\right)-V^{\beta}\left(\Delta+1,(k)_+,1\right)) &(95)\\
=&Q^{\beta}(\Delta,k,g;1)-Q^{\beta}(\Delta,k,g;0) &(96)\\
\leq&0 &(97)
\end{aligned}$$

The inequality (95) holds since $V^{\beta}\left(\Delta'+1,(k)_+,1\right)\geq V^{\beta}\left(\Delta+1,(k)_+,1\right)$ by Lemma 2. Thus, it is also optimal to transmit at (Δ',k,g). Hence, the unconstrained discounted Lagrange cost optimal policy is of threshold-type in AoI.

Let $\Delta_{\beta}^{\star}(k,g;\lambda)$ denote the threshold associated with k and g. That is, given k and g, it is optimal to transmit when $\Delta\geq\Delta_{\beta}^{\star}(k,g;\lambda)$. Let $\Delta_1=\Delta_{\beta}^{\star}(k,0;\lambda)$, we have

$$\begin{aligned}
&Q^{\beta}(\Delta_1,k,1;1)-Q^{\beta}(\Delta_1,k,1;0)\\
=&\lambda+\beta p_{11}\big(V^{\beta}\left(k,(k)_+,1\right)-V^{\beta}\left(\Delta_1+1,(k)_+,1\right)\big) &(98)\\
\leq&\lambda+\beta p_{01}\big(V^{\beta}\left(k,(k)_+,1\right)-V^{\beta}\left(\Delta_1+1,(k)_+,1\right)\big) &(99)\\
=&Q^{\beta}(\Delta_1,k,0;1)-Q^{\beta}(\Delta_1,k,0;0) &(100)\\
\leq&0. &(101)
\end{aligned}$$

The inequality (99) holds since $p_{01}\leq p_{11}$ by assumption and $V^{\beta}\left(k,(k)_+,1\right)-V^{\beta}\left(\Delta+1,(k)_+,1\right)\leq 0$ by Lemma 2. The inequality (101) holds by optimality.

Thus, we have $\Delta_{\beta}^{\star}(k,0;\lambda)=\Delta_1\geq\Delta_{\beta}^{\star}(k,1;\lambda)$. In other words, the threshold associated with good state is not larger than that associated with bad state.