Optimizing Information Freshness in Constrained IoT Systems: A Token-Based Approach

Erfan Delfani, *Graduate Student Member, IEEE*, and Nikolaos Pappas, *Senior Member, IEEE*.

***Abstract*—In Internet of Things (IoT) status update systems, where information is sampled and subsequently transmitted from a source to a destination node, the imperative necessity lies in maintaining the timeliness of information and updating the system with optimal frequency. Optimizing information freshness in resource-limited status update systems often involves *Constrained Markov Decision Process (CMDP)* problems with *update rate constraints*. Solving CMDP problems, especially with multiple constraints, is a challenging task. To address this, we present a *token-based approach* that transforms CMDP into an unconstrained MDP, simplifying the solution process. To demonstrate the comprehensiveness and effectiveness of the token-based approach, we apply this method to systems with one and two update rate constraints to optimize two distinct metrics: *Age of Incorrect Information (AoII)* and *Age of Information (AoI)*, respectively, and explore the analytical and numerical aspects. Additionally, we introduce an *iterative triangle bisection* method for solving the CMDP problems with two constraints, comparing its results with the token-based MDP approach. The structure of the optimal token-based solution is studied analytically. Our numerical results show that the token-based approach yields superior performance over baseline policies, converging to the optimal policy as the maximum number of tokens increases.**

***Index Terms*—Constrained Markov Decision Process (CMDP), Token-Based MDP, Status Update System, Update Rate, Optimal Policy, AoI, AoII.**

I. Introduction

Prudent control of status updates to provide timely information exchange in resource-limited IoT networks has emerged as an important research direction within semantics-aware goal-oriented communication [2], [3]. In these systems, *update packets* from an information source are generated and transmitted through a communication network to a destination node for subsequent processing and utilization. This process fulfills application-driven goals, including remote actuator control in industrial automation, steering autonomous vehicles in intelligent transportation systems, and monitoring health and environmental status [4], [5]. In such time-sensitive systems, the reliability and accuracy of control depend on the timeliness of information. Consequently, there is a high demand for the *freshness of information* and frequent generation of updates. However, due to resource limitations arising from factors such as energy limitations, channel accessibility, or processing capacity, it is crucial to control and optimize the timing and the amount of generated, transmitted, or processed data within a network. This results in problems addressing the optimization of information freshness under updating rate constraints, aiming to find *optimal policies* that answer the question of *when* to generate or transmit update packets. In the literature of status update systems, the quantification of information freshness is accomplished through timing metrics, including Age of Information (AoI) [6], Age of Incorrect Information (AoII) [7], Version AoI (VAoI) [8], each metric addressing distinct aspects of timing and importance of information. Optimizing these metrics under update rate constraints is often formulated as a Constrained Markov Decision Process (CMDP) [9], where the optimal policy determines the best actions regarding the scheduling of updates within the system.

The optimization of information freshness under constraints has been examined in many studies [7], [10]–[29]. Among these, [10] minimizes average AoI by scheduling updates over an error-prone channel with transmission constraints, studying optimal policies under various feedback mechanisms. [11] addresses minimizing average AoI under transmission power constraints, transforming the CMDP problem into an unconstrained MDP through Lagrangian relaxation. In [12], the AoI in a status update system is minimized under an average energy cost constraint for both sampling and transmission. The stochastic problem is formulated as a CMDP and is transformed into an MDP using the Lagrangian method. [7] introduces AoII metric and solves the CMDP problem with a Lagrangian approach. [14] optimizes AoII in a continuous-time setup by formulating a constrained semi-MDP problem and employing the Lagrangian method. [15] focuses on fresh data collection from power-constrained sensors in industrial IoT networks, optimizing average AoI with scheduling algorithms and Linear Programming (LP)-based optimization. The work [17] investigates the minimization of age in a two-hop relay system, subject to a constraint on the average number of forwarding actions at the relay, where the Lagrangian method transforms the CMDP into an MDP. [22] minimizes AoI in wireless networks with peak power-constrained base stations, using CMDP with strict and relaxed power constraints and proposing a truncated multi-user scheduling policy. [23] minimizes average AoI in resource-constrained IoT networks using a CMDP model and Lagrangian multipliers, offering an asymptotically optimal, low-complexity algorithm. [29] employs a similar Lagrangian approach to optimize the freshness and usefulness of information in a pull-based setup.

The works mentioned above primarily rely on formulating the constrained optimization problem using the CMDP framework and subsequently proceed to solve it through a Lagrangian approach, which is the most common method

The authors are with the Department of Computer and Information Science at Linköping University, Sweden, email: {erfan.delfani, nikolaos.pappas}@liu.se. This work has been supported in part by the Swedish Research Council (VR), ELLIIT, and the European Union (ETHER, 101096526, and ELIXIRION, 101120135). A shorter version has been published in [1].

for obtaining the exact solution, in contrast to other approximate or learning-based approaches (see [30]–[33]) or the online greedy approaches based on the virtual queue method and Lyapunov optimization [34]. However, solving a CMDP problem through the direct primary formulation or the Lagrangian dual approach presents a formidable challenge, particularly for problems with multiple constraints [9]. To address this challenge, we present a *token-based approach* for transforming the CMDP problem into an unconstrained MDP problem, following the same approach in [35]. The resulting MDP can be directly solved using popular iterative standard methods. Specifically, we convert the constrained problem into an unconstrained problem by defining new variables within the system model. These variables are defined in such a way as to ensure the update rate constraints. To elaborate, drawing inspiration from the well-known *token bucket* mechanism, a commonly referenced concept in the literature [35]–[38], we allocate a specific number of tokens to the system in each time slot (or assign one token with a specific probability) for potential updates. When the tokens accumulate sufficiently, the system can decide whether to spend them and update or refrain from updating. This way, the system can optimize update actions by spending tokens at the right time by solving an unconstrained MDP. For single rate constraint problems, this approach resembles the idea of modifying the system model and incorporating an Energy Harvesting (EH) sensor with a random energy arrival for sampling and status updating [39]–[46]. However, as we will demonstrate in subsequent sections by applying it to two rate constrained problems, this token-based approach can be effectively applied to general system models, without being limited to EH scenarios. Our results reveal a key finding that the solution of the proposed token-based unconstrained MDP problem converges to the solution of the primary CMDP problem. The main contributions of this paper are as follows:

- We present a token-based unconstrained MDP for optimizing AoII in a status update system derived from a single-rate CMDP and provide analytical results.
- We formulate a two-rate constrained MDP and its corresponding token-based unconstrained MDP for optimizing AoI in a status update system. Analytical results for the unconstrained problem are presented.
- We introduce an iterative triangle bisection method for solving the CMDP with two constraints.
- We examine and compare the stationary distributions of the optimal Markov chains obtained from both the optimal token-based and the CMDP policies.
- We provide numerical results to compare the performance of the optimal token-based policy across different setups with that of the optimal policy for the primary CMDP, as well as other baseline policies.
- Through numerical analysis, we demonstrate that the optimal token-based policy converges to the optimal policy of the primary CMDP problem as the maximum number of tokens in the system increases.

The remainder of the paper is structured as follows: First, in Section II, we present the status update system setup and the information freshness metrics, which will be utilized in the subsequent sections. Next, in Section III, we formulate and solve the proposed token-based approach for the single-rate constrained problem. Following this, we extend the approach to address a two-rate constrained problem in Section IV. The solutions obtained from the proposed token-based approach will be numerically compared to the optimal solutions for the main problems, as well as to other baseline policies, in Section V. Finally, the paper is concluded in Section VI.

II. General System Model

We consider a status update system depicted in Fig. 1, where update packets/samples originating from an information source are generated and transmitted from a source node (Tx) to a destination node (Rx) within a network, by an update policy. This policy is implemented on the transmitter side and dictates when the update packets are transmitted to the destination nodes, and our objective is to ascertain the optimal policy that maximizes performance by optimizing information freshness within the system while adhering to maximum allowable update rates. In this system model, the update packets are forwarded through a channel, which may be reliable or unreliable[1]. Additionally, we integrate a channel from the receiver to the transmitter, which can be utilized to either transmit acknowledgment feedback (in a *push-based* scenario) or request a new update (in a *pull-based* scenario), as in [29] and [47]. These feedback and request packets are typically small, and we assume the backward channel to be error-free and instantaneous[2].

The proposed token-based approach is versatile and can be applied to optimize various information freshness metrics across diverse system configurations. To demonstrate the comprehensiveness and effectiveness of this approach, we will apply it to two different metrics in two distinct setups. Specifically, we will investigate CMDP and token-based problems to optimize AoII and AoI under one and two update rate constraints, respectively, as discussed in Sections III and IV. In the first setup, we opt to employ push-based updating, while in the second setup, we utilize pull-based status updating.

AoI is a performance metric that quantifies information freshness. It is defined as the time elapsed since the generation time of the last successfully received update at the destination node. It can be mathematically represented as $\Delta^{AoI}(t) = t - u(t)$, where $u(t)$ denotes the generation time of the current update at the receiver. On the other hand, AoII quantifies the time elapsed since the latest instance when the source collected a sample with identical content to the current sample stored at the destination. Let us denote this time instance as $V(t)$. The mathematical representation of AoII is then given by

[1]We have considered a single source in the system model, which is directly applicable to multiple sources when each has a dedicated transmission channel. For multiple sources sharing a channel, reducing and optimizing the update rate of a source generally frees up more time slots for other sources to send their updates. This concept can be used to generalize the model to a multi-source scenario, which we propose as a potential area for future work.

[2]With imperfect feedback, the transmitter-side system agent cannot calculate accurate freshness metrics when feedback fails, leaving the system state not fully observable. Alternative approaches for partially observable states may be suggested, which go beyond the scope of this work.

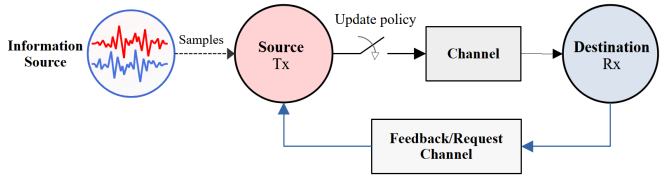


Fig. 1: A general system model for optimizing information freshness under update rate constraints.

$\Delta^{AoII}(t) = (t - V(t)) \times \mathbb{1}\{X_R(t) \neq X_S(t)\}$, where $X_R(t)$ denotes the state (content) of the information at the receiver at time t, $X_S(t)$ denotes the state (content) at the source at time t, and $\mathbb{1}\{\cdot\}$ is the indicator function. In this system model, we assume that the time is slotted and the slots are of equal duration.

III. A System with Single Rate Constraint

A. Problem Formulation

In a status update system, as depicted in Fig. 1, where the objective is to optimize the time average of a freshness metric or a *cost function* $C\big(s\,(t)\,,a\,(t)\,\big)$ subject to an update rate constraint, a CMDP problem can be defined as follows:

$$\min_{\pi\in\Pi}\ \limsup_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1}C\big(s^{\pi}(t),a^{\pi}(t)\big)\Big|s(0)\right], \qquad (1)$$
$$\text{s.t.}\ \limsup_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1}a^{\pi}(t)\Big|s(0)\right]\leq\alpha,$$

where the conditional expectation is taken over the stochastic processes in the system when the policy π is applied, given that the initial state of the system is $s(0)$. These stochastic processes represent the probabilistic model of the system components, including the information source, channel reliability, queries, and other variables that exhibit a stochastic nature. The *policy* π is a sequence of *actions*, i.e., $\pi \overset{\text{def}}{=} \big(a^{\pi}(0), a^{\pi}(1), a^{\pi}(2), \cdots\big)$, where $a^{\pi}(t) \in \{0,1\}$ represents the action at time t. Here, $a^{\pi}(t) = 1$ represents the *update* action, whereas $a^{\pi}(t) = 0$ indicates the *no-update* action. The term *update* refers to the process of generating and transmitting information from the source to the destination. Here, $s^{\pi}(t)$ is the *state* of the system under the policy π, including a timeliness or freshness variable. In this formulation, the parameter $0 < \alpha < 1$ restricts the average update rate. We omit the superscript π in our analysis to facilitate presentation.

To transform this CMDP problem into the token-based MDP one, we define a *new state vector* $\hat{s}(t) \overset{\text{def}}{=} [b(t), s(t)]$, where $b(t) \in \{0, 1, 2, \ldots, b_{\max}\}$ is a token variable evolving as:

$$b(t+1)=\begin{cases}\min\{b(t)-a(t)+1, b_{\max}\} & \text{with probability } \alpha,\\ b(t)-a(t) & \text{with probability } 1-\alpha,\end{cases} \qquad (2)$$

where we have assumed that each update consumes one token. The action $a(t)$ must be set to zero (indicating no update) when $b(t)$ is equal to zero. Otherwise, when $b(t) > 0$, the system will decide on the optimal action at each time slot. This optimal policy can be obtained by solving the following *token-based unconstrained MDP* problem:

$$\min_{\pi\in\Pi}\ \limsup_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1}C\Big(\hat{s}^{\pi}(t),a^{\pi}(t)\Big)\Big|\hat{s}(0)\right], \qquad (3)$$

where the cost function is the same as the CMDP problem (1), i.e., $C\left(\hat{s}^{\pi}(t), a^{\pi}(t)\right) = C\left(s^{\pi}(t), a^{\pi}(t)\right)$, and the transition probabilities of this new MDP problem can be simply derived using the transition probability of the main CMDP problem, via the following equation:

$$\begin{aligned}P\big[\hat{s}(t+1)\big|\hat{s}(t),a(t)\big] &= P\big[b(t+1),s(t+1)\big|b(t),s(t),a(t)\big]\\ &= \underbrace{P\big[b(t+1)\big|b(t),a(t)\big]}_{\text{Can be derived from (2)}} \times \underbrace{P\big[s(t+1)\big|s(t),a(t)\big]}_{\text{Given by CMDP problem (1)}}. \qquad (4)\end{aligned}$$

It is worth mentioning that the expected value of tokens added in each time slot equals the rate constraint, α, ensuring that the policy does not violate the rate constraints, even when all tokens are utilized. This is further elaborated in Section V-C.

B. Optimizing AoII under rate constraint in a push-based setup

In a scenario where the cost function is defined to be AoII, denoted by $\Delta^{\text{AoII}}(t)$, the CMDP problem (1) has previously been addressed for a specific system model [7]. The considered system model focuses on a communication system involving a transmitter-receiver pair where the transmitter sends status updates about a discrete Markov chain process to the receiver over an unreliable channel, as depicted in Fig. 1. The process has N states, and probabilities of staying in the same state (p_R) or transitioning to another state (p_t) are defined such that $p_R + (N-1)p_t = 1$. The unreliable channel follows a Bernoulli distribution with success (p_s) and failure ($p_f = 1 - p_s$) probabilities at each time slot. Successful transmission prompts an ACK feedback, while failure results in a NACK. The transmitter can perfectly estimate the receiver's information source using these packets. The transmitter follows a push-based scenario and generates updates at its discretion by sampling the current process state. In this setup, each update transmission spans a single time slot. The objective is to minimize the time-average AoII with the adopted transmission policy.

This problem can be formulated as a CMDP presented in (1), where the *state* at time t is characterized by the AoII, $s(t) \overset{\text{def}}{=} \Delta^{\text{AoII}}(t) \in \{0, 1, 2, \cdots, \Delta_{\max}\}$. We assume that the AoII is limited to $\Delta_{\max}$ since the stored data with a high AoII becomes too outdated to be useful, making higher values unnecessary (see Section V-A1 for details). The *action* at time t, denoted as $a(t) \in \{0, 1\}$, signifies an attempted update (value 1) or remaining idle (value 0). The *instantaneous cost* is defined as $C\left(s(t), a(t)\right) \overset{\text{def}}{=} \Delta^{\text{AoII}}(t)$, and the *transition probabilities* between states are detailed below:

$$P\left[s(t+1)=0\middle|s(t)=0,a(t)\right]=p_R,\ \forall a(t)\in\{0,1\} \quad (5a)$$
$$P\left[s(t+1)=1\middle|s(t)=0,a(t)\right]=1-p_R,\ \forall a(t)\in\{0,1\} \quad (5b)$$
$$P\left[s(t+1)=0\middle|s(t)\neq 0,a(t)=0\right]=p_t \quad (5c)$$
$$P\left[s(t+1)=s(t)+1\middle|s(t)\neq 0,a(t)=0\right]=1-p_t \quad (5d)$$
$$P\left[s(t+1)=0\middle|s(t)\neq 0,a(t)=1\right]=\beta \quad (5e)$$
$$P\left[s(t+1)=s(t)+1\middle|s(t)\neq 0,a(t)=1\right]=1-\beta, \quad (5f)$$

where $\beta \stackrel{\text{def}}{=} p_R p_s + p_f p_t$. For the main problem presented in [7], the assumption $p_R > p_t$ has been considered to prevent the trivial optimal scenario of *never update*. We consider the same assumption which results in: $\beta = p_R p_s + p_f p_t > p_t p_s + p_f p_t = p_t$.

Transforming this CMDP problem into a token-based MDP problem, the transition probabilities for the new state vector $\hat{s}$, i.e., $P\left[\hat{s}'|\hat{s},a\right] = P\left[b',s'\middle|b,s,a\right]$, can now be calculated using (4), (2) and (5).

- *Case 1.* $a=0$ and $s=0$.

$$P\left[\hat{s}'|\hat{s},a\right]=\begin{cases}\alpha p_R & b'=b+1,s'=0,\\ \alpha\bar{p}_R & b'=b+1,s'=1,\\ \bar{\alpha}p_R & b'=b,s'=0,\\ \bar{\alpha}\bar{p}_R & b'=b,s'=1.\end{cases} \quad (6)$$

- *Case 2.* $a=0$ and $s\neq 0$.

$$P\left[\hat{s}'|\hat{s},a\right]=\begin{cases}\alpha p_t & b'=b+1,s'=0,\\ \alpha\bar{p}_t & b'=b+1,s'=s+1,\\ \bar{\alpha}p_t & b'=b,s'=0,\\ \bar{\alpha}\bar{p}_t & b'=b,s'=s+1.\end{cases} \quad (7)$$

- *Case 3.* $a=1$ and $s=0$, while $b>0$.

$$P\left[\hat{s}'|\hat{s},a\right]=\begin{cases}\alpha p_R & b'=b,s'=0,\\ \alpha\bar{p}_R & b'=b,s'=1,\\ \bar{\alpha}p_R & b'=b-1,s'=0,\\ \bar{\alpha}\bar{p}_R & b'=b-1,s'=1.\end{cases} \quad (8)$$

- *Case 4.* $a=1$ and $s\neq 0$, while $b>0$.

$$P\left[\hat{s}'|\hat{s},a\right]=\begin{cases}\alpha\beta & b'=b,s'=0,\\ \alpha\bar{\beta} & b'=b,s'=s+1,\\ \bar{\alpha}\beta & b'=b-1,s'=0,\\ \bar{\alpha}\bar{\beta} & b'=b-1,s'=s+1.\end{cases} \quad (9)$$

We have defined $\bar{p}_R \stackrel{\text{def}}{=} 1-p_R$, $\bar{p}_t \stackrel{\text{def}}{=} 1-p_t$, $\bar{\alpha} \stackrel{\text{def}}{=} 1-\alpha$, and $\bar{\beta} \stackrel{\text{def}}{=} 1-\beta$. In the following, we present the structural results of the optimal token-based policy, and in Section V-A we compare its outcomes with the main CMDP solution for the same problem.

1) Analytical Results: We present analytical results regarding the existence and structure of the optimal token-based policy. Here, we omit the *hat* on $\hat{s}$ for the sake of simplicity.

Definition 1. *An MDP is weakly accessible (or weakly communicating) if its states can be divided into two subsets, S_t and S_c, where states in S_t are transient under any stationary policy, and for any two states s and s' in S_c, s' can be reached from s under some stationary policy.*

Proposition 1. *The token-based MDP problem* (3) *is weakly accessible.*

Proof. We demonstrate that any state $s' = \left(b', \Delta^{\text{AoII}'}\right) = (b', \Delta') \in \hat{S}$ is reachable from any other state $s = \left(b, \Delta^{\text{AoII}}\right) = (b, \Delta) \in \hat{S}$ under a stationary stochastic policy π, where the action $a \in \{0,1\}$ at each state is randomly selected with a positive probability. The state $b' < b$ is accessible from b with a positive probability (w.p.p.) by executing action $a=1$ for $(b-b')$ time slots, and $b' \geq b$ is reachable from b w.p.p. by realizing action $a=0$ for $(b'-b)$ slots. Upon reaching the state b', irrespective of subsequent actions, the state of the token variable can remain unchanged w.p.p. Consequently, for the remainder of the proof, we consider the token state to be b'. The state $\Delta' < \Delta$ is also attainable from Δ w.p.p. by executing action $a=1$ for one time slot, followed by executing action $a=0$ for Δ' slots. Meanwhile, the state $\Delta' \geq \Delta$ is accessible from Δ by executing action $a=0$ for $\Delta'-\Delta$ slots. □

Proposition 2. *In the token-based MDP problem* (3)*, the optimal average cost J^* achieved by an optimal policy π^* is the same for all initial states, and it satisfies the Bellman equation:*

$$J^* + V(s) = \min_{a\in\{0,1\}} \left\{ \sum_{s'\in\hat{S}} P\left(s'\middle|s,a\right)\left[C(s,a)+V(s')\right] \right\}, \quad (10)$$

$$\pi^*(s) \in \arg\min_{a\in\{0,1\}} \left\{ \sum_{s'\in\hat{S}} P\left(s'\middle|s,a\right)\left[C(s,a)+V(s')\right] \right\}. \quad (11)$$

where $V(s)$ denotes the value function of the MDP problem.

Proof. As per Proposition 1, the problem defined by equation (3) is weakly accessible. Consequently, by Proposition 4.2.3 in [48], the optimal average cost remains consistent across all initial states. Furthermore, based on Proposition 4.2.6 in [48], there exists an optimal policy. According to Proposition 4.2.1 in [48], if we can identify J^* and $V(s)$ satisfying (10), then the optimal policy can be determined using (11). □

The optimal policy, denoted as π^*, relies on $V(s)$, which typically lacks a closed-form solution. Standard and well-defined methods, such as the (Relative) Value Iteration and Policy Iteration algorithms, can be employed to solve this optimization problem. Additionally, it is a common practice to apply Reinforcement Learning (RL) methods to solve MDP problems in scenarios where the system parameters are unknown to the agent and must be learned [49]–[51]. Here, we proceed with models with known parameters, and, in what follows, we present the structure of the value function and the optimal policy of the token-based MDP. This analysis provides insights into the solution and suggests potential avenues for simplifying these algorithms, achieving lighter versions in terms of complexity and scalability.

Definition 2. *Suppose that there exists a $\Delta_T(b) > 0$ for each b such that the action $\pi(b, \Delta^{AoII})$ is 1 for $\Delta^{AoII} \geq \Delta_T(b)$, and 0 otherwise. In this case, the policy π is a threshold policy.*

Lemma 1. *The value function of the token-based MDP problem* (3) *is increasing in* Δ^{AoII}. *In other words, if* $0 < \Delta_1^{AoII} \leq \Delta_2^{AoII}$, *then* $V\left(b, \Delta_1^{AoII}\right) \leq V\left(b, \Delta_2^{AoII}\right)$.

Proof. The proof is presented in Appendix A. □

Theorem 1. *The optimal policy of the token-based MDP problem* (3) *is a threshold policy.*

Proof. The proof can be found in Appendix B. □

IV. A SYSTEM WITH TWO RATE CONSTRAINTS

Consider a scenario where the sensors of an IoT device measure physical variables and transmit updates to a user's monitoring node, such as a smartphone; it is reasonable to assume that higher demand for more frequent updates or fresher information arises when the user actively monitors the variables on their phone. Conversely, updates occur at a regular (or minimum) rate when there is no active monitoring. The system simply imposes two rate constraints in a pull-based scenario: one in response to user requests and another when there is no request. We aim to obtain an optimal update policy that minimizes the AoI in such a system while simultaneously satisfying the two rate constraints.

A. Problem Formulation

Considering the system model in Fig. 1, the optimization of average AoI under two rate constraints can be cast into an infinite-horizon average cost CMDP problem, given by,

$$\begin{aligned}
&\min_{\pi\in\Pi} \ \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} C\Big(s^\pi(t), a^\pi(t)\Big)\Big| s(0)\right], \qquad (12)\\
&\text{s.t.} \ \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} \Big(1-r(t)\Big) a^\pi(t)\Big| s(0)\right] \leq \alpha_0,\\
&\text{s.t.} \ \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} r(t) a^\pi(t)\Big| s(0)\right] \leq \alpha_1,
\end{aligned}$$

where the *state* is defined as $s(t) \stackrel{\text{def}}{=} [\Delta(t), r(t)]^T$, $\Delta(t)$ is the AoI, and $r(t)$ is the random process related to the user's requests, where $r(t) = 1$ when there is an request from Rx, and $r(t) = 0$ when there is no request. We assume that the request arrival process follows an i.i.d. Bernoulli distribution with parameter q over time slots. Same as the problem (1), π is a sequence of *actions* $\pi \stackrel{\text{def}}{=} (a^\pi(0), a^\pi(1), a^\pi(2), \ldots)$, $a^\pi(t) \in \{0, 1\}$, where $a^\pi(t) = 0$ denotes the *remain idle* action, and $a^\pi(t) = 1$ denotes the *update* action at time t. In this system model, the action is adopted at the transmitter (Tx) based on the state vector known at the beginning of each time slot, and the receiver is updated through an error-free channel with a lag normalized to one time slot. The transition probabilities are defined by $P\left[s(t+1)|s(t), a(t)\right] = P\left[\Delta(t+1)|\Delta(t), a(t)\right] \times P\left[r(t+1)\right]$ and calculated by the following equations:

$$\Delta(t+1) = \begin{cases} \min\left\{\Delta(t)+1, \Delta_{\max}\right\} & \text{when } a(t)=0,\\ 1 & \text{when } a(t)=1, \end{cases} \qquad (13)$$

$$r(t+1) = \begin{cases} 1 & \text{with probability (w.p.) } q,\\ 0 & \text{w.p. } 1-q. \end{cases} \qquad (14)$$

The *cost function* is defined to be equal to the AoI, i.e., $C\big(s(t), a(t)\big) \stackrel{\text{def}}{=} \Delta(t)$. In the formulation (12), α_1 and α_0 impose limits on the periods of time during which the system is transmitting updates, while simultaneously either receiving or not receiving a request from the receiver, respectively. These two parameters can also be represented as follows:

$$\begin{aligned}
\alpha_0 &= \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} \Big(1-r(t)\Big)\Big| s(0)\right] \times \alpha_{\min} = (1-q)\alpha_{\min},\\
\alpha_1 &= \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} r(t)\Big| s(0)\right] \times \alpha_{\max} = q\alpha_{\max}. \qquad (15)
\end{aligned}$$

Here, $\alpha_{\max}$ and $\alpha_{\min}$ denote the maximum desired update rates given that there is or is not a request, respectively.

1) Dual Lagrangian Problem: The *primary CMDP* problem (12) can be described by introducing two Lagrange multipliers $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$ and defining Lagrangian function as:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\lambda}, \pi) \qquad (16)\\
&\stackrel{\text{def}}{=} \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} C\big(s^\pi(t), a^\pi(t)\big)\Big| s(0)\right] + \lambda_0 c_0(\pi) + \lambda_1 c_1(\pi)\\
&= \limsup_{T\to\infty} \frac{1}{T} E\Bigg[\sum_{t=0}^{T-1} C\big(s^\pi(t), a^\pi(t)\big) + \lambda_0\big(1-r(t)\big)a^\pi(t)\\
&\qquad + \lambda_1 r(t) a^\pi(t)\Big| s(0)\Bigg] - \lambda_0\alpha_0 - \lambda_1\alpha_1,
\end{aligned}$$

where $\boldsymbol{\lambda} \stackrel{\text{def}}{=} [\lambda_0 \ \lambda_1]^T$ is the vector of Lagrangian multipliers. Here, we have defined,

$$\begin{aligned}
c_0(\pi) &\stackrel{\text{def}}{=} \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} \Big(1-r(t)\Big) a^\pi(t)\Big| s(0)\right] - \alpha_0\\
c_1(\pi) &\stackrel{\text{def}}{=} \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} r(t) a^\pi(t)\Big| s(0)\right] - \alpha_1, \qquad (17)
\end{aligned}$$

where $c_0(\pi) \leq 0$ and $c_1(\pi) \leq 0$ represent the constraints of the CMDP problem (12). The Lagrangian *dual* problem then is given by:

$$\sup_{\boldsymbol{\lambda}\geq 0} \underbrace{\min_{\pi\in\Pi} \mathcal{L}(\boldsymbol{\lambda}, \pi)}_{\stackrel{\text{def}}{=} g(\boldsymbol{\lambda})}, \qquad (18)$$

where $g(\boldsymbol{\lambda}) = g(\lambda_0, \lambda_1) = \mathcal{L}(\boldsymbol{\lambda}, \pi^*_{\boldsymbol{\lambda}})$ represents the Lagrange dual function, where $\pi^*_{\boldsymbol{\lambda}}$ is a $\boldsymbol{\lambda}$-*optimal policy* derived from an unconstrained MDP problem for a given $\boldsymbol{\lambda}$, as follows:

$$\pi^*_{\boldsymbol{\lambda}} \in \arg\min_{\pi\in\Pi} \mathcal{L}(\boldsymbol{\lambda}, \pi). \qquad (19)$$

The new cost function for the dual problem (18), denoted by $C_{\boldsymbol{\lambda}}\big(s(t), a(t)\big)$, can be redefined according to (16):

$$C_{\boldsymbol{\lambda}}\big(s(t), a(t)\big) \stackrel{\text{def}}{=} C\big(s(t), a(t)\big) + \lambda_0\big(1-r(t)\big)a(t) + \lambda_1 r(t) a(t). \qquad (20)$$

The solution to the dual Lagrangian problem (18) generally provides a lower bound for the primary CMDP problem (12). However, given that the state space of the problem $\mathcal{S}$ is a finite

set, the growth condition [9, Eq. 11.21] is satisfied. Additionally, as the cost function $C\big(s(t),a(t)\big) \geq 0$ is bounded from below, the conditions outlined in [9, Corollary 12.2] are met, ensuring the equivalence of the optimal solutions for both the dual and primary problems. Consequently, the optimal solution for the primary CMDP (18) can be obtained by solving:

$$\sup_{\boldsymbol{\lambda}\geq 0} \mathcal{L}(\boldsymbol{\lambda}, \pi^*_{\boldsymbol{\lambda}}), \tag{21}$$

where $\pi^*_{\boldsymbol{\lambda}}$ is derived from (19). More specifically, the optimal policy can be determined in two steps: first, by solving the unconstrained MDP (19) and identifying the $\boldsymbol{\lambda}$-optimal policy $\pi^*_{\boldsymbol{\lambda}}$; second, by obtaining the optimal value of the Lagrangian vector $\boldsymbol{\lambda}$ as per (21). It is important to note that according to (20), the cost of action $a(t)=1$ increases with the increase of $\boldsymbol{\lambda}$ multipliers, while it remains constant for action $a(t)=0$. Thus, the cost function $C_{\boldsymbol{\lambda}}\big(s(t),a(t)\big)$ will increase while the rates will decrease with the increase of $\boldsymbol{\lambda}$. On the other hand, the dual function $g(\boldsymbol{\lambda})$ is decreasing in $\boldsymbol{\lambda}$ [52, Lemma 3.1]. Therefore, the search is for the lowest value of $\boldsymbol{\lambda}$ that results in the maximum allowable rates. In Section IV-C, we present an iterative algorithm to obtain the optimal policy and $\boldsymbol{\lambda}$.

2) Formulation of the Token-based MDP problem: We convert the primary CMDP problem (12) into a token-based MDP problem by including token variables within the state vector, as mentioned earlier. This token-based MDP is characterized by a tuple $< \hat{\mathcal{S}}, \mathcal{A}, P, C >$, where $\hat{\mathcal{S}}$ is the state space, $\mathcal{A}$ is the set of actions, P is the state transition probability function, and C is the cost of MDP.

- States: the new state vector $\hat{s}(t)$ is defined as $\hat{s}(t) \stackrel{\text{def}}{=} [b_0(t), b_1(t), \Delta(t), r(t)]^T \in \hat{\mathcal{S}}$, where $b_0(t)$ and $b_1(t)$ are token variables for two constraints, respectively, taking value in the set $B = \{0,1,2,\ldots,b_{\max}\}$. Moreover, $\Delta(t) \in \{1,2,3,\cdots,\Delta_{\max}\}$ is the AoI at the receiver, and $r(t) \in \{0,1\}$ is the request process at time slot t; $r(t)$ is 1 when there is a request from the destination node and 0 otherwise. The state space, $\hat{\mathcal{S}} = \big\{(b_0,b_1,\Delta,r):\ b_0 \in B, b_1 \in B, \Delta \in \{1,2,\cdots,\Delta_{\max}\}$, and $r \in \{0,1\}\big\}$ is a finite set. The time evolution of these state variables is given by the following equations:

$$b_0(t+1) = \begin{cases} \min\{b_0(t)-a(t)+1, b_{\max}\} & \text{w.p. } \alpha_{\min} \text{ when } r(t)=0, \\ b_0(t)-a(t) & \text{w.p. } 1-\alpha_{\min} \text{ when } r(t)=0, \\ b_0(t) & \text{w.p. 1 when } r(t)=1. \end{cases} \tag{22}$$

$$b_1(t+1) = \begin{cases} \min\{b_1(t)-a(t)+1, b_{\max}\} & \text{w.p. } \alpha_{\max} \text{ when } r(t)=1, \\ b_1(t)-a(t) & \text{w.p. } 1-\alpha_{\max} \text{ when } r(t)=1, \\ b_1(t) & \text{w.p. 1 when } r(t)=0. \end{cases} \tag{23}$$

$$\Delta(t+1) = \begin{cases} \min\{\Delta(t)+1, \Delta_{\max}\} & \text{when } a(t)=0, \\ 1 & \text{when } a(t)=1. \end{cases} \tag{24}$$

$$r(t+1) = \begin{cases} 1 & \text{w.p. } q, \\ 0 & \text{w.p. } 1-q. \end{cases} \tag{25}$$

- Actions: at time t, $a(t)=0$ represents the action of staying idle, while $a(t)=1$ represents the action of transmitting an update. The action $a(t)$ is forced to be 0 when there is no token, i.e., when $b_0(t)=0$ and $r(t)=0$, or $b_1(t)=0$ and $r(t)=1$.
- Transition probabilities: given the following equation,

$$\begin{aligned} P\left[\hat{s}(t+1)|\hat{s}(t),a(t)\right] &= P\left[b_0(t+1)|b_0(t),r(t),a(t)\right] \\ &\times P\left[b_1(t+1)|b_1(t),r(t),a(t)\right] \\ &\times P\left[\Delta(t+1)|\Delta(t),a(t)\right] \times P\left[r(t+1)\right], \end{aligned} \tag{26}$$

  the transition probabilities can simply be written using the equations (22) to (25).
- Cost function: the cost function is identical to the primary CMDP problem: $C\big(\hat{s}(t),a(t)\big) \stackrel{\text{def}}{=} \Delta(t)$.

The resulting token-based unconstrained MDP optimization problem is given by,

$$\min_{\pi\in\Pi} \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} C\Big(\hat{s}^{\pi}(t), a^{\pi}(t)\Big)\Big|\hat{s}(0)\right], \tag{27}$$

where Π is the set of all feasible policies, and $\hat{s}(0)$ is the initial state of the system.

B. Analytical Results

We provide analytical results concerning the existence and structure of the optimal policy for the token-based MDP problem (27). Here, we omit the *hat* on $\hat{s}$ for simplicity.

Proposition 3. *The token-based MDP problem* (27) *is weakly accessible.*

Proof. We establish that any state $s' = (b_0', b_1', r', \Delta') \in \hat{\mathcal{S}}$ is reachable from any other state $s = (b_0, b_1, r, \Delta) \in \hat{\mathcal{S}}$ under a stationary stochastic policy π, where the action $a \in \{0,1\}$ at each state is randomly selected. The token state $b_0' < b_0$ ($b_1' < b_1$) can be accessed from b_0 (b_1) with a positive probability by executing action $a=1$ for $b_0 - b_0'$ ($b_1 - b_1'$) time slots. Conversely, $b_0' \geq b_0$ ($b_1' \geq b_1$) is reachable from b_0 (b_1) with a positive probability (w.p.p.) by implementing action $a=0$ for $b_0' - b_0$ ($b_1' - b_1$) slots. Upon reaching the state b_0' (b_1'), the subsequent actions have no impact, and the token variables' state can remain unchanged with positive probability. Consequently, for the remainder of the proof, we consider the tokens' state as (b_0', b_1'). Additionally, the state $r' \in \{0,1\}$ is reached w.p.p. irrespective of the previous or current system state. Finally, the state $\Delta' < \Delta$ is attainable from Δ w.p.p. by executing action $a=1$ for one time slot, followed by executing action $a=0$ for Δ' slots. Meanwhile, the state $\Delta' \geq \Delta$ is accessible from Δ by executing action $a=0$ for $\Delta' - \Delta$ slots. Consequently, s' is accessible from s, thereby concluding the proof. □

Proposition 4. *In the token-based MDP problem* (27)*, the optimal average cost J^* achieved by an optimal policy π^* is the same for all initial states, satisfying the Bellman equations given by* (10) *and* (11)*.*

Proof. Given that the MDP problem (27) is weakly accessible, the proof is the same as the proof of Proposition 2, which has been omitted for the sake of space. □

Standard methods, like the (Relative) Value Iteration and Policy Iteration algorithms, can be employed to solve this optimization problem. We also present the properties of the value function and the structure of the optimal policy.

Lemma 2. *The value function of the token-based MDP problem* (27) *is increasing in* Δ. *In other words, if* $0 < \Delta_1 \leq \Delta_2$, *then* $V(b_0, b_1, \Delta_1, r) \leq V(b_0, b_1, \Delta_2, r)$.

Proof. The proof is the same as that of Lemma 1, which has been omitted for space. □

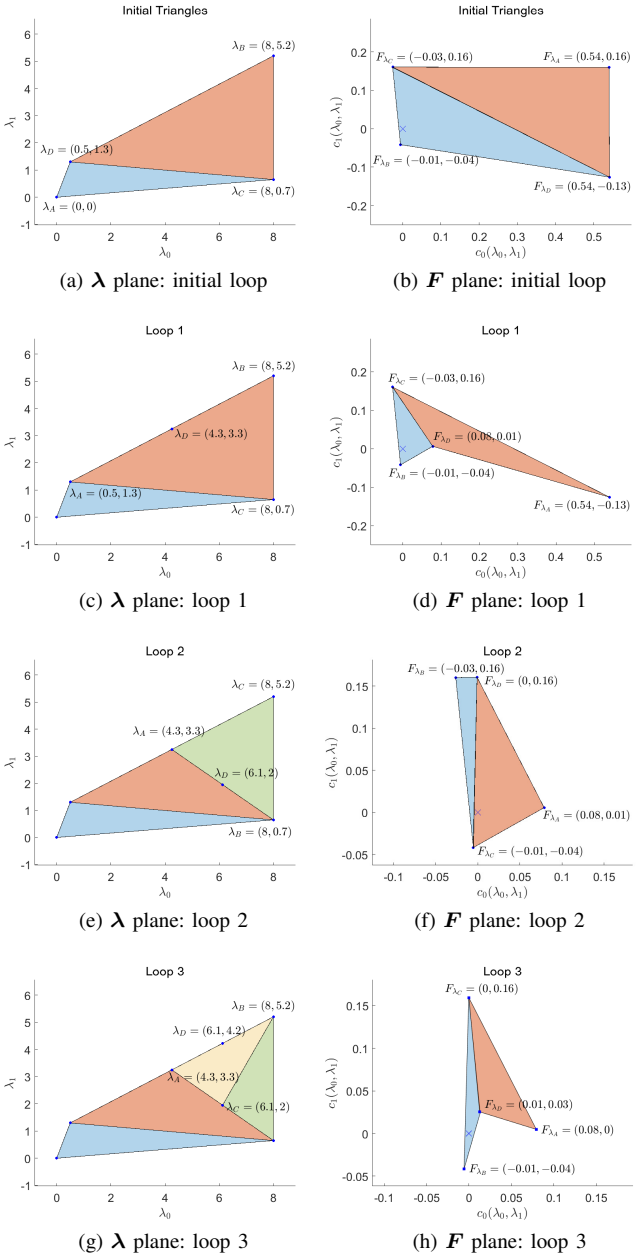
Theorem 2. *The optimal policy* π^* *of the token-based MDP problem* (27) *is a threshold policy. This means that for each combination of* (b_0, b_1, r) *there exists an age threshold* Δ_T *such that* $\pi^*(b_0, b_1, \Delta, r)$ *is* 1 *for* $\Delta \geq \Delta_T$, *and* 0 *otherwise.*

Proof. The proof is the same as the proof of Theorem 1, which has been omitted for the sake of space. □

C. Iterative triangle bisection algorithm for solving the CMDP problem

We present an algorithm to obtain the optimal solution for the problem (12). For a CMDP with a single constraint, the optimal policy can be found by applying an iterative algorithm approach as presented in [10]–[12], [23]. The algorithm optimizes the Lagrange dual problem (18) resulting from the CMDP by iterating through two loops: the inner loop computes the $\boldsymbol{\lambda}$-optimal policy for a given Lagrange multiplier $\boldsymbol{\lambda}$ using the Relative Value Iteration Algorithm (RVIA), and the external loop finds the optimal Lagrange multiplier $\boldsymbol{\lambda}^*$ for a given update policy through a bisection search. However, the aforementioned bisection search is not applicable to cases with two Lagrangian multipliers. To address this issue in the external loop, we provide a generalized version of the bisection method referred to as *iterative triangle bisection*, which can be employed for solving two-dimensional nonlinear equations. This method is a concise version of the algorithm introduced in [53] and outlines an algorithm for solving a system $\boldsymbol{F}_{\boldsymbol{\lambda}} \overset{\text{def}}{=} \begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}}) \\ c_1(\pi^*_{\boldsymbol{\lambda}}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, where $c_0(\cdot)$ and $c_1(\cdot)$ have been defined in Sections IV-A1. In this method, a triangle $(\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C)$ in a two-dimensional plane is iteratively bisected into two triangles ($\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_D\boldsymbol{\lambda}_C$ and $\Delta\boldsymbol{\lambda}_D\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C$), and the search involves identifying the triangle containing a point $\boldsymbol{\lambda}_E \overset{\text{def}}{=} (\lambda_{0E}, \lambda_{1E})$ that satisfies $\boldsymbol{F}_{\boldsymbol{\lambda}_E} = \begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_E}) \\ c_1(\pi^*_{\boldsymbol{\lambda}_E}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The verification that the point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ lies within a triangle $\Delta\boldsymbol{F}_{\boldsymbol{\lambda}_A}\boldsymbol{F}_{\boldsymbol{\lambda}_D}\boldsymbol{F}_{\boldsymbol{\lambda}_C}$ can be accomplished using various methods, such as the *L-test* introduced in [53].

An illustration of this iterative triangle bisection is presented in Fig. 2. The algorithm is initiated with a set of four initial points $\boldsymbol{\lambda}_A$, $\boldsymbol{\lambda}_B$, $\boldsymbol{\lambda}_C$, and $\boldsymbol{\lambda}_D$ such that $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_A}) \geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_A}) \geq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_B}) \leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_B}) \leq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_C}) \geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_C}) \leq 0 \end{bmatrix}$, and $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_D}) \leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_D}) \geq 0 \end{bmatrix}$. These four points constitute two triangles, R ($\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_D\boldsymbol{\lambda}_C$) and S ($\Delta\boldsymbol{\lambda}_D\boldsymbol{\lambda}_B\boldsymbol{\lambda}C$). The $\boldsymbol{\lambda}$-optimal policies are then calculated at the vertices of R and S, followed by the calculation of the corresponding $\boldsymbol{F}_{\boldsymbol{\lambda}}$ for each, as depicted in the $\boldsymbol{F}$-plane on the right-hand side of Fig. 2. The point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ lies within one of these resulting triangles in the $\boldsymbol{F}$-plane, where the algorithm



(a) $\boldsymbol{\lambda}$ plane: initial loop (b) $\boldsymbol{F}$ plane: initial loop

(c) $\boldsymbol{\lambda}$ plane: loop 1 (d) $\boldsymbol{F}$ plane: loop 1

(e) $\boldsymbol{\lambda}$ plane: loop 2 (f) $\boldsymbol{F}$ plane: loop 2

(g) $\boldsymbol{\lambda}$ plane: loop 3 (h) $\boldsymbol{F}$ plane: loop 3

Fig. 2: The first three iterations of the iterative triangle bisection algorithm are shown, with the $\boldsymbol{\lambda}$-plane on the left and the $\boldsymbol{F}$-plane on the right. In each iteration, the triangle in the $\boldsymbol{F}$-plane containing the point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ (marked by a cross) is selected, and its corresponding triangle in the $\boldsymbol{\lambda}$-plane is bisected in the subsequent iteration.

continues with the corresponding triangle in the $\boldsymbol{\lambda}$-plane. This selected triangle, renamed as $\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C$, is bisected into two new triangles, $\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_D\boldsymbol{\lambda}_C$ and $\Delta\boldsymbol{\lambda}_D\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C$, where $\boldsymbol{\lambda}_D$ is the midpoint of the $\boldsymbol{\lambda}_A\boldsymbol{\lambda}_B$ edge. The algorithm proceeds to the next iteration with these two triangles. The proposed method is presented in Algorithm 1. The inner loop of RVIA is also provided in Algorithm 2.

Since the action space is discrete, the optimal policy $\pi^*_{\boldsymbol{\lambda}}$ may not yield the maximum allowable rates but rather achieves approximately equal rates. For justification, and in the case of a CMDP with a single constraint, it is well established that the ultimate optimal policy is a mixture of two non-

randomized stationary policies [52]: one policy is tailored to values exceeding (albeit nearly equal to) the constraint, while the other is designed for values below (yet again, nearly equal to) the maximum allowable rate. Through this approach, upon mixing, the equality constraint can be met. Here, we adopt a generalized approach where a set of four *mixed policies* [9, Section 6.3], denoted by $\pi^*_{\boldsymbol{\lambda}^{++}}$, $\pi^*_{\boldsymbol{\lambda}^{+-}}$, $\pi^*_{\boldsymbol{\lambda}^{-+}}$ and $\pi^*_{\boldsymbol{\lambda}^{--}}$, is employed. These four policies represent the four nearest $\boldsymbol{\lambda}$-optimal policies, achieved through a gradual decrease (increase) of λ_0^* and/or λ_1^* until the conditions of the greater than (smaller than) inequalities are met, such that $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}^{++}})\geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}^{++}})\geq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}^{--}})\leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}^{--}})\leq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}^{+-}})\geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}^{+-}})\leq 0 \end{bmatrix}$, and $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}^{-+}})\leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}^{-+}})\geq 0 \end{bmatrix}$. The gradual decrease (increase) of the λ multipliers can simply be achieved by iteratively dividing (multiplying) them by $1+\gamma$, where γ is chosen to be a small positive real number close to zero, e.g., 0.1. The mixing probability of these policies can be expressed by introducing probabilities ρ_0 and ρ_1 where $\pi^*_{\boldsymbol{\lambda}^{++}}$, $\pi^*_{\boldsymbol{\lambda}^{+-}}$, $\pi^*_{\boldsymbol{\lambda}^{-+}}$ and $\pi^*_{\boldsymbol{\lambda}^{--}}$ are assigned probabilities $\rho_0\rho_1$, $\rho_0(1-\rho_1)$, $(1-\rho_0)\rho_1$, and $(1-\rho_0)(1-\rho_1)$, respectively, satisfying the following system of equations:

$$\begin{aligned}\rho_0\rho_1 c_0(\pi^*_{\boldsymbol{\lambda}^{++}})+\rho_0(1-\rho_1)c_0(\pi^*_{\boldsymbol{\lambda}^{+-}})+(1-\rho_0)\rho_1 c_0(\pi^*_{\boldsymbol{\lambda}^{-+}})\\ +(1-\rho_0)(1-\rho_1)c_0(\pi^*_{\boldsymbol{\lambda}^{--}})=0 \quad (28)\\ \rho_0\rho_1 c_1(\pi^*_{\boldsymbol{\lambda}^{++}})+\rho_0(1-\rho_1)c_1(\pi^*_{\boldsymbol{\lambda}^{+-}})+(1-\rho_0)\rho_1 c_1(\pi^*_{\boldsymbol{\lambda}^{-+}})\\ +(1-\rho_0)(1-\rho_1)c_1(\pi^*_{\boldsymbol{\lambda}^{--}})=0 \quad (29)\end{aligned}$$

D. Complexity of RVIA and iterative triangle bisection

The computational complexity of each iteration of the Value Iteration Algorithm is $\mathcal{O}\left(|\mathcal{S}|^2|\mathcal{A}|\right)$ [54]; therefore, the RVIA has a total computational complexity of $\mathcal{O}\left(N_{\text{RVIA}}|\mathcal{S}|^2|\mathcal{A}|\right)$, where $|\mathcal{S}|$ is the size of the state space, $|\mathcal{A}|$ is the size of the action space, and N_{RVIA} is the number of iterations for termination. The computational complexity of the iterative triangle bisection algorithm is $\mathcal{O}\left(N_{\text{E}}N_{\text{RVIA}}|\mathcal{S}|^2|\mathcal{A}|\right)$, where N_{E} is the iteration number of the external loop. Both N_{E} and N_{RVIA} depend on initial points and termination parameters ε_λ and ε_V, respectively. For the CMDP problem with two rate constraints (12), the iterative triangle bisection algorithm's complexity is $\mathcal{O}\left(N_{\text{E}}N_{\text{RVIA}}|\Delta_{\max}|^2\right)$. For the corresponding token-based MDP, the RVIA complexity is $\mathcal{O}\left(\hat{N}_{\text{RVIA}}|b_{\max}|^4|\Delta_{\max}|^2\right)$.

E. Distribution of the resulting optimal Markov chains

Optimal policies for the token-based MDP and the main CMDP problem can be derived using the RVIA and the iterative triangle bisection algorithm, respectively. By employing the optimal policies derived for these two problems, two corresponding Markov chains (or Markov reward chains, if we consider the costs at each state) result. We can characterize these two Markov chains as follows.

1) Optimal Markov chain from CMDP problem (MC_{CMDP}): This Markov chain can be described by the tuple $< \mathcal{S}, P_{MC} >$, along with an arbitrary initial distribution. Here, $\mathcal{S}$ represents the same state space as defined in Section IV-A, spanning all the state vectors $s = [\Delta, r]^T$, and P_{MC} denotes the transition probability under the optimal action. Specifically, $P_{MC}\left[s'|s\right]$ is determined by $P\left[s'|s, a=\pi^*(s)\right]$, where $P\left[s'|s,a\right]$ is elaborated in Section IV-A.

2) Optimal Markov chain from token-based problem ($MC_{Token-based}$): This Markov chain is characterized by the tuple $< \hat{\mathcal{S}}, \hat{P}_{MC} >$ with an arbitrary initial distribution, where $\hat{\mathcal{S}}$ is the same state space as defined in Section IV-A2, spanning all the state vectors $\hat{s} = [b_0, b_1, \Delta, r]^T$, and $\hat{P}_{MC}$ is the transition probability under the optimal action, i.e., $\hat{P}_{MC}\left[\hat{s}'|\hat{s}\right] = P\left[\hat{s}'|\hat{s}, a=\pi^*(\hat{s})\right]$, where $P\left[\hat{s}'|\hat{s},a\right]$ is defined in Section IV-A2.

The *steady-state* or *stationary distribution* of a Markov chain is a probability distribution over the states that remains unchanged as time progresses. We denote the stationary probability of states s and $\hat{s}$ by $\mu(s)$ and $\hat{\mu}(\hat{s})$, respectively, and the stationary distributions by row vectors $\boldsymbol{\mu} = [\mu(s_1)\ \mu(s_2)\ \cdots\ \mu(s_N)]$ and $\hat{\boldsymbol{\mu}} = \left[\hat{\mu}(\hat{s}_1)\ \hat{\mu}(\hat{s}_2)\ \cdots\ \hat{\mu}(\hat{s}_{\hat{N}})\right]$, respectively. The stationary distribution of the aforementioned Markov chains can be calculated using the following *balance equations*:

$$MC_{CMDP}: \begin{cases} \boldsymbol{\mu}\boldsymbol{P_{MC}} = \boldsymbol{\mu} \\ \sum_{i=1}^{N}\mu(s_i) = 1 \end{cases} \quad (30)$$

$$MC_{Token-based}: \begin{cases} \hat{\boldsymbol{\mu}}\hat{\boldsymbol{P}}_{\boldsymbol{MC}} = \hat{\boldsymbol{\mu}} \\ \sum_{i=1}^{\hat{N}}\hat{\mu}(\hat{s}_i) = 1 \end{cases} \quad (31)$$

where N and $\hat{N}$ are the sizes (number of elements) of the state spaces $\mathcal{S}$ and $\hat{\mathcal{S}}$, respectively. Here, $\boldsymbol{P_{MC}}$ ($\hat{\boldsymbol{P}}_{\boldsymbol{MC}}$) is the transition probability matrix whose $(i,j)^{th}$ element represents the transition probability from state s_i ($\hat{s}_i$) to s_j ($\hat{s}_j$), i.e., $P_{MC}\left[s_j|s_i\right]$ ($\hat{P}_{MC}\left[\hat{s}_j|\hat{s}_i\right]$). After solving the balance equations (30) and (31), we obtain the stationary distributions $\mu(s) = \mu(\Delta, r)$ and $\hat{\mu}(\hat{s}) = \hat{\mu}(b_0, b_1, \Delta, r)$. Additionally, we present the marginal distribution $\hat{\mu}(\Delta, r) = \sum_{b_0=0}^{b_{\max}}\sum_{b_1=0}^{b_{\max}}\hat{\mu}(b_0, b_1, \Delta, r)$, as we will compare this marginal distribution with $\mu(\Delta, r)$ in Section V-B.

V. Numerical Results

In this section, we evaluate the performance of the proposed token-based policy through simulation results. We execute our algorithms on MATLAB over 2×10^4 time slots and average them over 400 runs. All numerical results are obtained using a standard laptop with an Intel(R) Core(TM) i7-1355U 1.7 GHz processor and 32 GB of RAM.

A. Single Rate Constrained Problem

We compare the results of the optimal policies for the primary CMDP problem (1), and the corresponding token-based MDP problem (3). The first policy is derived through the *optimal threshold finder* presented in [7], while the second policy is obtained using the RVIA algorithm. In Figs. 3 and 4 the average AoII for both policies is depicted for various sets of system parameters. The optimal average AoII for various values of α has been illustrated in Fig. 3, with $p_R = 0.5$ and $N = 8$. In this figure, we have modified the value of $b_{\max}$

Algorithm 1 Iterative triangle bisection approach to solve CMDP problem with two constraints

Require: System parameters q, $\alpha_{\min}$, $\alpha_{\max}$, $\Delta_{\max}$, ε_λ and a set of four initial points $\boldsymbol{\lambda}_A$, $\boldsymbol{\lambda}_B$, $\boldsymbol{\lambda}_C$, and $\boldsymbol{\lambda}_D$ such that $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_A})\geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_A})\geq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_B})\leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_B})\leq 0 \end{bmatrix}$, $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_C})\geq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_C})\leq 0 \end{bmatrix}$, and $\begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}_D})\leq 0 \\ c_1(\pi^*_{\boldsymbol{\lambda}_D})\geq 0 \end{bmatrix}$.

Initialize the triangles: $R \leftarrow \Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_D\boldsymbol{\lambda}_C$, $S \leftarrow \Delta\boldsymbol{\lambda}_D\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C$.
Initialize the $\boldsymbol{\lambda}$ vector: $\boldsymbol{\lambda}_E^{\text{new}} \leftarrow \boldsymbol{\lambda}_D$, $\boldsymbol{\lambda}_E^{\text{old}} \leftarrow \begin{bmatrix} \inf \\ \inf \end{bmatrix}$.
while $|\boldsymbol{\lambda}_E^{\text{new}} - \boldsymbol{\lambda}_E^{\text{old}}| \geq \varepsilon_\lambda$ **do** ▷ *External loop: Triangle bisection*
  Run RVIA($\boldsymbol{\lambda}$) at the vertices of R ($\boldsymbol{\lambda}_{RA}$, $\boldsymbol{\lambda}_{RB}$, and $\boldsymbol{\lambda}_{RC}$) to obtain $\boldsymbol{\lambda}$-optimal policies: $\pi^*_{\boldsymbol{\lambda}_{RA}} \leftarrow \text{RVIA}(\boldsymbol{\lambda}_{RA})$, $\pi^*_{\boldsymbol{\lambda}_{RB}} \leftarrow \text{RVIA}(\boldsymbol{\lambda}_{RB})$, and $\pi^*_{\boldsymbol{\lambda}_{RC}} \leftarrow \text{RVIA}(\boldsymbol{\lambda}_{RC})$.
  Calculate $\boldsymbol{F}_{\boldsymbol{\lambda}} = \begin{bmatrix} c_0(\pi^*_{\boldsymbol{\lambda}}) \\ c_1(\pi^*_{\boldsymbol{\lambda}}) \end{bmatrix}$ at the vertices of R to obtain triangle $\Delta\boldsymbol{F}_{\boldsymbol{\lambda}_{RA}}\boldsymbol{F}_{\boldsymbol{\lambda}_{RB}}\boldsymbol{F}_{\boldsymbol{\lambda}_{RC}}$
  if ($\mathbf{0}$ lies within triangle $\Delta\boldsymbol{F}_{\boldsymbol{\lambda}_{RA}}\boldsymbol{F}_{\boldsymbol{\lambda}_{RB}}\boldsymbol{F}_{\boldsymbol{\lambda}_{RC}}$) **then** $T \leftarrow R$
  else $T \leftarrow S$
  end if
  Rotate T such that its first edge ($\boldsymbol{\lambda}_{TA}\boldsymbol{\lambda}_{TB}$) becomes the longest edge of triangle T.
  Update $\Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C \leftarrow T$
  Update $\boldsymbol{\lambda}_E^{\text{old}} \leftarrow \boldsymbol{\lambda}_E^{\text{new}}$, $\boldsymbol{\lambda}_E^{\text{new}} \leftarrow \frac{\boldsymbol{\lambda}_A+\boldsymbol{\lambda}_B+\boldsymbol{\lambda}_C}{3}$, and $\boldsymbol{\lambda}_D \leftarrow \frac{\boldsymbol{\lambda}_A+\boldsymbol{\lambda}_B}{2}$
  Update the triangles: $R \leftarrow \Delta\boldsymbol{\lambda}_A\boldsymbol{\lambda}_D\boldsymbol{\lambda}_C$, $S \leftarrow \Delta\boldsymbol{\lambda}_D\boldsymbol{\lambda}_B\boldsymbol{\lambda}_C$.
end while
$\boldsymbol{\lambda}^* \leftarrow \boldsymbol{\lambda}_E^{\text{new}}$ and $\pi^*_{\boldsymbol{\lambda}} \leftarrow \text{RVIA}(\boldsymbol{\lambda}^*)$
return $\boldsymbol{\lambda}^*$ and $\pi^*_{\boldsymbol{\lambda}}$

Algorithm 2 RVIA($\boldsymbol{\lambda}$) function

Require: State space S, parameters q, ε_V, and the input $\boldsymbol{\lambda}$
Initialize $V_0(s) = v_0(s) = 0,\ \forall s \in S$, and set $t = 0$.
repeat ▷ *Inner loop: RVIA iteration*
  $t \leftarrow t + 1$
  for $s \in S$ **do**
    $v_t(s) = \min_{a\in\{0,1\}} \sum_{s'\in S} P\left(s'|s,a\right)\left[C_{\boldsymbol{\lambda}}(s,a) + V_{t-1}(s')\right]$
    $\pi^*_{\boldsymbol{\lambda}}(s) = \arg\min_{a\in\{0,1\}} \sum_{s'\in S} P\left(s'|s,a\right)\left[C_{\boldsymbol{\lambda}}(s,a) + V_{t-1}(s')\right]$
    $V_t(s) = v_t(s) - v_t(s_0)$ ▷ *s_0: any arbitrary state*
  end for
until $\max_{s\in S}\{V_t(s) - V_{t-1}(s)\} - \min_{s\in S}\{V_t(s) - V_{t-1}(s)\} < \varepsilon_V$
return $\pi^*_{\boldsymbol{\lambda}}$

from 5 to 20. As evident, the optimal AoII for the token-based MDP problem converges to the optimal AoII of the main CMDP problem. In Fig. 4, the optimal average AoII for two policies is depicted as a function of p_R, for $N = 8$ and $\alpha = 0.1$. The optimal average AoII for the token-based policy converges to the optimal average AoII of the main CMDP problem as the value of $b_{\max}$ increases. In both figures, we considered $\Delta_{\max} = 30$.

1) The Impact of Δ_{max}: Our results regarding the optimal policies and optimal average AoII/AoI are obtained under the assumption of a maximum age $\Delta_{\max}$. This approach is a common practice for truncating the age to a maximum value, as age values exceeding a certain high threshold are considered too stale from a systemic perspective. Consequently, these values can be represented or truncated by that high threshold, provided the truncation level is sufficiently high. This approach aids both the theoretical and numerical analysis of the CMDP/MDP problem without compromising the accuracy of the results. For further investigation, we have illustrated the

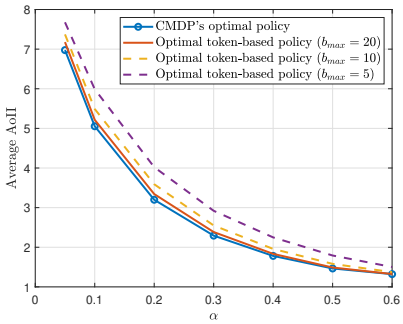


Fig. 3: Optimal average AoII for two policies vs. α in the single rate constrained problem.

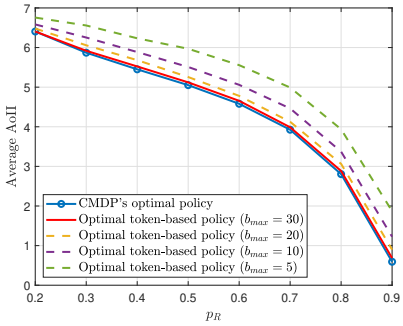


Fig. 4: Optimal average AoII for two policies vs. p_R in the single rate constrained problem.

impact of $\Delta_{\max}$ on the results for the single-rate constrained problem, noting that the two-rate problem exhibits similar behavior. In Fig. 5, the average AoII is depicted against $\Delta_{\max}$ for two values of α. Other system parameters are $N = 8$, $p_R = 0.5$, and $b_{\max} = 10$. It is observed that the average AoII values converge towards their ultimate values as $\Delta_{\max}$ exceeds 20 or 30. For higher values of $\Delta_{\max}$, increases in $\Delta_{\max}$ result in negligible changes. This indicates that $\Delta_{\max} = 30$ yields true age values in our results.

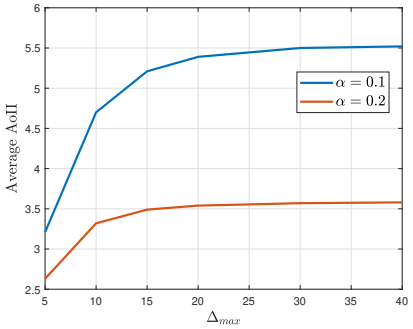


Fig. 5: Average AoII vs. $\Delta_{\max}$.

B. Two Rate Constrained Problem

We compare the optimal token-based policy of the MDP problem (27) with the optimal policy derived from the primary CMDP problem (12), along with two additional baseline policies. We utilize the RVIA algorithm to solve the token-based

MDP problem and the iterative triangle bisection algorithm to solve the CMDP problem. Subsequently, we compare the resulting average AoI for these two policies across various system parameters q (Fig. 6) and $\alpha_{\max}$ (Fig. 7), while keeping $\alpha_{\min}$ fixed at 0.1. In Figs. 6 and 7, we also present the average AoI resulting from a uniform two-rate policy and a random policy. By a uniform two-rate policy, we mean that when $r = 0$, the system is updated at a uniform rate of $\alpha_{\min}$, and when $r = 1$, the system is updated at a uniform rate of $\alpha_{\max}$. Whereas, according to a random policy, the system is updated with probability $\alpha_{\min}$ ($\alpha_{\max}$) when $r = 0$ ($r = 1$). The parameters $\Delta_{\max}$, $\varepsilon_{\boldsymbol{\lambda}}$, and ε_V have been set to 20, 0.1, and 0.1, respectively. The results are summarized as follows:

- With an increase in $b_{\max}$, the optimal average AoI for the token-based MDP problem converges to that of the primary CMDP problem. For further illustration, in Fig. 8, we plot the *optimality gap* between the token-based policy and the optimal policy for the main CMDP problem as a function of $b_{\max}$, where $\alpha_{\max} = 0.5$ and q is either 0.2 or 0.5.
- Even for low levels of $b_{\max}$ (e.g. $b_{\max} = 5$), the performance of the optimal token-based policy is close to the optimal CMDP policy, and it outperforms both the uniform and random policies. In Section IV-D, we observed that when $b_{\max}$ is low, the RVIA's complexity in finding the optimal token-based policy may be preferable to the iterative approach for the primary CMDP. In Fig. 9, the execution time of the token-based RVIA is depicted alongside the execution time of the iterative triangle bisection for the primary CMDP on a logarithmic scale. It can be confirmed that, for low values of $b_{\max}$, the token-based approach demonstrates lower complexity. This is particularly significant because *the token-based approach offers a straightforward formulation and solution without becoming entangled in the Lagrangian formulation and the intricate aspects of the iterative bisection algorithm for the primary CMDP problem.*

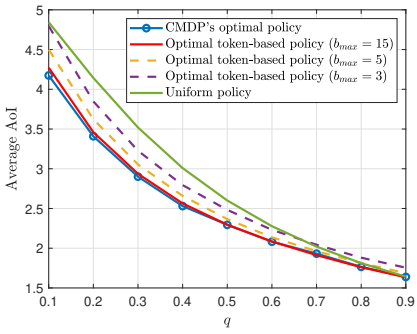


Fig. 6: Average AoI for different policies vs. q ($\alpha_{\max} = 0.5$) in the two rate constrained problem.

To reinforce the convergence of the token-based solution to the main CMDP solution, we present the stationary distributions of the resulting optimal Markov chains introduced in Section IV-E. We compare the stationary probabilities of being in state (Δ, r) for both Markov chains, denoted as $\mu(\Delta, r)$ and $\hat{\mu}(\Delta, r)$, for $\Delta \in \{1, 3, 5, 7\}$ and $r \in \{0, 1\}$.

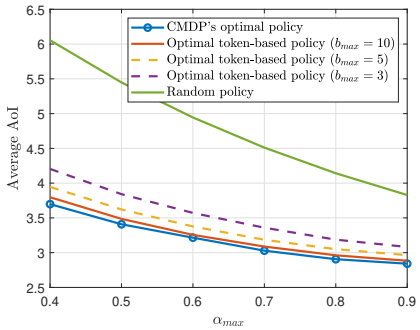


Fig. 7: Average AoI for different policies vs. $\alpha_{\max}$ ($q = 0.2$) in the two rate constrained problem.

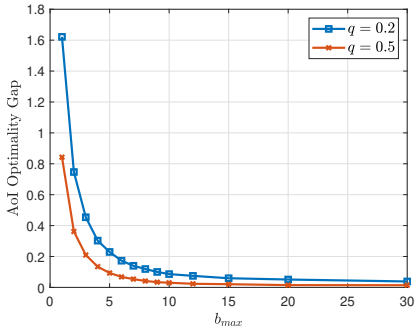


Fig. 8: Optimality gap vs. $b_{\max}$ in the two rate constrained problem.

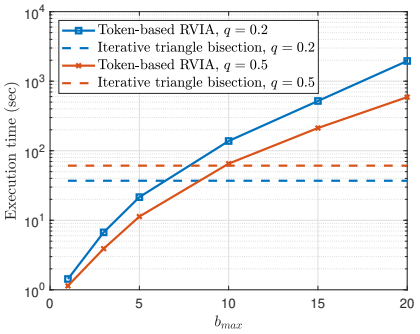


Fig. 9: The execution time vs. $b_{\max}$ for the token-based approach and the iterative triangle bisection method for the primary CMDP.

The results are depicted in Fig. 10 as a function of $b_{\max}$, with $\alpha_{\min} = 0.1$, $\alpha_{\max} = 0.5$, $q = 0.2$, and $\Delta_{\max} = 20$. It can be observed that for different combinations of (Δ, r) values, as we increase $b_{\max}$, the marginal stationary distribution of the optimal Markov chain resulting from the token-based problem ($\hat{\mu}(\Delta, r)$, shown as a solid red line) converges to that of the main CMDP problem ($\mu(\Delta, r)$, shown as a dashed blue line).

C. Discussion on How the Token-Based Approach Satisfies the Constraints

The first point to note is that the expected value of the added tokens in the system at each time slot is equal to the rate constraint (denoted as α in the single-rate problem, and α_1 and α_2 in the two-rate problem), as specified in equations

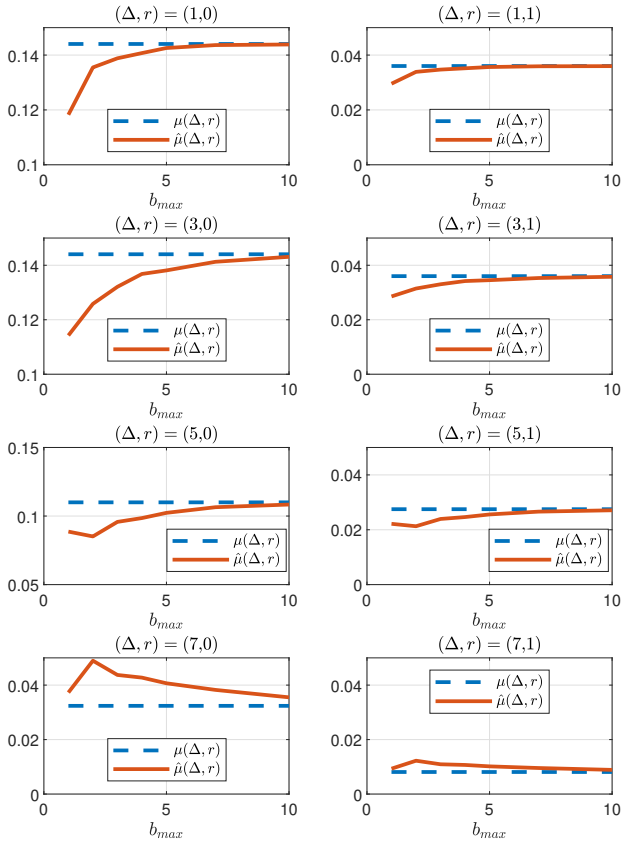


Fig. 10: The stationary distributions of the resulting optimal Markov chains are depicted as a function of $b_{\max}$ for various Δ and r.

(2), (22), and (23). Therefore, even if the policy utilizes all the tokens, the rate constraints will not be violated.

On the other hand, as the maximum number of tokens, denoted as $b_{\max}$, increases in the system, the degrees of freedom in taking actions and selecting the age thresholds (with one threshold for each $b \in \{0, 1, 2, \cdots, b_{\max}\}$) also increase. Consequently, the optimal policy achieves improved optimization results. This increased degree of freedom is illustrated in Figs. 11 and 12, which show the optimal actions for the single rate constrained problem in Section V-A with two different $b_{\max}$ values. In fact, for a given threshold policy (with threshold $\Delta_T(b)$ for each b), the average update rate can be characterized by analyzing the resulting Markov chain and is expressed as $\sum_{b=1}^{b_{\max}} Prob\left[\Delta \geq \Delta_T(b)\right]$. Thus, as $b_{\max}$ increases, the number of optimal $\Delta_T(b)$ values that can be selected to adjust the update rate in the system (as a function of age and b) also increases.

The impact of these optimized results is reflected in the ultimate average update rate used in the system according to the optimal policy. In Fig. 13, we illustrate the update rate versus $b_{\max}$ for different α values in the single-rate problem. It can be observed that as $b_{\max}$ increases, the update rate approaches the equality constraint. It is intuitively reasonable because sending the maximum possible number of updates in the system will enhance the freshness metric, as this approach provides the freshest possible updates within the system. However, it

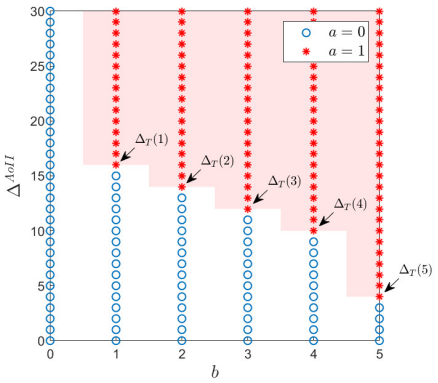


Fig. 11: Optimal actions for $b_{\max} = 5$ ($\alpha = 0.1$).

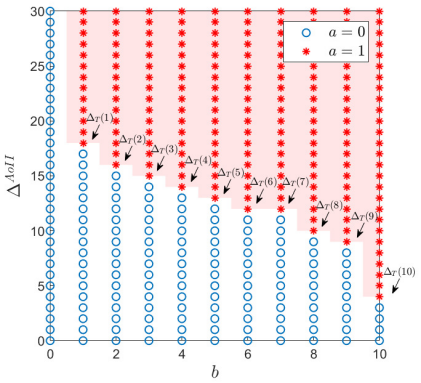


Fig. 12: Optimal actions for $b_{\max} = 10$ ($\alpha = 0.1$).

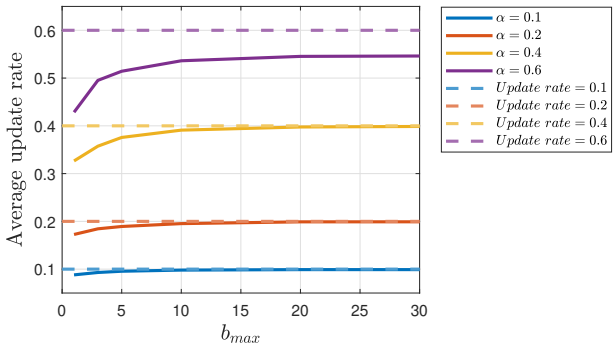


Fig. 13: Average update rate vs. $b_{\max}$.

should be noted that sending the maximum possible number of updates (or, equivalently, using the maximum average update rate) is not always necessary. In some cases, sending additional updates results in uninformative data that does not improve the average AoII or other semantics-aware metrics used as the cost/objective function. This is demonstrated in the same setup with $\alpha = 0.6$, where a lower average update rate, less than 0.6, yields the optimal outcome. However, it is also noteworthy to mention that, in this case, the increase in $b_{\max}$ also brings the average update rate closer to the optimal value.

VI. Conclusion

In this work, we explored a token-based approach for transforming CMDP problems into unconstrained MDP prob-

lems, explicitly focusing on optimizing information freshness in status update systems. We have applied this approach to systems with one and two update rate constraints and examined the analytical structure of the optimal token-based policy. Additionally, we have introduced an iterative triangle bisection algorithm for solving CMDP problems with two constraints. Our findings demonstrate the superior performance of the optimal policy compared to baseline policies and the convergence of the optimal token-based policy to the optimal policy of the main CMDP problem as the maximum number of tokens in the system increases. As future work, we may consider incorporating more realistic channel models, either in the forward channel or in the feedback channel. Additionally, given that the structure of the optimal policies has been proven to be threshold-based, lightweight approximate or learning-based algorithms can also be explored to further enhance the applicability of the token-based approach in practical implementations.

REFERENCES

[1] E. Delfani and N. Pappas, "Optimizing information freshness in iot systems with update rate constraints: A token-based approach," in *IFIP/IEEE Networking Conference*, 2024.

[2] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, 2021.

[3] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret *et al.*, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, vol. 36, no. 4, 2022.

[4] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, 2021.

[5] N. Pappas, M. A. Abd-Elmagid, B. Zhou, W. Saad, and H. S. Dhillon, *Age of Information: Foundations and Applications*. Cambridge University Press, 2023.

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

[7] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, 2020.

[8] R. D. Yates, "The age of gossip in networks," in *IEEE ISIT*, 2021.

[9] E. Altman, *Constrained Markov decision processes*. Vol. 7, CRC Press, 1999.

[10] E. T. Ceran, D. Gündüz, and A. György, "Average age of information with hybrid arq under a resource constraint," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, 2019.

[11] Q. Wang, H. Chen, Y. Li, Z. Pang, and B. Vucetic, "Minimizing age of information for real-time monitoring in resource-constrained industrial iot networks," in *IEEE 17th INDIN*, vol. 1, 2019.

[12] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 7468–7482, 2019.

[13] M. Ma and V. W. Wong, "Age of information driven cache content update scheduling for dynamic contents in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8427–8441, 2020.

[14] I. Cosandal, N. Akar, and S. Ulukus, "Multi-threshold aoii-optimum sampling policies for ctmc information sources," *arXiv preprint arXiv:2407.08592*, 2024.

[15] H. Tang, J. Wang, L. Song, and J. Song, "Minimizing age of information with power constraints: Multi-user opportunistic scheduling in multi-state time-varying channels," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, 2020.

[16] Q. Wang, H. Chen, Y. Gu, Y. Li, and B. Vucetic, "Minimizing the age of information of cognitive radio-based iot systems under a collision constraint," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8054–8067, 2020.

[17] Y. Gu, Q. Wang, H. Chen, Y. Li, and B. Vucetic, "Optimizing information freshness in two-hop status update systems under a resource constraint," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1380–1392, 2021.

[18] R. Li, Q. Ma, J. Gong, Z. Zhou, and X. Chen, "Age of processing: Age-driven status sampling and processing offloading for edge-computing-enabled real-time iot applications," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14 471–14 484, 2021.

[19] H. Huang, D. Qiao, and M. C. Gursoy, "Age-energy tradeoff optimization for packet delivery in fading channels," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 179–190, 2021.

[20] J. P. Champati, M. Skoglund, M. Jansson, and J. Gross, "Detecting state transitions of a markov source: Sampling frequency and age trade-off," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3081–3095, 2022.

[21] S. Jayanth and R. V. Bhat, "Age of processed information minimization over fading multiple access channels," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 1664–1676, 2022.

[22] G. Chen, Y. Chen, J. Wang, and J. Song, "Minimizing age of information in downlink wireless networks with time-varying channels and peak power constraint," *IEEE Transactions on Vehicular Technology*, 2023.

[23] M. Hatami, M. Leinonen, Z. Chen, N. Pappas, and M. Codreanu, "On-demand aoi minimization in resource-constrained cache-enabled iot networks with energy harvesting sensors," *IEEE Transactions on Communications*, vol. 70, no. 11, 2022.

[24] B. Yu, Y. Cai, X. Diao, and K. Cheng, "Adaptive packet length adjustment for minimizing age of information over fading channels," *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6641–6653, 2023.

[25] B. Yu, Y. Cai, X. Diao, and Y. Chen, "Aoi minimization scheme for short-packet communications in energy-constrained iiot," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp. 20 188–20 200, 2023.

[26] A. Zakeri, M. Moltafet, M. Leinonen, and M. Codreanu, "Minimizing the aoi in resource-constrained multi-source relaying systems: Dynamic and learning-based scheduling," *IEEE Transactions on Wireless Communications*, 2023.

[27] S. S. Vilni, M. Moltafet, M. Leinonen, and M. Codreanu, "Multi-source aoi-constrained resource minimization under harq: Heterogeneous sampling processes," *IEEE Transactions on Vehicular Technology*, 2023.

[28] J. Wang and D. Qiao, "How to minimize the weighted sum aoi in multi-source status update systems: Tdma or noma?" *IEEE Transactions on Vehicular Technology*, 2023.

[29] P. Agheli, N. Pappas, P. Popovski, and M. Kountouris, "Effective communication: When to pull updates?" in *IEEE ICC*, 2024.

[30] C. Tessler, D. J. Mankowitz, and S. Mannor, "Reward constrained policy optimization," in *International Conference on Learning Representations*, 2018.

[31] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained policy optimization," in *International conference on machine learning*, 2017.

[32] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, "Safe exploration in continuous action spaces," *arXiv preprint arXiv:1801.08757*, 2018.

[33] S. Levine and V. Koltun, "Guided policy search," in *International conference on machine learning*, 2013.

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

[35] G. J. Stamatakis, O. Simeone, and N. Pappas, "Optimizing information freshness over a channel that wears out," *57th Asilomar Conference on Signals, Systems, and Computers*, 2023.

[36] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, 2011.

[37] V. Raghunathan, S. Ganeriwal, M. Srivastava, and C. Schurgers, "Energy efficient wireless packet scheduling and fair queuing," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 3, no. 1, 2004.

[38] F. Chiariotti, J. Holm, A. E. Kalør, B. Soret, S. K. Jensen, T. B. Pedersen, and P. Popovski, "Query age of information: Freshness in pull-based communication," *IEEE Transactions on Communications*, vol. 70, no. 3, 2022.

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

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

[41] E. Delfani, G. J. Stamatakis, and N. Pappas, "State-aware timeliness in energy harvesting iot systems monitoring a markovian source," *IEEE Transactions on Green Communications and Networking*, 2024.

[42] Z. Chen, N. Pappas, E. Björnson, and E. G. Larsson, "Optimizing information freshness in a multiple access channel with heterogeneous devices," *IEEE Open Journal of the Communications Society*, vol. 2, 2021.

[43] E. Gindullina, L. Badia, and D. Gündüz, "Age-of-information with information source diversity in an energy harvesting system," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, 2021.

[44] M. Hatami, M. Leinonen, and M. Codreanu, "Aoi minimization in status update control with energy harvesting sensors," *IEEE Transactions on Communications*, vol. 69, no. 12, 2021.

[45] E. Delfani and N. Pappas, "Version age-optimal cached status updates in a gossiping network with energy harvesting sensor," *IEEE Transactions on Communications*, 2024.

[46] S. Dongare, A. Jovovic, W. de Sombre, A. Ortiz, and A. Klein, "Minimizing the age of incorrect information for status update systems with energy harvesting," in *ICC 2024-IEEE International Conference on Communications*, 2024, pp. 5323–5328.

[47] G. Stamatakis, N. Pappas, A. Fragkiadakis, N. Petroulakis, and A. Traganitis, "Semantics-aware active fault detection in status updating systems," *IEEE Open Journal of the Communications Society*, 2024.

[48] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. II*, 3rd ed. Athena Scientific, 2007.

[49] X. Chen, C. Wu, T. Chen, H. Zhang, Z. Liu, Y. Zhang, and M. Bennis, "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Transactions on wireless communications*, vol. 19, no. 4, pp. 2268–2281, 2020.

[50] E. T. Ceran, D. Gündüz, and A. György, "A reinforcement learning approach to age of information in multi-user networks with harq," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1412–1426, 2021.

[51] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in rf-powered communication systems," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4747–4760, 2020.

[52] F. J. Beutler and K. W. Ross, "Optimal policies for controlled markov chains with a constraint," *Journal of mathematical analysis and applications*, vol. 112, no. 1, 1985.

[53] C. Harvey and F. Stenger, "A two-dimensional analogue to the method of bisections for solving nonlinear equations," *Quarterly of Applied Mathematics*, vol. 33, no. 4, 1976.

[54] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the complexity of solving markov decision problems," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 394–402.

[55] J. N. Tsitsiklis, "Np-hardness of checking the unichain condition in average cost mdps," *Operations research letters*, vol. 35, no. 3, pp. 319–323, 2007.

[56] H. C. Tijms, *A first course in stochastic models*. John Wiley and sons, 2003.

Appendix A
Proof of Lemma 1

Proof. We use the Value Iteration Algorithm to prove the lemma. The iteration at time t updates the value function of $s \in S$ as follows:

$$V_{t+1}(s) = \min_{a\in\{0,1\}} \left\{ C(s) + \sum_{s'\in S} P\left[s'|s,a\right] V_t(s') \right\}. \quad (32)$$

VIA converges to the value function of the Bellman equation regardless of the initial value[3] of $V_0(s)$, i.e., $\lim_{t\to\infty} V_t(s) = V(s)\ \forall s \in S$. Therefore, it is sufficient to prove that $\forall\ 0 < \Delta_1 \le \Delta_2$ and $\forall\ t = 0, 1, 2, \cdots$:

$$V_t\left(b, \Delta_1, \right) \le V_t\left(b, \Delta_2\right). \quad (33)$$

Without loss of generality, we assume $V_0\left(b, \Delta\right) = 0,\ \forall (b, \Delta) \in S$. Thus, (33) is true for $t = 0$. Next, with the assumption that (33) holds till $t > 0$, we show it holds for $t + 1$. Let us define x, y, u, and w as follows:

$$x = \Delta_1 + \sum_{s'\in S} P\left[s'|s_1, a = 0\right] V_t(s'), \quad (34)$$

$$y = \Delta_1 + \sum_{s'\in S} P\left[s'|s_1, a = 1\right] V_t(s'), \quad (35)$$

$$u = \Delta_2 + \sum_{s'\in S} P\left[s'|s_2, a = 0\right] V_t(s'), \quad (36)$$

$$w = \Delta_2 + \sum_{s'\in S} P\left[s'|s_2, a = 1\right] V_t(s'), \quad (37)$$

where $s_1 = (b, \Delta_1)$, $s_2 = (b, \Delta_2)$. So we have $V_{t+1}\left(s_1\right) = \min\left\{x, y\right\}$ and $V_{t+1}\left(s_2\right) = \min\left\{u, w\right\}$. Using the transition probabilities, we can show that:

$$\begin{aligned} x = \Delta_1 &+ \alpha p_t V_t\left(b{+}1, 0\right) + \alpha \bar{p}_t V_t\left(b{+}1, \min\{\Delta_1{+}1, \Delta_{\max}\}\right) \\ &+ \bar{\alpha} p_t V_t\left(b, 0\right) + \bar{\alpha}\bar{p}_t V_t\left(b, \min\{\Delta_1{+}1, \Delta_{\max}\}\right), \end{aligned} \quad (38)$$

$$\begin{aligned} u = \Delta_2 &+ \alpha p_t V_t\left(b{+}1, 0\right) + \alpha \bar{p}_t V_t\left(b{+}1, \min\{\Delta_2{+}1, \Delta_{\max}\}\right) \\ &+ \bar{\alpha} p_t V_t\left(b, 0\right) + \bar{\alpha}\bar{p}_t V_t\left(b, \min\{\Delta_2{+}1, \Delta_{\max}\}\right). \end{aligned} \quad (39)$$

Since $\Delta_1 \le \Delta_2$ and $V_t\left(b, \Delta_1\right) \le V_t\left(b, \Delta_2\right)$, each term in x is smaller than (or equal to) the corresponding term in u. Therefore, $x \le u$. Similarly, according to the following equations, $y \le w$.

$$\begin{aligned} y = \Delta_1 &+ \alpha\beta V_t\left(b, 0\right) + \alpha\bar{\beta} V_t\left(b, \min\{\Delta_1{+}1, \Delta_{\max}\}\right) \\ &+ \bar{\alpha}\beta V_t\left(b{-}1, 0\right) + \bar{\alpha}\bar{\beta} V_t\left(b{-}1, \min\{\Delta_1{+}1, \Delta_{\max}\}\right), \end{aligned} \quad (40)$$

$$\begin{aligned} w = \Delta_2 &+ \alpha\beta V_t\left(b, 0\right) + \alpha\bar{\beta} V_t\left(b, \min\{\Delta_2{+}1, \Delta_{\max}\}\right) \\ &+ \bar{\alpha}\beta V_t\left(b{-}1, 0\right) + \bar{\alpha}\bar{\beta} V_t\left(b{-}1, \min\{\Delta_2{+}1, \Delta_{\max}\}\right). \end{aligned} \quad (41)$$

If $x \le u$ and $y \le w$, then $\min\left\{x, y\right\} \le \min\left\{u, w\right\}$; therefore, $V_{t+1}\left(s_1\right) \le V_{t+1}\left(s_2\right)$ and the proof is complete. □

Appendix B
Proof of Theorem 1

Proof. The Bellman equation at state $s = (b, \Delta)$ can be simplified as follows:

$$\begin{aligned} J^* + V(s) &= \min_{a\in\{0,1\}} \left\{ \Delta + \sum_{s'\in S} P\left[s'|s,a\right] V(s') \right\} \\ &= \Delta + \min_{a\in\{0,1\}} \left\{ \sum_{s'\in S} P\left[s'|s,a\right] V(s') \right\}. \end{aligned} \quad (42)$$

Therefore, the optimal action can be obtained by:

$$a^*(s) = \arg\min_{a\in\{0,1\}} \left\{ \sum_{s'\in S} P\left[s'|s,a\right] V(s') \right\} = \begin{cases} 0, & DV(s) \ge 0, \\ 1, & DV(s) < 0, \end{cases} \quad (43)$$

[3] According to [48, Proposition 4.3.1], when weak accessibility holds, value iteration converges to the optimal result in the general case. However, stronger conditions are required to guarantee the convergence of the value iteration algorithm, including the unichain condition and aperiodicity of the transition matrix of the MDP for every policy. While the unichain condition is intuitive, it is not clear how to check it without enumerating all policies, which presents an NP-hard problem [55]. Some weaker conditions have also been discussed in the literature, including the weak unichain assumption, which is practically always satisfied in real-world applications [56].

where we have defined $V^0(s) = \sum_{s' \in S} P\left[s'|s, a=0\right] V(s')$, $V^1(s) = \sum_{s' \in S} P\left[s'|s, a=1\right] V(s')$, and $DV(s) = V^1(s) - V^0(s)$.

As can be observed, the optimal action $a^*(s)$ is related to the sign of $DV(s)$. To demonstrate that the optimal policy is a threshold policy, we consider two cases for the states: $\Delta = 0$ and $0 < \Delta \leq \Delta_{\max}$. When $\Delta = 0$ or $b = 0$, it can be shown that $DV(s) = 0$; therefore, the action $a = 0$ is optimal. In the other cases where $b > 0$ and $0 < \Delta \leq \Delta_{\max}$, we have:

$$\begin{aligned} V^0(s) &= \alpha p_t V(b+1, 0) + \alpha \bar{p}_t V(b+1, \min\{\Delta+1, \Delta_{\max}\}) \\ &\quad + \bar{\alpha} p_t V(b, 0) + \bar{\alpha} \bar{p}_t V(b, \min\{\Delta+1, \Delta_{\max}\}), \quad (44a) \\ V^1(s) &= \alpha\beta V(b, 0) + \alpha\bar{\beta} V(b, \min\{\Delta+1, \Delta_{\max}\}) \quad (44b) \\ &\quad + \bar{\alpha}\beta V(b-1, 0) + \bar{\alpha}\bar{\beta} V(b-1, \min\{\Delta+1, \Delta_{\max}\}), \\ DV(s) &= DV(b, \Delta) = V^1(s) - V^0(s). \quad (44c) \end{aligned}$$

It is obvious that $DV(b, \Delta_{\max}) = DV(b, \Delta_{\max} - 1)$. Therefore, the optimal action at the state $(b, \Delta_{\max})$ will be identical to the optimal action at the state $(b, \Delta_{\max} - 1)$. Thus, we can limit the remainder of the proof to the states where $0 < \Delta < \Delta_{\max}$, allowing us to omit the operator $\min\{\cdot, \Delta_{\max}\}$.

In what follows, we demonstrate that $DV(s) = DV(b, \Delta)$ is a decreasing (non-increasing) function of Δ, i.e., for $\Delta^- \leq \Delta^+$, we show that $DV(b, \Delta^+) \leq DV(b, \Delta^-)$ or $DV(b, \Delta^+) - DV(b, \Delta^-) \leq 0$. This leads to a threshold policy, as a negative value of $DV(b, \Delta)$ for a threshold $\Delta_T(b)$ implies that it will also be negative for $\Delta \geq \Delta_T(b)$, maintaining the optimal action at 1. By simplification of $DV(b, \Delta^+)$ and $DV(b, \Delta^-)$ based on (44) we obtain the following equations:

$$\begin{aligned} &DV(b, \Delta^+) - DV(b, \Delta^-) \\ &= V^1(b, \Delta^+) - V^1(b, \Delta^-) - \big[V^0(b, \Delta^+) - V^0(b, \Delta^-)\big] \\ &= \alpha\bar{\beta}\big[V(b, \Delta^+ + 1) - V(b, \Delta^- + 1)\big] \\ &\quad + \bar{\alpha}\bar{\beta}\big[V(b-1, \Delta^+ + 1) - V(b-1, \Delta^- + 1)\big] \\ &\quad - \alpha\bar{p}_t\big[V(b+1, \Delta^+ + 1) - V(b+1, \Delta^- + 1)\big] \\ &\quad - \bar{\alpha}\bar{p}_t\big[V(b, \Delta^+ + 1) - V(b, \Delta^- + 1)\big] \\ &= \alpha\Big\{\bar{\beta}\big[V(b, \Delta^+ + 1) - V(b, \Delta^- + 1)\big] \\ &\quad - \bar{p}_t\big[V(b+1, \Delta^+ + 1) - V(b+1, \Delta^- + 1)\big]\Big\} \\ &\quad + \bar{\alpha}\Big\{\bar{\beta}\big[V(b-1, \Delta^+ + 1) - V(b-1, \Delta^- + 1)\big] \\ &\quad - \bar{p}_t\big[V(b, \Delta^+ + 1) - V(b, \Delta^- + 1)\big]\Big\}. \quad (45) \end{aligned}$$

According to (45), to confirm the inequality $DV(b, \Delta^+) - DV(b, \Delta^-) \leq 0$, it suffice to demonstrate that $\bar{\beta}\big[V(b-1, \Delta^+) - V(b-1, \Delta^-)\big] - \bar{p}_t\big[V(b, \Delta^+) - V(b, \Delta^-)\big] \leq 0$, for $b > 0$ and $1 < \Delta^- \leq \Delta^+$. We utilize the VIA and mathematical induction to proceed with the proof. VIA converges to the value function of the Bellman equation irrespective of the initial value assigned to $V_0(s)$, i.e., $\lim_{k \to \infty} V_k(s) = V(s)\ \forall s \in S$. Therefore, it suffices to establish the following inequality for all $k \in \{0, 1, 2, \cdots\}$:

$$\begin{aligned} &\bar{\beta}\big[V_k(b-1, \Delta^+) - V_k(b-1, \Delta^-)\big] \\ &\qquad - \bar{p}_t\big[V_k(b, \Delta^+) - V_k(b, \Delta^-)\big] \leq 0. \quad (46) \end{aligned}$$

Assuming $V_0(s) = 0$ for all $s \in S$, equation (46) holds true for $k = 0$. Now, with the same assumption extending up to $k > 0$, we prove its validity for $k + 1$, i.e.,

$$\begin{aligned} &\bar{\beta}\big[V_{k+1}(b-1, \Delta^+) - V_{k+1}(b-1, \Delta^-)\big] \\ &\qquad - \bar{p}_t\big[V_{k+1}(b, \Delta^+) - V_{k+1}(b, \Delta^-)\big] \leq 0. \quad (47) \end{aligned}$$

By defining $V^0_{k+1}(s) = \sum_{s' \in S} P\left[s'|s, a=0\right] V_k(s')$, $V^1_{k+1}(s) = \sum_{s' \in S} P\left[s'|s, a=1\right] V_k(s')$,

$$\begin{aligned} V^0_{k+1}(s) &= \alpha p_t V_k(b+1, 0) + \alpha \bar{p}_t V_k(b+1, \Delta+1) \\ &\quad + \bar{\alpha} p_t V_k(b, 0) + \bar{\alpha} \bar{p}_t V_k(b, \Delta+1), \quad (48a) \\ V^1_{k+1}(s) &= \alpha\beta V_k(b, 0) + \alpha\bar{\beta} V_k(b, \Delta+1) \\ &\quad + \bar{\alpha}\beta V_k(b-1, 0) + \bar{\alpha}\bar{\beta} V_k(b-1, \Delta+1), \quad (48b) \end{aligned}$$

the VIA equation is given by $V_{k+1}(b, \Delta) = \Delta + \min\{V^0_{k+1}(b, \Delta), V^1_{k+1}(b, \Delta)\}$; thus the inequality (47) can further be simplified:

$$\begin{aligned} &\bar{\beta}\Big[\Delta^+ - \Delta^- + \min\{V^0_{k+1}(b-1, \Delta^+), V^1_{k+1}(b-1, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b-1, \Delta^-), V^1_{k+1}(b-1, \Delta^-)\}\Big] \\ &- \bar{p}_t\Big[\Delta^+ - \Delta^- + \min\{V^0_{k+1}(b, \Delta^+), V^1_{k+1}(b, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b, \Delta^-), V^1_{k+1}(b, \Delta^-)\}\Big] \leq 0 \\ &\Leftrightarrow \underbrace{(p_t - \beta)}_{<0} \underbrace{(\Delta^+ - \Delta^-)}_{\geq 0} \\ &+ \bar{\beta}\Big[\min\{V^0_{k+1}(b-1, \Delta^+), V^1_{k+1}(b-1, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b-1, \Delta^-), V^1_{k+1}(b-1, \Delta^-)\}\Big] \\ &- \bar{p}_t\Big[\min\{V^0_{k+1}(b, \Delta^+), V^1_{k+1}(b, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b, \Delta^-), V^1_{k+1}(b, \Delta^-)\}\Big] \leq 0. \quad (49) \end{aligned}$$

The first term in (49) is non-positive; thus, it is sufficient to demonstrate that:

$$\begin{aligned} &\bar{\beta}\Big[\min\{V^0_{k+1}(b-1, \Delta^+), V^1_{k+1}(b-1, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b-1, \Delta^-), V^1_{k+1}(b-1, \Delta^-)\}\Big] \\ &- \bar{p}_t\Big[\min\{V^0_{k+1}(b, \Delta^+), V^1_{k+1}(b, \Delta^+)\} \\ &\qquad - \min\{V^0_{k+1}(b, \Delta^-), V^1_{k+1}(b, \Delta^-)\}\Big] \leq 0. \quad (50) \end{aligned}$$

To proceed with the proof, we consider four cases.

Case 1. $V^0_{k+1}(b-1, \Delta^-) \leq V^1_{k+1}(b-1, \Delta^-)$ and $V^0_{k+1}(b, \Delta^+) \leq V^1_{k+1}(b, \Delta^+)$.

Case 2. $V^0_{k+1}(b-1, \Delta^-) \leq V^1_{k+1}(b-1, \Delta^-)$ and $V^0_{k+1}(b, \Delta^+) > V^1_{k+1}(b, \Delta^+)$.

Case 3. $V^0_{k+1}(b-1, \Delta^-) > V^1_{k+1}(b-1, \Delta^-)$ and $V^0_{k+1}(b, \Delta^+) \leq V^1_{k+1}(b, \Delta^+)$.

Case 4. $V^0_{k+1}(b-1, \Delta^-) > V^1_{k+1}(b-1, \Delta^-)$ and $V^0_{k+1}(b, \Delta^+) > V^1_{k+1}(b, \Delta^+)$.

Case 1. $V^0_{k+1}(b-1,\Delta^-) \leq V^1_{k+1}(b-1,\Delta^-)$ and $V^0_{k+1}(b,\Delta^+) \leq V^1_{k+1}(b,\Delta^+)$. In this case, equation (50) is further simplified:

$$\bar{\beta}\Big[\min\{V^0_{k+1}(b-1,\Delta^+), V^1_{k+1}(b-1,\Delta^+)\} - V^0_{k+1}(b-1,\Delta^-)\Big] - \bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - \min\{V^0_{k+1}(b,\Delta^-), V^1_{k+1}(b,\Delta^-)\}\Big] \leq 0. \tag{51}$$

We know that $\min\{x,y\} = x + \min\{0, y-x\}$, so we can simplify further:

$$\begin{aligned}&\bar{\beta}\Big[V^0_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^-)\Big]\\ &+\underbrace{\bar{\beta}\min\{0, V^1_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^+)\}}_{\leq 0}\\ &-\bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - V^0_{k+1}(b,\Delta^-)\Big]\\ &+\underbrace{\bar{p}_t\min\{0, V^1_{k+1}(b,\Delta^-) - V^0_{k+1}(b,\Delta^-)\}}_{\leq 0} \leq 0,\end{aligned} \tag{52}$$

where the second and last terms are negative (non-positive), thus it suffices to show that:

$$\begin{aligned}&\bar{\beta}\Big[V^0_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^-)\Big]\\ &\qquad - \bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - V^0_{k+1}(b,\Delta^-)\Big] \leq 0.\end{aligned} \tag{53}$$

According to (48), it can be written as follows:

$$\begin{aligned}&\bar{\beta}\Big\{\alpha\bar{p}_t\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{p}_t\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\Big\}\\ &-\bar{p}_t\Big\{\alpha\bar{p}_t\big[V_k(b+1,\Delta^+ +1) - V_k(b+1,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{p}_t\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\Big\} \leq 0\\ &\Leftrightarrow \alpha\bar{p}_t\Big\{\bar{\beta}\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\\ &-\bar{p}_t\big[V_k(b+1,\Delta^+ +1) - V_k(b+1,\Delta^- +1)\big]\Big\}\\ &+\bar{\alpha}\bar{p}_t\Big\{\bar{\beta}\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\\ &-\bar{p}_t\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\Big\} \leq 0.\end{aligned} \tag{54}$$

both the expressions within the braces are negative according to (46), and the proof is complete for Case 1.

Case 2. $V^0_{k+1}(b-1,\Delta^-) \leq V^1_{k+1}(b-1,\Delta^-)$ and $V^0_{k+1}(b,\Delta^+) > V^1_{k+1}(b,\Delta^+)$. In this case, equation (50) is further simplified:

$$\begin{aligned}&\bar{\beta}\Big[\min\{V^0_{k+1}(b-1,\Delta^+), V^1_{k+1}(b-1,\Delta^+)\} - V^0_{k+1}(b-1,\Delta^-)\Big]\\ &-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+) - \min\{V^0_{k+1}(b,\Delta^-), V^1_{k+1}(b,\Delta^-)\}\Big] \leq 0\\ &\Leftrightarrow \bar{\beta}\Big[V^0_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^-)\Big]\\ &+\underbrace{\bar{\beta}\min\{0, V^1_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^+)\}}_{\leq 0}\\ &-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+) - V^1_{k+1}(b,\Delta^-)\Big]\\ &+\underbrace{\bar{p}_t\min\{V^0_{k+1}(b,\Delta^-) - V^1_{k+1}(b,\Delta^-), 0\}}_{\leq 0} \leq 0.\end{aligned} \tag{55}$$

It suffices to show that:

$$\begin{aligned}&\bar{\beta}\Big[V^0_{k+1}(b-1,\Delta^+) - V^0_{k+1}(b-1,\Delta^-)\Big]\\ &-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+) - V^1_{k+1}(b,\Delta^-)\Big] \leq 0.\end{aligned} \tag{56}$$

According to (48), it can be written as follows:

$$\begin{aligned}&\bar{\beta}\Big\{\alpha\bar{p}_t\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{p}_t\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\Big\}\\ &-\bar{p}_t\Big\{\alpha\bar{\beta}\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{\beta}\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\Big\} = 0,\end{aligned} \tag{57}$$

where the equation is equal to zero, and it concludes the proof for Case 2.

Case 3. $V^0_{k+1}(b-1,\Delta^-) > V^1_{k+1}(b-1,\Delta^-)$ and $V^0_{k+1}(b,\Delta^+) \leq V^1_{k+1}(b,\Delta^+)$. In this case, equation (50) is further simplified:

$$\begin{aligned}&\bar{\beta}\Big[\min\{V^0_{k+1}(b-1,\Delta^+), V^1_{k+1}(b-1,\Delta^+)\} - V^1_{k+1}(b-1,\Delta^-)\Big]\\ &-\bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - \min\{V^0_{k+1}(b,\Delta^-), V^1_{k+1}(b,\Delta^-)\}\Big] \leq 0\\ &\Leftrightarrow \bar{\beta}\Big[V^1_{k+1}(b-1,\Delta^+) - V^1_{k+1}(b-1,\Delta^-)\Big]\\ &+\underbrace{\bar{\beta}\min\{V^0_{k+1}(b-1,\Delta^+) - V^1_{k+1}(b-1,\Delta^+), 0\}}_{\leq 0}\\ &-\bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - V^0_{k+1}(b,\Delta^-)\Big]\\ &+\underbrace{\bar{p}_t\min\{0, V^1_{k+1}(b,\Delta^-) - V^0_{k+1}(b,\Delta^-)\}}_{\leq 0} \leq 0.\end{aligned} \tag{58}$$

It suffices to show that:

$$\begin{aligned}&\bar{\beta}\Big[V^1_{k+1}(b-1,\Delta^+) - V^1_{k+1}(b-1,\Delta^-)\Big]\\ &-\bar{p}_t\Big[V^0_{k+1}(b,\Delta^+) - V^0_{k+1}(b,\Delta^-)\Big] \leq 0.\end{aligned} \tag{59}$$

According to (48), it can be written as follows:

$$\begin{aligned}&\bar{\beta}\Big\{\alpha\bar{\beta}\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{\beta}\big[V_k(b-2,\Delta^+ +1) - V_k(b-2,\Delta^- +1)\big]\Big\}\\ &-\bar{p}_t\Big\{\alpha\bar{p}_t\big[V_k(b+1,\Delta^+ +1) - V_k(b+1,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{p}_t\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\Big\} \leq 0,.\end{aligned}$$

By adding and subtracting the following expression:

$$\begin{aligned}&\bar{p}_t\Big\{\alpha\bar{\beta}\big[V_k(b,\Delta^+ +1) - V_k(b,\Delta^- +1)\big]\\ &+\bar{\alpha}\bar{\beta}\big[V_k(b-1,\Delta^+ +1) - V_k(b-1,\Delta^- +1)\big]\Big\},\end{aligned} \tag{60}$$

we have:

$$\begin{aligned}
&\bar{\beta}\Big\{\alpha\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\big[V_k(b{-}2,\Delta^+{+}1)-V_k(b{-}2,\Delta^-{+}1)\big]\Big\}\\
&-\bar{p}_t\Big\{\alpha\bar{\beta}\big[V_k(b,\Delta^+{+}1)-V_k(,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\Big\}\\
&+\bar{p}_t\Big\{\alpha\bar{\beta}\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\Big\}\\
&-\bar{p}_t\Big\{\alpha\bar{p}_t\big[V_k(b{+}1,\Delta^+{+}1)-V_k(b{+}1,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{p}_t\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\Big\}\le 0, \qquad (61)
\end{aligned}$$

and by rearranging the equation:

$$\begin{aligned}
&\alpha\bar{\beta}\Big\{\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\\
&-\bar{p}_t\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\Big\}\\
&+\bar{\alpha}\bar{\beta}\Big\{\bar{\beta}\big[V_k(b{-}2,\Delta^+{+}1)-V_k(b{-}2,\Delta^-{+}1)\big]\\
&-\bar{p}_t\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\Big\}\\
&+\alpha\bar{p}_t\Big\{\bar{\beta}\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\\
&-\bar{p}_t\big[V_k(b{+}1,\Delta^+{+}1)-V_k(b{+}1,\Delta^-{+}1)\big]\Big\}\\
&+\bar{\alpha}\bar{p}_t\Big\{\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\\
&+\bar{p}_t\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\Big\}\le 0, \qquad (62)
\end{aligned}$$

where all the expressions within the braces are negative according to (46), and it concludes the proof for Case 3.

Case 4. $V^0_{k+1}(b-1,\Delta^-) > V^1_{k+1}(b-1,\Delta^-)$ and $V^0_{k+1}(b,\Delta^+) > V^1_{k+1}(b,\Delta^+)$. In this case, equation (50) is further simplified:

$$\begin{aligned}
&\bar{\beta}\Big[\min\{V^0_{k+1}(b{-}1,\Delta^+),V^1_{k+1}(b{-}1,\Delta^+)\}{-}1V^1_{k+1}(b{-}1,\Delta^-)\Big]\\
&-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+)-\min\{V^0_{k+1}(b,\Delta^-),V^1_{k+1}(b,\Delta^-)\}\Big]\le 0\\
&\Leftrightarrow \bar{\beta}\Big[V^1_{k+1}(b{-}1,\Delta^+){-}V^1_{k+1}(b{-}1,\Delta^-)\Big]\\
&+\underbrace{\bar{\beta}\min\{V^0_{k+1}(b{-}1,\Delta^+){-}V^1_{k+1}(b{-}1,\Delta^+),0\}}_{\le 0}\\
&-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+){-}V^1_{k+1}(b,\Delta^-)\Big]\\
&+\underbrace{\bar{p}_t\min\{V^0_{k+1}(b,\Delta^-){-}V^1_{k+1}(b,\Delta^-),0\}}_{\le 0}\le 0. \qquad (63)
\end{aligned}$$

It suffices to show that:

$$\begin{aligned}
&\bar{\beta}\Big[V^1_{k+1}(b{-}1,\Delta^+){-}V^1_{k+1}(b{-}1,\Delta^-)\Big]\\
&-\bar{p}_t\Big[V^1_{k+1}(b,\Delta^+){-}V^1_{k+1}(b,\Delta^-)\Big]\le 0, \qquad (64)
\end{aligned}$$

According to (48), it can be written as follows:

$$\begin{aligned}
&\bar{\beta}\Big\{\alpha\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\big[V_k(b{-}2,\Delta^+{+}1)-V_k(b{-}2,\Delta^-{+}1)\big]\Big\}\\
&-\bar{p}_t\Big\{\alpha\bar{\beta}\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1){-}V_k(b{-}1,\Delta^-{+}1)\big]\Big\}\le 0,\\
&\Leftrightarrow \alpha\bar{\beta}\Big\{\bar{\beta}\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\\
&-\bar{p}_t\big[V_k(b,\Delta^+{+}1)-V_k(b,\Delta^-{+}1)\big]\\
&+\bar{\alpha}\bar{\beta}\Big\{\bar{\beta}\big[V_k(b{-}2,\Delta^+{+}1){-}V_k(b{-}2,\Delta^-{+}1)\big]\\
&-\bar{p}_t\big[V_k(b{-}1,\Delta^+{+}1)-V_k(b{-}1,\Delta^-{+}1)\big]\Big\}\le 0, \qquad (65)
\end{aligned}$$

both expressions within the braces are negative according to (46), concluding the proof for Case 3 and Theorem 1. □