

State-Aware Timeliness in Energy Harvesting IoT Systems Monitoring a Markovian Source

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

Abstract—In this study, we investigate the optimal transmission policies within an energy harvesting status update system, where the demand for status updates depends on the state of the source. The system monitors a source with two contextual states characterized by a Markovian stochastic process, which can be in either a *normal* state or an *alarm* state, with a higher demand for fresh updates when the source is in the alarm state. We propose a metric to capture the freshness of status updates for each state of the stochastic process by introducing two Age of Information (AoI) variables, extending the definition of AoI to account for the state changes of the stochastic process. We formulate the problem as a Markov Decision Process (MDP), utilizing a transition cost function that applies linear and non-linear penalties based on AoI and the state of the stochastic process. Through analytical investigation, we delve into the structure of the optimal transmission policy for the resulting MDP problem. Furthermore, we evaluate the derived policies via numerical results and demonstrate their effectiveness in reserving energy in anticipation of forthcoming alarm states.

I. INTRODUCTION

Timely communication of status updates is critically essential for applications providing monitoring services in cyber-physical systems [2]. These applications form the foundation of the intelligent infrastructure enabled by the Internet of Things (IoT). Instances of such applications encompass, but are not restricted to, smart cities, intelligent factories and grids, advanced agriculture, parking and traffic control, e-Health, and environmental monitoring [2], [3].

A pivotal finding in the field indicated that metrics like throughput and delay do not adequately address the goal of timely status updating. In addressing this issue, the authors in [4] introduced a novel metric known as the Age of Information (AoI). Since its introduction, the optimal determination of status update generation and transmission to minimize AoI metrics has garnered considerable attention from the research community [5]–[7]. The scope of AoI has been extended to encompass other metrics, including the Value of Information [8], cost of update delay [9], Age of Synchronization [10], non-linear AoI [11], Age of Incorrect Information [12], Version Age of Information [13], and Age of Actuation [14].

E. Delfani and N. Pappas are with the Department of Computer and Information Science Linköping University, Sweden, email: {erfan.delfani, nikolaos.pappas}@liu.se. G. Stamatakis is with the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), email: gstam@ics.forth.gr. This work has been supported in part by the Swedish Research Council (VR), ELLIIT, Zenith, and the European Union (ETHER, 101096526, ELIXIRION, 101120135, and SOVEREIGN, 101131481). A shorter version has been published in [1].

Another notable challenge in the field involves selecting a suitable energy source for remote sensors. With their finite lifespan, batteries pose the risk of high replacement costs, particularly when dealing with numerous sensors located in remote or inaccessible areas. To tackle this issue, energy harvesting (EH) technologies have been devised to provide the required power to remote sensors [15]. Regardless of whether energy harvesting, batteries, or both are employed, the stored energy must be judiciously managed to ensure an adequate supply when most crucial.

In [16], the paper explores the optimization of transmitting updates from an EH source to a receiver, aiming to minimize the average age of updates. Similar studies can be found in [17]–[24]. The paper [25] explores a monitoring system where nodes are powered wirelessly and send updates to a central node to maintain data freshness. It aims to minimize the average AoI by optimizing energy transfer and update scheduling. Using deep reinforcement learning, the paper proposes an efficient solution and analyzes its properties. It also compares the optimal policy with one maximizing throughput and studies the impact of system parameters. In [26], the study examines the average AoI in a wireless power transfer sensor network. Additionally, [27] investigates the interplay of throughput/delay and AoI in a two-user multiple access channel with a single energy harvesting source. In [28], the study examines the average AoI for status updates from an EH transmitter with a finite-capacity battery. The research investigates optimal scheduling policies under known channel and EH statistics. In cases of unknown environments, the authors propose an adaptive reinforcement learning algorithm to learn system parameters and update policies in real-time. In [29], the study focuses on a cognitive radio system with a secondary user as an EH sensor, deciding between spectrum sensing and status updating in each time slot. The sequential decision-making problem is framed as a Partially Observable MDP (POMDP) and solved using dynamic programming, with an exploration of the optimal policy's structural properties. Another study [30] tackles real-time IoT applications using EH sensors, aiming to minimize the Age of Correlated Information (AoCI) at the data fusion center. The approach involves formulating the dynamic status update as a POMDP and introducing a DRL algorithm to solve the problem. The study [31] focuses on optimizing wireless communication of stochastic process samples to minimize distortion at the destination while maintaining a specified AoI and cost of actions. It introduces a stationary randomized policy (SRP) solution and highlights challenges related

to rapid source changes and channel states. Additionally, a constrained POMDP formulation for the problem has been defined. The article [32] optimizes IoT systems by minimizing AoI and distortion through effective policies, including save-and-transmit and fixed power transmission. Causal EH information is addressed with an MDP for optimal policy. The study reveals that the optimal transmit power is a bivalued function of the current age and distortion. The authors of [33] study an EH monitoring node managing updates from diverse sources with different energy consumption and AoI values. The objective is to minimize average AoI through optimal actions (requesting an update from a source or staying idle) formulated as an MDP, with the optimal policy determined using the Value Iteration algorithm. In [34], the focus is on minimizing on-demand AoI in a multi-user IoT energy harvesting network, using an MDP formulation. The study proposes an iterative algorithm for optimal status updates, with a low-complexity alternative for scenarios with numerous sensors. In [35], the problem is tackled without transmission constraints, employing a model-free Q-learning method within an MDP framework. [36] introduces a pull-based communication model using the Age of Information at Query (QAoI) metric in an MDP, determining the optimal status updating policy for a monitoring scenario with periodic queries from a server to an EH sensor at an edge node. The paper [37] investigates online scheduling in wireless-powered communication networks for IoT devices. It focuses on minimizing the Expected Weighted Sum Age of Information (EWSAoI) by proposing a Max-Weight policy based on Lyapunov optimization theory. This policy schedules sensor nodes to transmit their data to a mobile edge server efficiently, considering wireless power transfer and channel fading effects. Additionally, The work [38] examines and optimizes a real-time IoT network, considering energy harvesting, caching, and gossiping. It focuses on minimizing the average Version AoI in a destination gossiping network while managing energy constraints for the EH sensor and responding to network requests, utilizing the MDP framework. The work [39] deals with updating information efficiently for an EH IoT receiver that interacts with a variable-rate information source. It aims to minimize the average AoI by optimizing when the receiver turns on or off. The study uses the MDP framework to find optimal scheduling policies and introduces a state-adapted waiting policy.

In this study, we demonstrate the close connection between the challenge of reserving energy for *critical* use and the issue of ensuring timely status updates. Specifically, we examine an energy harvesting (EH) status update system that monitors a source with two *contextual* or *operational* states, a *normal* state, and an *alarm* state. This framework encompasses systems where events occur with a certain probability at defined time intervals during normal operation, while the probability increases significantly during alarm operation. For instance, this scenario is applicable to networks, where the rate of packet arrivals during a denial of service attack contrasts with normal operation. Additionally, our focus is on systems where the demand for fresh status updates is considerably higher during alarm periods than in normal operation. To address this heightened demand, the system needs to account for the

characteristics of the energy arrival process and strategically reserve energy when feasible.

To the best of our knowledge, this is the first work to consider an AoI-based status update system for a two-state stochastic process and study the impact of constrained energy resources on the optimal status update transmission policies. For an effective representation of the problem, we introduce two AoI variables, each corresponding to a state of the stochastic process. We expand the AoI definition to encompass scenarios where the state of the stochastic process changes without the monitoring application being informed of the change. Finally, our results illustrate how the optimal policy is influenced by the probabilities of energy harvesting, successful status update transmission, and the probability of the monitored process changing state from its current state.

II. SYSTEM MODEL

The system we consider is presented in Fig. 1 and comprises an Energy Harvesting sensor responsible for monitoring a source, whose contextual or operational states are represented by a stochastic process, and sending status updates to a destination node, denoted as Rx. We assume that Rx is one hop away from the sensor, time is slotted, and each slot has a duration of T . At the beginning of each time slot, the stochastic

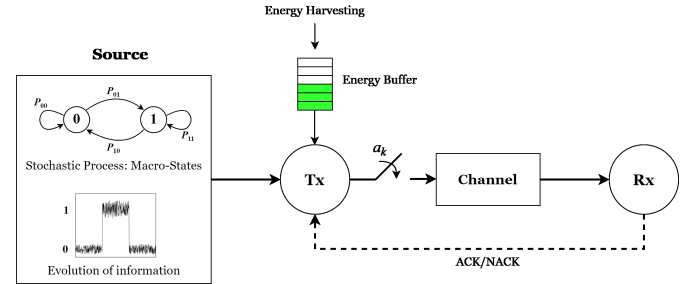


Fig. 1: An EH status update system for a stochastic process with normal and alarm states.

process can exist in one of two states. The first state, 0, indicates a *normal* operational state. In contrast, the second state, 1, signifies an *alarm* operational state. An illustrative representation of such a process is presented in Fig. 1. It is anticipated that a monitoring application for this stochastic process should deliver more frequent status updates during alarm periods. Let $\{Z_k\}$, where $k = 0, 1, \dots$, represent the sequence of states of the stochastic process over time. We assume that the state of the stochastic process remains constant throughout a time slot. At the onset of the $(k+1)$ -th time-slot, the state of the stochastic process transitions from $Z_k = z$ to $Z_{k+1} = z'$, governed by transition probabilities $P_{zz'}$, where $z, z' \in \{0, 1\}$, as depicted in Fig. 1.

At the beginning of each time slot, the sensor generates a new status update and subsequently decides whether to transmit it to the destination. The sensor has an energy buffer capable of storing an integer number of energy units, with a maximum capacity of E_{max} energy units. The sensor has a probability P_e of harvesting an energy unit in a given time slot. We assume that each status update transmission

consumes one energy unit, and no transmission is possible if the energy buffer is empty. For the purposes of this study, we do not consider energy costs associated with other sensor functions, such as sensing, processing, and data storage in memory. Each transmission has an independent probability of success, denoted as P_s , and this probability is unaffected by the outcomes of previous transmissions. Additionally, we assume that acknowledgment of a packet transmission occurs instantaneously.

We employ the AoI metric to quantify the timeliness of status updates reaching the destination. AoI, as defined in [4], represents the time elapsed since the generation of the last successfully decoded status update. However, our study must also account for state changes in the stochastic process. The destination remains unaware of any such state change until it receives a fresh status update. Additionally, the sensor node faces the challenge of deciding when to transmit a new status update, considering both the increased (or decreased) demand during alarm (normal) states of the stochastic process and the limited energy resources in the buffer. The sensor must leverage its knowledge of the stochastic process's state changes and the AoI value at the destination to achieve this objective.

To address this scenario, we employ two distinct AoI variables, each corresponding to a different state of the stochastic process. We represent the AoI for the z -th state of the stochastic process at time k as Δ_k^z , $z \in \{0, 1\}$. Additionally, we use the sequence of time indices where a state change occurs, denoted as $\{\tau_n : Z_{\tau_n} \neq Z_{\tau_n-1}, n = 1, 2, \dots\}$, and define τ_N as the time index of the most recent state change for the stochastic process by time k , where $N = \max\{n : \tau_n < k\}$. Lastly, let Z_k^d represent the state that the destination *knows* as the process's state at time k , indicating the state of the stochastic process included in the most recent status update received by the destination. The definition of AoI is then as follows,

$$\Delta_k^z = \begin{cases} \min\{k - U_k, \Delta_{max}^z\}, & \text{if } z = Z_k^d, \\ \min\{k - \tau_N, \Delta_{max}^z\}, & \text{if } z \neq Z_k^d \text{ and } z = Z_k, \\ 0, & \text{if } z \neq Z_k^d \text{ and } z \neq Z_k, \end{cases} \quad (1)$$

where U_k denotes the timestamp of the most recent packet received at the destination by time k , and Δ_{max}^z represents the maximum value of AoI associated with the highest level of staleness.

The first branch of (1) applies to the AoI variable associated with the state of the stochastic process known at the destination by time k , aligning with the definition of AoI as presented in [4]. The second branch of (1) is applicable in scenarios where *one or more state changes* have occurred, leading to the current state of the stochastic process differing from the one recognized at the destination ($Z_k \neq Z_k^d$). In such instances, the AoI for $z = Z_k$, denoted as Δ_k^z , is defined as the time elapsed since the last state change (τ_N). Finally, the third branch of the equation applies for AoI Δ_k^z when z is neither the state known by the destination nor the currently active state, i.e., the state known to the destination at the k -th time slot, Z_k^d , is equal to the actual state of the stochastic process. In such a case, an AoI value of zero is assigned.

Since each AoI variable Δ_k^z is utilized only when the active state of the stochastic process is equal to z , i.e., $z = Z_k$, the last branch of (1) becomes irrelevant. Consequently, the definition of the two random variables Δ_k^z for $z \in \{0, 1\}$ can be briefly presented as shown in (3).

$$\Delta_k^z = \begin{cases} \min\{k - U_k, \Delta_{max}^z\}, & \text{if } z = Z_k = Z_k^d, \\ \min\{k - \tau_N, \Delta_{max}^z\}, & \text{if } z \neq Z_k^d \text{ and } z = Z_k. \end{cases} \quad (2)$$

An illustration in Fig. 2 depicts the evolution of Δ_k^0 and Δ_k^1 over time. Status updates occur at t_k ($k \geq 0$), reaching the destination at time points denoted as t'_k . τ_c ($c \geq 0$) indicates times of stochastic process state changes. At $k = 2$, the destination receives a status update indicating $Z_0 = 0$, causing Δ_k^0 to increase while Δ_k^1 remains zero. At $k = 5$ (τ_1), the process shifts to $Z_5 = 1$, incrementing both Δ_k^0 and Δ_k^1 following the first and second branches of (1), respectively. At $k = 9$, the destination receives an update confirming a state change at $\tau_1 = 5$, resetting Δ_k^0 to zero according to the third branch of (1), and continuing the increment of Δ_k^1 following the first branch. Finally, at $k = 12$ (τ_2), another state change occurs, repeating the process.

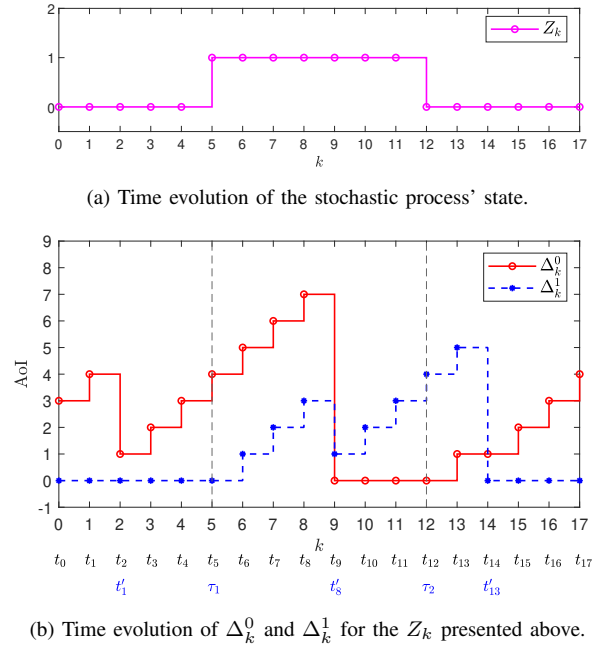


Fig. 2: The first sub-figure presents the time evolution of the stochastic process' state. The second sub-figure presents the evolution of the AoI for each state of the stochastic process.

By employing these two AoI variables, we will be able to formulate various metrics or cost (reward) functions related to the staleness (freshness) of a system in two distinct states with different demands. The subsequent sections will further elucidate these metrics.

A. Contextual States: Macro- and Micro-Evolution of the Source

It is important to note that in this system setup, the information source has two *operational/contextual* states, or *macro-states*, within which underlying events, or *micro-states*,

evolve. These micro-states represent the raw evolution of the source. The contextual macro-states, represented by the stochastic process Z_k , primarily determine the level of demand for timeliness, with a higher freshness required in the alarm state. This work addresses the timeliness of the source's micro-evolution under the contextual macro-states, i.e., the AoI during the different macro-states. Another interpretation involves transmitting the exact macro-states (normal and alarm states) as the primary information packets. In other words, this scenario considers the stochastic process as the primary source of information. In this context, both the accuracy of the information-whether correct or incorrect-and the timeliness of updates are considered. The AoI variables evaluate and penalize both the staleness of correct information and the duration of incorrect information separately for the macro-states. The penalty is applied both when the receiver is synchronized or in the correct macro-state (as indicated by the regular AoI metric according to the first branch of (1)) and when the receiver is in an incorrect macro-state (due to the state change at the source, as indicated by the second branch of (1)). The key point is that this penalty has a greater impact when the source is in the alarm state. This is achieved by defining appropriate metrics, as presented in Section III.

Regarding this latter interpretation, an additional penalty for having an incorrect macro-state can also be incorporated into the defined AoI variables. The first branch of (1) captures the staleness of information when the system is in the correct macro-state ($Z_k = Z_k^d$), whereas the second branch captures the duration of time that the system is in an incorrect macro-state ($Z_k \neq Z_k^d$). By introducing a weight $\alpha \geq 0$ in the first branch of the equation, we can control the balance between the penalties associated with these two system dynamics. For instance, when $\alpha = 0$, Δ_k^z transforms to the Age of Incorrect Information (AoII) [12] for a two-state system, as it solely measures the elapsed time during which the receiver has incorrect information.

$$\Delta_k^z = \begin{cases} \min\{\alpha(k - U_k), \Delta_{max}^z\}, & \text{if } z = Z_k = Z_k^d, \\ \min\{k - \tau_N, \Delta_{max}^z\}, & \text{if } z \neq Z_k^d \text{ and } z = Z_k. \end{cases} \quad (3)$$

III. PROBLEM FORMULATION

Our objective is to model and optimize the freshness of information in the aforementioned status update system in Fig. 1 with a two-state Markovian source, where the demand for fresher updates in the alarm state is higher than in the normal state. To achieve this, we formulate an MDP problem in this section, using a proposed metric as the cost function, as detailed below. In the subsequent sections, we will analyze the optimal transmission policy for the system. We first present the state, action, and random variable spaces of the system, as well as the system's transition and cost functions.

States: At the beginning of the k -th time-slot the state of the system is represented by the following state vector,

$$s_k = [Z_k, Z_k^d, E_k, \Delta_k^0, \Delta_k^1]^T, \quad (4)$$

where $Z_k \in \{0, 1\}$ represents the state of the stochastic process, Z_k^d signifies the state known by the destination at time k , $E_k = \{0, 1, \dots, E_{max}\}$ is the energy in the buffer, and

$\Delta_k^z, z \in \{0, 1\}$ is the AoI at the destination for the z -th state of the stochastic process as defined by (1), with T denoting the transpose operator. The set of all system states is denoted as S .

Actions: When at least one energy unit is in the buffer, the sensing node can choose to transmit a fresh status update or conserve energy for later use. The action taken by the sensing node is denoted as $a_k \in \{0, 1\}$, where 0 indicates not transmitting a status update, and 1 indicates transmitting one. If the energy buffer is empty, the sensor is restricted to action 0. We use $a^*(s)$ to denote the optimal action in state s , A for the set of all actions, and $A(s)$ to represent the set of permissible actions at state s .

Random variables: Given the system's state and the sensor's action, a stochastic transition to a new state occurs, determined by three random variables. The first, $W_k^s \in \{0, 1\}$, signifies the random event of a successful transmission over the noisy channel, assumed to happen with probability P_s . If the sensor opts not to transmit at time-slot k , W_k^s is forced to be zero. The second variable, $W_k^e \in \{0, 1\}$, represents the random event of an energy unit arrival, assumed to occur with probability P_e during a time slot. The third, $W_k^z \in \{0, 1\}$, denotes the new state of the random process, determined by transition probabilities presented in Fig. 1. These random variables' values become known to the sensor at the end of the k -th time-slot, as typical in optimal control theory [40]. Lastly, we assume independence among W_k^s , W_k^e , and W_k^z , with their values being independent of previous time slots and identically distributed across all time slots. The random column vector $W_k = [W_k^s, W_k^e, W_k^z]^T$ collectively refers to the system's random variables.

System Dynamics: Given the current state of the system $s_k = [Z_k, Z_k^d, E_k, \Delta_k^0, \Delta_k^1]^T$ and the action a_k , the next state of system $s_{k+1} = [Z_{k+1}, Z_{k+1}^d, E_{k+1}, \Delta_{k+1}^0, \Delta_{k+1}^1]^T$ is determined by the realization of random vector $W_k = [W_k^s, W_k^e, W_k^z]$. More specifically the state of the stochastic process at the $(k+1)$ -th time-slot is provided by the random variable W_k^z whose value becomes known by the end of the k -th time-slot.

$$Z_{k+1} = W_k^z, \quad (5)$$

while the state of the stochastic process known by the destination assumes a new value only in the case of a successful status update transmission,

$$Z_{k+1}^d = \begin{cases} Z_k^d & W_k^s = 0, \\ Z_k & W_k^s = 1. \end{cases} \quad (6)$$

The energy stored in the energy buffer at the beginning of the $(k+1)$ -th time-slot depends on whether the sensor transmitted a status update and an energy unit was harvested during the k -th time-slot,

$$E_{k+1} = \min\{E_k + W_k^e - a_k, E_{max}\}. \quad (7)$$

Here, we present a recursive definition for Δ_{k+1}^z , although the evolution of the AoI variables over time was described in (1),

$$\Delta_{k+1}^z = \begin{cases} 0 & (z \neq Z_k, z \neq Z_k^d, W_k^s = 0) \\ & \text{or } (z \neq Z_k, W_k^s = 1), \\ \min \{\Delta_k^z + 1, \Delta_{max}^z\} & (z = Z_k = Z_k^d, W_k^s = 0) \\ & \text{or } (Z_k \neq Z_k^d, z \in \{0, 1\}, W_k^s = 0), \\ 1 & (z = Z_k, W_k^s = 1). \end{cases} \quad (8)$$

Transition Probabilities: The transition probability of the system can be represented by the total probability theorem as follows:

$$\begin{aligned} P[s_{k+1}|s_k, a_k] &= \sum_{W_k} P[s_{k+1}, W_k|s_k, a_k] \\ &= \sum_{W_k} P[s_{k+1}|s_k, a_k, W_k] P[W_k|s_k, a_k] \\ &= \sum_{[W_k^s, W_k^e, W_k^z]} P[s_{k+1}|s_k, a_k, W_k^s, W_k^e, W_k^z] P[W_k^s, W_k^e, W_k^z|s_k, a_k]. \end{aligned} \quad (9)$$

We can simplify the conditional probabilities in 9:

$$\begin{aligned} P[s_{k+1}|s_k, a_k, W_k^s, W_k^e, W_k^z] &= P[Z_{k+1}, Z_{k+1}^d, E_{k+1}, \Delta_{k+1}^0, \Delta_{k+1}^1|s_k, a_k, W_k^s, W_k^e, W_k^z] \\ &= P[Z_{k+1}|W_k^z] \times P[Z_{k+1}^d|Z_k^d, W_k^s] \times P[E_{k+1}|E_k, a_k, W_k^e] \\ &\times P[\Delta_{k+1}^0|Z_k, Z_k^d, \Delta_k^0, W_k^s] \times P[\Delta_{k+1}^1|Z_k, Z_k^d, \Delta_k^1, W_k^s], \end{aligned} \quad (10)$$

$$\begin{aligned} P[W_k^s, W_k^e, W_k^z|s_k, a_k] &= P[W_k^s|s_k, a_k] P[W_k^e|s_k, a_k] P[W_k^z|s_k, a_k] \\ &= P[W_k^s|E_k, a_k] P[W_k^e] P[W_k^z|Z_k], \end{aligned} \quad (11)$$

where:

$$P[Z_{k+1}|W_k^z] = \begin{cases} 1 & Z_{k+1} = W_k^z, \\ 0 & \text{otherwise}, \end{cases} \quad (12)$$

$$P[Z_{k+1}^d|Z_k^d, W_k^s] = \begin{cases} 1 & Z_{k+1}^d = Z_k^d, W_k^s = 0, \\ 1 & Z_{k+1}^d = Z_k, W_k^s = 1, \\ 0 & \text{otherwise}, \end{cases} \quad (13)$$

$$P[E_{k+1}|E_k, a_k, W_k^e] = \begin{cases} 1 & E_{k+1} = \min\{E_k + W_k^e - a_k, E_{max}\}, \\ 0 & \text{otherwise}, \end{cases} \quad (14)$$

$$\begin{aligned} P[\Delta_{k+1}^z|Z_k, Z_k^d, \Delta_k^z, W_k^s] &= \begin{cases} 1 & \Delta_{k+1}^z = 0, (z \neq Z_k, z \neq Z_k^d, W_k^s = 0) \\ & \text{or } (z \neq Z_k, W_k^s = 1), \\ 1 & \Delta_{k+1}^z = \min\{\Delta_k^z + 1, \Delta_{max}^z\}, (z = Z_k = Z_k^d, W_k^s = 0) \\ & \text{or } (Z_k \neq Z_k^d, z \in \{0, 1\}, W_k^s = 0), \\ 1 & \Delta_{k+1}^z = 1, (z = Z_k, W_k^s = 1). \end{cases} \end{aligned} \quad (15)$$

$$P[W_k^s|E_k, a_k] = \begin{cases} 1 & W_k^s = 0, (a_k = 0 \text{ or } E_k = 0), \\ P_s & W_k^s = 1, a_k = 1, \\ 1 - P_s & W_k^s = 0, a_k = 1, \end{cases} \quad (16)$$

$$P[W_k^e] = \begin{cases} P_e & W_k^e = 1, \\ 1 - P_e & W_k^e = 0, \end{cases} \quad (17)$$

$$P[W_k^z|Z_k] = \begin{cases} P_{00} & W_k^z = 0, Z_k = 0, \\ P_{01} & W_k^z = 1, Z_k = 0, \\ P_{10} & W_k^z = 0, Z_k = 1, \\ P_{11} & W_k^z = 1, Z_k = 1. \end{cases} \quad (18)$$

By substituting equations (12) to (15) into (10), and equations (16) to (18) into (11), the transition probability (9) is determined.

Transition cost function: We define a general metric as the cost function as follows:

$$g(s_k, a_k, w_k) = (1 - Z_k) \cdot f(\Delta_k^0) + Z_k \cdot h(\Delta_k^1), \quad (19)$$

where $f(\cdot)$ and $h(\cdot)$ are two real-valued functions defined on non-negative integers, with the condition that $h(\cdot)$ is larger than $f(\cdot)$, i.e., $h(\Delta_k) \geq f(\Delta_k), \forall \Delta_k \in \{0, 1, 2, \dots\}$. Here, for simplicity, we consider the linear and square functions for $f(\cdot)$ and $h(\cdot)$, respectively, where the cost associated with each state transition is given by,

$$g(s_k, a_k, w_k) = g(s_k, a_k) = (1 - Z_k) \cdot \Delta_k^0 + Z_k \cdot (\Delta_k^1)^2, \quad (20)$$

where w_k is the realization of random vector W_k at the k -th time-slot. From (20) we observe that when the stochastic process is in the normal state ($Z_k = 0$), the transition cost increases linearly with Δ_k^0 , while when the stochastic process is in the alarm state ($Z_k = 1$) the transition cost increases with the square of Δ_k^1 . Thus, the transition cost function captures the increased demand for status updates when $Z_k = 1$.

Total cost function: We aim to minimize the cumulative cost over an infinite time span,

$$J_\mu(s_0) = \lim_{N \rightarrow \infty} \mathbb{E}_{W_k, k=0,1,\dots} \left\{ \sum_{k=0}^{N-1} \gamma^k g(s_k, a_k, w_k) | s_0 \right\}, \quad (21)$$

where s_0 denotes the initial state of the system, the expectation $\mathbb{E}\{\cdot\}$ is computed based on the joint probability distribution of random variables W_k for $k \in \{0, 1, \dots\}$, and γ serves as a discount factor (where $0 < \gamma < 1$), indicating diminishing importance of induced cost over time. Lastly, let $\mu = \{u_0, u_1, u_2, \dots, u_k, \dots\}$ represent a deterministic policy mapping each state s_k to a specific action $a_k = u_k(s_k)$ at each time slot k .

Our objective is to obtain an optimal policy $\mu^* = \{u_0^*, u_1^*, u_2^*, \dots\}$ that minimizes (21).

IV. ANALYTICAL RESULTS

A. Optimal Policy

The dynamic system outlined in section III is characterized by finite state, control, and probability spaces. State transitions rely on s_k , a_k , and w_k , independent of their previous values. Furthermore, the probability distribution of random variables remains constant over time. The cost linked to a state transition is bounded, and the cost function $J(\cdot)$ accumulates additively over time. These structural characteristics establish the considered dynamic system as a Markov Decision Process (MDP), where the state transition probabilities completely describe its dynamics. Specifically, the problem (21) is an infinite horizon discounted cost MDP problem with bounded cost per stage [40, Sec. 1.2]. For the MDP under consideration,

given that $0 < \gamma < 1$, there exists an optimal stationary policy μ^* which is characterized by Bellman's equation [40, Prop. 1.2.5, pg. 17]. Specifically, when the system is in state s , the optimal stationary policy μ^* always applies the same control $a^*(s)$ that minimizes (21), i.e.,

$$\mu^* = \arg \min_{\mu \in \mathcal{M}} J_\mu(s), \quad (22)$$

where $a^*(s) = u^*(s)$, for all $s \in S$, and \mathcal{M} is the set of all policies. Let $V^*(s) = J^*(s)$ be the infinite horizon discounted cost attained when the optimal policy μ^* is applied and the system begins at state s . The optimal cost $V^*(s)$ and the optimal action $a^*(s)$ satisfy the Bellman's equation:

$$V^*(s) = \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) [g(s, a) + \gamma V^*(\tilde{s})] \right\}, \forall s \in S, \quad (23)$$

$$a^*(s) = \arg \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) [g(s, a) + \gamma V^*(\tilde{s})] \right\}, \forall s \in S, \quad (24)$$

where $s = [Z, Z^d, E, \Delta^0, \Delta^1]^T$, $\tilde{s} = [\tilde{Z}_k, \tilde{Z}_k^d, \tilde{E}_k, \tilde{\Delta}_k^0, \tilde{\Delta}_k^1]^T$ and $V(s)$ is the value function of the MDP problem.

Given that the transition cost $g(s, a)$ is bounded and that $0 < \gamma < 1$, the operator,

$$(TV)(s) = \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) [g(s, a) + \gamma V(\tilde{s})] \right\}, \quad (25)$$

is a contraction mapping [40, Assumption D, Prop. 1.2.1, pg. 14] and starting with an arbitrarily initialized vector $V(s)$, $s \in S$, and repeatedly applying transformation (TV) for all states $s \in S$ we attain the optimal cost V^* and at the same time derive the optimal policy μ^* for all $s \in S$ according to [40, Prop. 1.2.1, pg. 14] which states that,

$$V^*(s) = \lim_{m \rightarrow \infty} (T^m V)(s), \quad (26)$$

where $(T^m V)(s) = (T(T^{m-1} \dots (T^0 V)))(s)$ and $(T^0 V)(s) = V(s)$. (25) is a formal description of the Value Iteration (VI) algorithm [40, Section 2.2, pg. 84].

B. Threshold Policy

Definition 1. Policy μ is a threshold policy if for each combination of values for Z , Z^d , and E there exists a threshold $\delta_T = (\Delta_T^0, \Delta_T^1)$ such that then sensor will transmit, i.e. $a(s) = 1$, only if $\delta = (\Delta^0, \Delta^1) \geq (\Delta_T^0, \Delta_T^1) = \delta_T$, where \geq is meant to hold element-wise.

Theorem 1. An optimal policy of the MDP problem is a threshold policy.

Proof. The proof can be found in appendix A. \square

Demonstrating that the optimal policy is a threshold policy holds significant theoretical and practical value. A threshold policy, defined by its threshold levels, can simplify the theoretical analysis and algorithmic approach to solving MDPs. For a one-dimensional state variable, using a threshold policy

with a specific threshold transforms the MDP into a Markov chain, facilitating easier analysis and optimization. However, this approach becomes more complex with multi-dimensional state variables. Despite this, identifying the optimal policy as a threshold policy could streamline iterative algorithms (such as Value Iteration and Policy Iteration) or inspire new algorithms for MDPs. We will defer this exploration to future research and proceed with the standard Value Iteration Algorithm (VIA) to solve the MDP problem.

V. NUMERICAL RESULTS

In this section, we conduct a numerical evaluation of the optimal infinite horizon discounted cost, $J^*(\cdot)$, under different system parameter configurations. For consistency across all experiments, we fix the discount factor at $\gamma = 0.99$, set both AoI variables' upper bounds (Δ_{max}^0 and Δ_{max}^1) to 10, and for ease of interpretation, assume a constant initial state s_0 for the system. More specifically, we assume the deployment of the sensor during the normal state ($Z_k = 0$) of the random process, with this information known to the destination ($Z_k^d = 0$). The energy buffer starts empty ($E_k = 0$), and the initial state s_0 is defined as $[0, 0, 0, 1, 0]^T$ with AoI counters Δ_k^0 and Δ_k^1 set to 1 and 0, respectively.

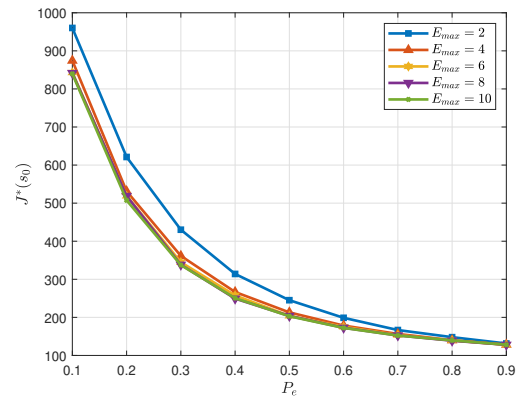


Fig. 3: Impact of energy buffer's capacity, E_{max} , on $J^*(s_0)$.

In Fig. 3, we display $J^*(s_0)$ across various capacities of the energy buffer E_{max} while the energy harvesting probability P_e varies. In all experiments depicted in Fig. 3, the state transition probabilities of the stochastic process were configured as follows:

$$P_z = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}, \quad (27)$$

and the transmission success probability P_s was set to 0.8. Fig. 3 illustrates that being in an environment with a high probability of energy harvesting and having a larger capacity energy buffer contributes positively to reducing $J^*(s_0)$. The results in Fig. 3 also indicate that the influence of the energy buffer's capacity on $J^*(s_0)$ becomes negligible when E_{max} exceeds a certain threshold for the given system configuration.

Fig. 4 illustrates $J^*(s_0)$ in relation to P_e for various transmission success probabilities P_s . In this series of experiments, P_z corresponds to the matrix defined in (27), and E_{max} was

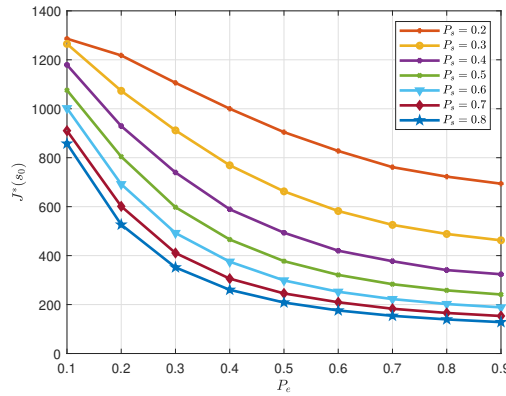


Fig. 4: Impact of transmission success probability P_s on $J^*(s_0)$.

fixed at 5. The figure indicates that an increase in P_s consistently leads to a reduction in $J^*(s_0)$. Additionally, the results suggest that, given the energy buffer's capacity, targeting a higher P_s value in environments with a low probability of energy harvesting is advisable.

Fig. 5 depicts $J^*(s_0)$ for various combinations of state transition probabilities governing the stochastic process. In the figure, the probability P_{01} (or P_{10}) represents the probability of the stochastic process transitioning from the normal (alarm) state to the alarm (normal) state by the end of a time slot. The probabilities P_{00} and P_{11} are calculated as $1 - P_{01}$ and $1 - P_{10}$ respectively. The highest value of $J^*(s_0)$ is

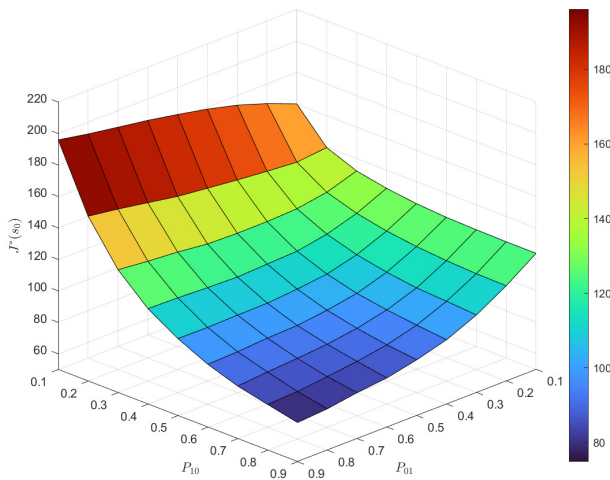


Fig. 5: Impact of different stochastic process's state transition probabilities on $J^*(s_0)$.

observed when $(P_{01}, P_{10}) = (0.9, 0.1)$, indicating a situation where the stochastic process is highly likely to transition from the normal to the alarm state and, once in the alarm state, has a low probability of returning to the normal state. The mentioned cost decreases as the probability of returning to the normal state, P_{10} , increases. One might anticipate a similar cost reduction when decreasing the values of P_{01} ; however, the results in Fig. 5 demonstrate that decreasing P_{01} could lead to

an increase in $J^*(s_0)$. To be specific, the minimum value of $J^*(s_0)$ is observed when $(P_{01}, P_{10}) = (0.9, 0.9)$, and $J^*(s_0)$ actually rises as P_{01} decreases from 0.9 to 0.1. Initially, this may seem counterintuitive, as one might expect that when P_{01} is small, the system will spend less time in the alarm state, resulting in a smaller cost $J^*(s_0)$. However, the underlying logic behind this phenomenon is that when P_{01} and P_{10} are large, the stochastic process spends only a limited number of time slots in each state. If the transmission success probability P_s is high, neither Δ_k^0 nor Δ_k^1 will reach significant values, resulting in low transition costs as defined by (20).

In Fig. 6, we illustrate the optimal policy μ^* for two scenarios and two states of the stochastic process. In the first scenario, energy is harvested with a high probability at each time slot ($P_e = 0.8$). In contrast, in the second scenario, P_e is set to a lower value of 0.4. In both experiments, the transmission success probability P_s was fixed at 0.8, the energy buffer capacity was set to 5, and the stochastic process' state transition probabilities P_z were defined as in (27). Specifically, Fig. 6a presents the optimal transmitter actions based on the number of energy units E_k stored in the energy buffer and the value of the AoI counter Δ_k^0 when $Z_k = 0$ and $P_e = 0.8$. Figure 6b shows the corresponding results for the case where the stochastic process is in the alarm state ($Z_k = 1$). Figures 6c and 6d present the corresponding results for the second scenario with $P_e = 0.4$.

Comparing Figures 6a and 6b, it is evident that when the probability of harvesting energy is high, the actions prescribed by the optimal policy exhibit minimal differences between the two states of the stochastic process. Specifically, the optimal policy μ^* still tends to reserve energy when $Z_k = 0$ by refraining from transmitting a status update ($a^* = 0$) when $(E_k, \Delta_k^0) \in \{(2, 1), (3, 1)\}$, representing the only distinction between the two cases. *The emphasis on energy reservation, anticipating alarm periods, becomes more pronounced when the probability of harvesting energy is lower.* Comparing Figures 6c and 6d, we observe that the optimal policy restrains the transmitter from sending status updates when $Z_k = 0$, even with a substantial number of energy units stored in the energy buffer. This strategy aims to avoid the quadratic cost associated with Z_k^1 . The transition probability values of the stochastic process further support the justification for this optimal policy. Matrix P_z indicates that once the stochastic process enters an alarm state, it will likely remain in that state ($P_{11} = 0.8$). *Therefore, reserving energy becomes essential to accommodate potentially extended periods during which the stochastic process remains in the alarm state.*

In Fig. 7, we have also examined the situation in which the AoI variable related to a state other than the current state of the stochastic process, i.e., Δ_k^z for $z \neq Z_k$, has a high value. Comparing Figs. 7a and 7b with Figs. 6c and 6d, we find that when Z_k and Z_k^d are identical, this AoI variable becomes irrelevant. However, in cases where $Z_k \neq Z_k^d$, it influences optimal actions and reduces the AoI thresholds at various energy levels. This is because both AoI variables increase concurrently, making it reasonable to transmit fresh updates when one of them has a high level, particularly when Δ_k^1 is elevated (as shown in the right figure in Fig. 7a).

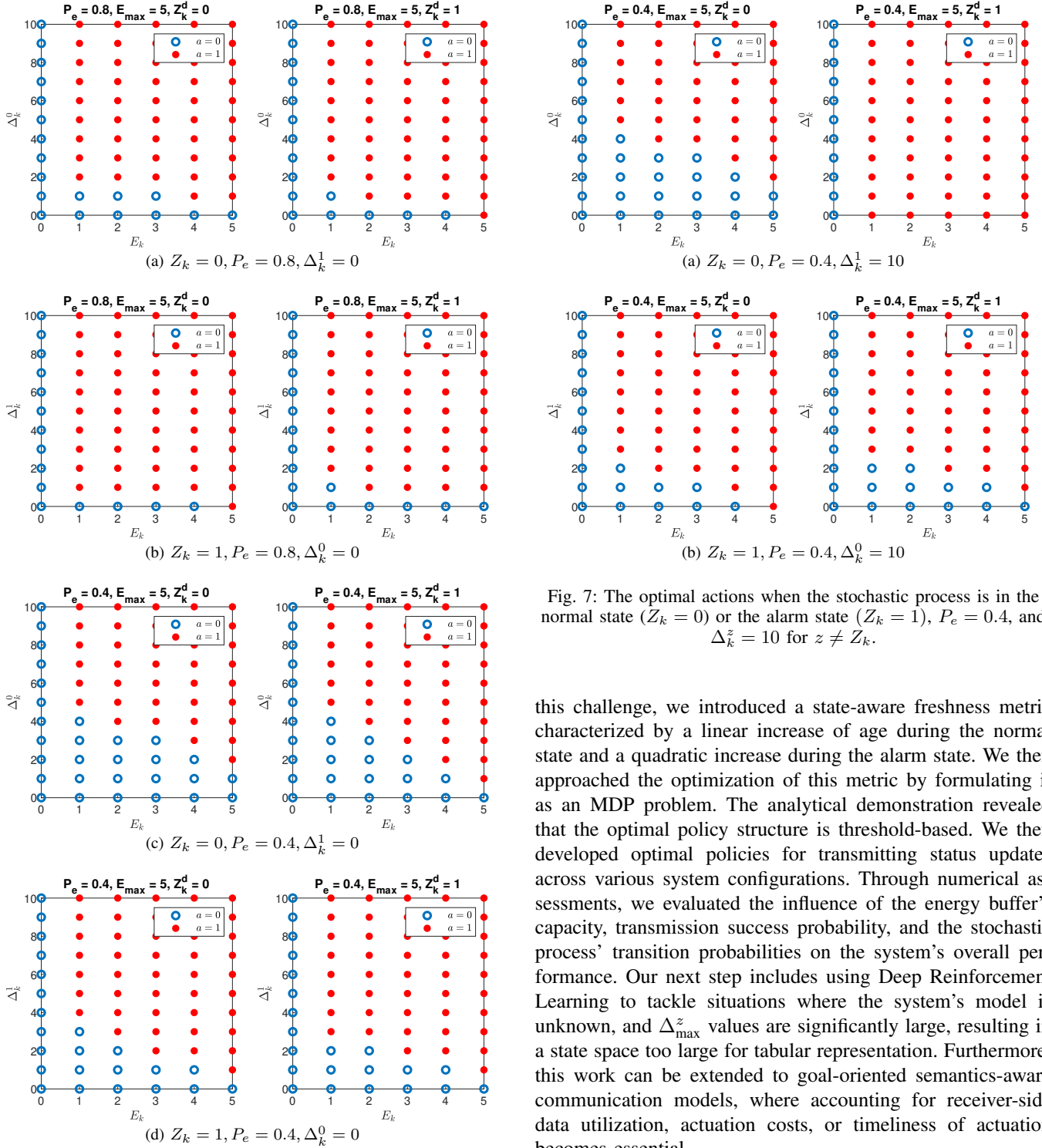


Fig. 6: The optimal actions when the stochastic process is in the normal state ($Z_k = 0$) or the alarm state ($Z_k = 1$) and P_e is either 0.8 or 0.4.

VI. CONCLUSIONS AND FUTURE WORK

This study examined a status update system incorporating an energy harvesting sensor to monitor a stochastic process. This process can exist in either a normal or an alarm operational state, each demanding different levels of timeliness. To address

Fig. 7: The optimal actions when the stochastic process is in the normal state ($Z_k = 0$) or the alarm state ($Z_k = 1$), $P_e = 0.4$, and $\Delta_k^z = 10$ for $z \neq Z_k$.

this challenge, we introduced a state-aware freshness metric characterized by a linear increase of age during the normal state and a quadratic increase during the alarm state. We then approached the optimization of this metric by formulating it as an MDP problem. The analytical demonstration revealed that the optimal policy structure is threshold-based. We then developed optimal policies for transmitting status updates across various system configurations. Through numerical assessments, we evaluated the influence of the energy buffer's capacity, transmission success probability, and the stochastic process' transition probabilities on the system's overall performance. Our next step includes using Deep Reinforcement Learning to tackle situations where the system's model is unknown, and Δ_{\max}^z values are significantly large, resulting in a state space too large for tabular representation. Furthermore, this work can be extended to goal-oriented semantics-aware communication models, where accounting for receiver-side data utilization, actuation costs, or timeliness of actuation becomes essential.

REFERENCES

- [1] G. Stamatakis, N. Pappas, and A. Traganitis, "Control of status updates for energy harvesting devices that monitor processes with alarms," in *IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.
- [2] 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, pp. 1183–1210, 2021.
- [3] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the internet of things," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, 2019.

- [4] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM*, March 2012.
- [5] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, November 2017.
- [6] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the wiener process for remote estimation over a channel with random delay," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, 2019.
- [7] G. Stamatakis, N. Pappas, and A. Traganitis, "Optimal policies for status update generation in an iot device with heterogeneous traffic," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5315–5328, 2020.
- [8] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *IEEE ISIT*, June 2017.
- [9] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks (JCN)*, vol. 21, no. 3, pp. 204–219, June 2019.
- [10] J. Zhong, R. D. Yates, and E. Soljanin, "Two freshness metrics for local cache refresh," in *IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1924–1928.
- [11] X. Zheng, S. Zhou, Z. Jiang, and Z. Niu, "Closed-form analysis of non-linear age of information in status updates with an energy harvesting transmitter," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4129–4142, 2019.
- [12] 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, pp. 2215–2228, 2020.
- [13] R. D. Yates, "The age of gossip in networks," in *IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2984–2989.
- [14] A. Nikkhah, A. Ephremides, and N. Pappas, "Age of actuation in a wireless power transfer system," in *IEEE INFOCOM Workshops*, 2023.
- [15] O. B. Akan, O. Cetinkaya, C. Koca, and M. Ozger, "Internet of hybrid energy harvesting things," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 736–746, 2017.
- [16] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *IEEE ISIT*, June 2015.
- [17] A. Arafat, S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *IEEE GLOBECOM*, December 2017.
- [18] A. Arafat, J. Yang, and S. Ulukus, "Age-minimal online policies for energy harvesting sensors with random battery recharges," in *IEEE ICC*, May 2018.
- [19] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 193–204, March 2018.
- [20] S. Farazi, A. G. Klein, and D. R. Brown, "Age of information in energy harvesting status update systems: When to preempt in service?" in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018.
- [21] S. Feng and J. Yang, "Minimizing age of information for an energy harvesting source with updating failures," in *IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2431–2435.
- [22] A. Arafat, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 534–556, 2019.
- [23] B. T. Bacinoglu, Y. Sun, E. Uysal, and V. Mutlu, "Optimal status updating with a finite-battery energy harvesting source," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 280–294, 2019.
- [24] S. Feng and J. Yang, "Age of information minimization for an energy harvesting source with updating erasures: Without and with feedback," *IEEE Transactions on Communications*, vol. 69, no. 8, pp. 5091–5105, 2021.
- [25] 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.
- [26] I. Krikidis, "Average age of information in wireless powered sensor networks," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 628–631, April 2019.
- [27] 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, pp. 456–470, 2021.
- [28] E. T. Ceran, D. Gündüz, and A. György, "Reinforcement learning to minimize age of information with an energy harvesting sensor with HARQ and sensing cost," in *IEEE INFOCOM Workshops*, April 2019.
- [29] S. Leng and A. Yener, "Minimizing age of information for an energy harvesting cognitive radio," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6.
- [30] C. Xu, X. Zhang, H. H. Yang, X. Wang, N. Pappas, D. Niyato, and T. Q. Quek, "Optimal status updates for minimizing age of correlated information in iot networks with energy harvesting sensors," *IEEE Transactions on Mobile Computing*, 2023.
- [31] S. Jayanth, N. Pappas, and R. V. Bhat, "Distortion minimization with age of information and cost constraints," in *21st International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2023.
- [32] Y. Dong, P. Fan, and K. B. Letaief, "Energy harvesting powered sensing in iot: Timeliness versus distortion," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10897–10911, 2020.
- [33] 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, pp. 1529–1540, 2021.
- [34] 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, pp. 7446–7463, 2022.
- [35] 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, pp. 8335–8351, 2021.
- [36] J. Holm, A. E. Kalør, F. Chiariotti, B. Soret, S. K. Jensen, T. B. Pedersen, and P. Popovski, "Freshness on demand: Optimizing age of information for the query process," in *IEEE International Conference on Communications*, 2021, pp. 1–6.
- [37] H. Hu, K. Xiong, H.-C. Yang, Q. Ni, B. Gao, P. Fan, and K. B. Letaief, "Aoi-minimal online scheduling for wireless powered iot: A lyapunov optimization-based approach," *IEEE Transactions on Green Communications and Networking*, 2023.
- [38] E. Delfani and N. Pappas, "Version age-optimal cached status updates in a gossiping network with energy harvesting sensor," in *21st International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2023.
- [39] P. Rafiee, Z. Ju, and M. Doroslovački, "Adaptive on/off scheduling to minimize age of information in an energy harvesting receiver," *IEEE Sensors Journal*, 2023.
- [40] D. Bertsekas, *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific Belmont, MA, 2012, vol. 2.

APPENDIX A PROOF OF THEOREM 1

Proof. Since $g(s, a) = g(s) = (1 - Z)\Delta^0 + Z(\Delta^1)^2$ is not a function of a , the Bellman's equation can be simplified as follows:

$$V(s) = g(s) + \arg \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} \gamma P(\tilde{s}|s, a) V(\tilde{s}) \right\}, \quad (28)$$

$$a^*(s) = \arg \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) V(\tilde{s}) \right\}. \quad (29)$$

We have dropped the asterisk superscript above V for the sake of simplicity. Let us define $V^1(s) = \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=1) V(\tilde{s})$, $V^0(s) = \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=0) V(\tilde{s})$, and $\Delta V(s) = V^1(s) - V^0(s)$. Thus, we have:

$$a^*(s) = \begin{cases} 0 & \Delta V(s) \geq 0, \\ 1 & \Delta V(s) < 0. \end{cases} \quad (30)$$

In what follows, we show that $\Delta V(s)$ is a decreasing function of (Δ^0, Δ^1) for each combination of (Z, Z^d, E) . Thus, $\Delta V(s)$ can become negative for sufficiently large values

of (Δ^0, Δ^1) , leading to the action $a = 1$ for $(\Delta^0, \Delta^1) \geq (\Delta_T^0, \Delta_T^1)$.

$$\begin{aligned} \Delta V(s) &= V^1(s) - V^0(s) \\ &= \sum_{\tilde{s} \in S} [P(\tilde{s}|s, a=1) - P(\tilde{s}|s, a=0)] V(\tilde{s}). \end{aligned} \quad (31)$$

When $E = 0$, then $\Delta V(s) = 0$ for each combination of $(Z, Z^d, \Delta^0, \Delta^1)$, so the action $a = 0$ is optimal. We therefore consider the cases where $E > 0$. The second term in (31) can be determined using the equations (9), (10), and (11), as presented in (32).

If the sensor decides against a transmission, then $W^s = 0$ with probability one, so we have:

$$\begin{aligned} P(\tilde{s}|s, a=0) &= \sum_{[W^e, W^z]} P[\tilde{s}|s, a=0, W^s=0, W^e, W^z] P[W^e] P[W^z|Z]. \end{aligned} \quad (34)$$

According to (12) - (15), $P[\tilde{s}|s, a=0, W^s=0, W^e, W^z] = 1$ for those $\tilde{s} \in S$ that hold all of the following conditions; it is 0 otherwise.

$$\tilde{Z} = W^z, \quad (35a)$$

$$\tilde{Z}^d = Z^d, \quad (35b)$$

$$\tilde{E} = E + W^e, \quad (35c)$$

$$\begin{cases} (\tilde{\Delta}^0 = 0, Z \neq 0, Z^d \neq 0) \\ \text{or } (\tilde{\Delta}^0 = \Delta^0 + 1, (Z = Z^d = 0 \text{ or } Z \neq Z^d)), \end{cases} \quad (35d)$$

$$\begin{cases} (\tilde{\Delta}^1 = 0, Z \neq 1, Z^d \neq 1) \\ \text{or } (\tilde{\Delta}^1 = \Delta^1 + 1, (Z = Z^d = 1 \text{ or } Z \neq Z^d)). \end{cases} \quad (35e)$$

We omitted the $\min\{\cdot, \Delta_{max}\}$ term due to space constraints, as it does not affect the proof of the theorem. The first term in (31) can also be simplified using the equations (9), (10), and (11), as presented in (33).

The summation (33a) is the same as (34), and is equal to 1 if the conditions (33) are satisfied; except that condition (35c) is replaced by $\tilde{E} = E + W^e - 1$. In addition, according to (12) - (15), $P[\tilde{s}|s, a=1, W^s=1, W^e, W^z]$ in (33b) will be equal to 1 for those $\tilde{s} \in S$ that hold all of the following conditions; it will be 0 otherwise.

$$\tilde{Z} = W^z, \quad (36a)$$

$$\tilde{Z}^d = Z, \quad (36b)$$

$$\tilde{E} = E + W^e - 1, \quad (36c)$$

$$(\tilde{\Delta}^0 = 0, Z \neq 0) \text{ or } (\tilde{\Delta}^0 = 1, Z = 0), \quad (36d)$$

$$(\tilde{\Delta}^1 = 0, Z \neq 1) \text{ or } (\tilde{\Delta}^1 = 1, Z = 1). \quad (36e)$$

Now, we can write $\Delta V(s)$ using the equation (31) for different values of Z and Z^d .

Case 1. $Z = 0$ and $Z^d = 0$. In this case, we have:

$$\begin{aligned} P[\tilde{s}|s, a=0, W^s=0, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (37)$$

$$\begin{aligned} P[\tilde{s}|s, a=1, W^s=0, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (38)$$

$$\begin{aligned} P[\tilde{s}|s, a=1, W^s=1, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = 1, \tilde{\Delta}^1 = 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (39)$$

then we have $\Delta V(s) = V^1(s) - V^0(s)$, where:

$$V^0(s) = \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=0) V(\tilde{s}) \quad (40)$$

$$\begin{aligned} &= \sum_{\tilde{s} \in S} \sum_{[W^e, W^z]} P[\tilde{s}|s, a=0, W^s=0, W^e, W^z] P[W^e] P[W^z|Z] V(\tilde{s}) \\ &= \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e, \Delta^0 + 1, 0) P[W^e] P[W^z|Z], \end{aligned} \quad (41)$$

$$\begin{aligned} V^1(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=1) V(\tilde{s}) \\ &= (1 - P_s) \sum_{\tilde{s} \in S} \sum_{[W^e, W^z]} P[\tilde{s}|s, a=1, W^s=0, W^e, W^z] P[W^e] P[W^z|Z] V(\tilde{s}) \\ &\quad + P_s \sum_{\tilde{s} \in S} \sum_{[W^e, W^z]} P[\tilde{s}|s, a=1, W^s=1, W^e, W^z] P[W^e] P[W^z|Z] V(\tilde{s}) \\ &= (1 - P_s) \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e - 1, \Delta^0 + 1, 0) P[W^e] P[W^z|Z] \\ &\quad + P_s \sum_{[W^e, W^z]} V(W^z, Z, E + W^e - 1, 1, 0) P[W^e] P[W^z|Z]. \end{aligned} \quad (42)$$

Case 2. $Z = 0$ and $Z^d = 1$. In this case, we have:

$$\begin{aligned} P[\tilde{s}|s, a=0, W^s=0, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (43)$$

$$\begin{aligned} P[\tilde{s}|s, a=1, W^s=0, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (44)$$

$$\begin{aligned} P[\tilde{s}|s, a=1, W^s=1, W^e, W^z] &= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = 1, \tilde{\Delta}^1 = 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (45)$$

then we have $\Delta V(s) = V^1(s) - V^0(s)$, where:

$$\begin{aligned} V^0(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=0) V(\tilde{s}) \\ &= \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e, \Delta^0 + 1, \Delta^1 + 1) P[W^e] P[W^z|Z], \end{aligned} \quad (46)$$

$$\begin{aligned} V^1(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=1) V(\tilde{s}) \\ &= (1 - P_s) \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e - 1, \Delta^0 + 1, \Delta^1 + 1) P[W^e] P[W^z|Z] \\ &\quad + P_s \sum_{[W^e, W^z]} V(W^z, Z, E + W^e - 1, 1, 0) P[W^e] P[W^z|Z]. \end{aligned} \quad (47)$$

$$P(\tilde{s}|s, a = 0) = \sum_{[W^s, W^e, W^z]} P[\tilde{s}|s, a = 0, W^s, W^e, W^z] P[W^s|E, a = 0] P[W^e] P[W^z|Z]. \quad (32)$$

$$\begin{aligned} P(\tilde{s}|s, a = 1) &= \sum_{[W^s, W^e, W^z]} P[\tilde{s}|s, a = 1, W^s, W^e, W^z] P[W^s|E, a = 1] P[W^e] P[W^z|Z] \\ &= \sum_{[W^e, W^z]} P[\tilde{s}|s, a = 1, W^s = 0, W^e, W^z] P[W^s = 0|E, a = 1] P[W^e] P[W^z|Z] \\ &\quad + \sum_{[W^e, W^z]} P[\tilde{s}|s, a = 1, W^s = 1, W^e, W^z] P[W^s = 1|E, a = 1] P[W^e] P[W^z|Z] \\ &= (1 - P_s) \sum_{[W^e, W^z]} P[\tilde{s}|s, a = 1, W^s = 0, W^e, W^z] P[W^e] P[W^z|Z] \quad (33a) \\ &\quad + P_s \sum_{[W^e, W^z]} P[\tilde{s}|s, a = 1, W^s = 1, W^e, W^z] P[W^e] P[W^z|Z]. \quad (33b) \end{aligned}$$

Case 3. $Z = 1$ and $Z^d = 0$. In this case, we have:

$$\begin{aligned} P[\tilde{s}|s, a = 0, W^s = 0, W^e, W^z] \quad (48) \\ = \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

$$\begin{aligned} P[\tilde{s}|s, a = 1, W^s = 0, W^e, W^z] \quad (49) \\ = \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = \Delta^0 + 1, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

$$\begin{aligned} P[\tilde{s}|s, a = 1, W^s = 1, W^e, W^z] \quad (50) \\ = \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = 0, \tilde{\Delta}^1 = 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

then we have $\Delta V(s) = V^1(s) - V^0(s)$, where:

$$\begin{aligned} V^0(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a = 0) V(\tilde{s}) \quad (51) \\ &= \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e, \Delta^0 + 1, \Delta^1 + 1) P[W^e] P[W^z|Z], \end{aligned}$$

$$\begin{aligned} V^1(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a = 1) V(\tilde{s}) \quad (52) \\ &= (1 - P_s) \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e - 1, \Delta^0 + 1, \Delta^1 + 1) P[W^e] P[W^z|Z] \\ &\quad + P_s \sum_{[W^e, W^z]} V(W^z, Z, E + W^e - 1, 0, 1) P[W^e] P[W^z|Z]. \end{aligned}$$

Case 4. $Z = 1$ and $Z^d = 1$. In this case, we have:

$$\begin{aligned} P[\tilde{s}|s, a = 0, W^s = 0, W^e, W^z] \quad (53) \\ = \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e, \\ & \tilde{\Delta}^0 = 0, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

$$P[\tilde{s}|s, a = 1, W^s = 0, W^e, W^z] \quad (54)$$

$$= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z^d, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = 0, \tilde{\Delta}^1 = \Delta^1 + 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$P[\tilde{s}|s, a = 1, W^s = 1, W^e, W^z] \quad (55)$$

$$= \begin{cases} 1 & \tilde{Z} = W^z, \tilde{Z}^d = Z, \tilde{E} = E + W^e - 1, \\ & \tilde{\Delta}^0 = 0, \tilde{\Delta}^1 = 1, \\ 0 & \text{otherwise,} \end{cases}$$

then we have $\Delta V(s) = V^1(s) - V^0(s)$, where:

$$\begin{aligned} V^0(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a = 0) V(\tilde{s}) \quad (56) \\ &= \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e, 0, \Delta^1 + 1) P[W^e] P[W^z|Z], \end{aligned}$$

$$\begin{aligned} V^1(s) &= \sum_{\tilde{s} \in S} P(\tilde{s}|s, a = 1) V(\tilde{s}) \quad (57) \\ &= (1 - P_s) \sum_{[W^e, W^z]} V(W^z, Z^d, E + W^e - 1, 0, \Delta^1 + 1) P[W^e] P[W^z|Z] \\ &\quad + P_s \sum_{[W^e, W^z]} V(W^z, Z, E + W^e - 1, 0, 1) P[W^e] P[W^z|Z]. \end{aligned}$$

We will proceed with the theorem's proof for case 3, as the proof for other cases follows a similar approach. We aim to show that $\Delta V(s)$ is decreasing in (Δ^0, Δ^1) for each (Z, Z^d, E) .

$$\Delta V(s) = V^1(s) - V^0(s) \quad (58)$$

$$\begin{aligned} &= \sum_{[W^e, W^z]} \left\{ (1 - P_s) V(W^z, Z^d, E + W^e - 1, \Delta^0 + 1, \Delta^1 + 1) \right. \\ &\quad \left. - V(W^z, Z^d, E + W^e, \Delta^0 + 1, \Delta^1 + 1) \right. \\ &\quad \left. + P_s V(W^z, Z, E + W^e - 1, 0, 1) \right\} P[W^e] P[W^z|Z]. \quad (59) \end{aligned}$$

Let us define $s^+ = [Z, Z^d, E, \Delta^{0+}, \Delta^{1+}]^T$ and $s^- = [Z, Z^d, E, \Delta^{0-}, \Delta^{1-}]^T$ such that $(\Delta^{0+}, \Delta^{1+}) \geq (\Delta^{0-}, \Delta^{1-})$, element-wise. We will therefore prove that $\Delta V(s^+) \leq \Delta V(s^-)$.

$$\begin{aligned}
\Delta V(s^+) &\leq \Delta V(s^-) \quad (60) \\
\Leftrightarrow \sum_{[W^e, W^z]} &\left\{ (1-P_s)V(W^z, Z^d, E+W^e-1, \Delta^{0+}+1, \Delta^{1+}+1) \right. \\
&\quad - V(W^z, Z^d, E+W^e, \Delta^{0+}+1, \Delta^{1+}+1) \\
&\quad \left. + P_s V(W^z, Z, E+W^e-1, 0, 1) \right\} P[W^e] P[W^z|Z] \\
&\leq \sum_{[W^e, W^z]} \left\{ (1-P_s)V(W^z, Z^d, E+W^e-1, \Delta^{0-}+1, \Delta^{1-}+1) \right. \\
&\quad - V(W^z, Z^d, E+W^e, \Delta^{0-}+1, \Delta^{1-}+1) \\
&\quad \left. + P_s V(W^z, Z, E+W^e-1, 0, 1) \right\} P[W^e] P[W^z|Z] \\
&\Leftrightarrow (1-P_s)V(W^z, Z^d, E+W^e-1, \Delta^{0+}+1, \Delta^{1+}+1) \\
&\quad - V(W^z, Z^d, E+W^e, \Delta^{0+}+1, \Delta^{1+}+1) \\
&\leq (1-P_s)V(W^z, Z^d, E+W^e-1, \Delta^{0-}+1, \Delta^{1-}+1) \\
&\quad - V(W^z, Z^d, E+W^e, \Delta^{0-}+1, \Delta^{1-}+1) \\
&\Leftrightarrow (1-P_s) \left[V(W^z, Z^d, E+W^e-1, \Delta^{0+}+1, \Delta^{1+}+1) \right. \\
&\quad \left. - V(W^z, Z^d, E+W^e-1, \Delta^{0-}+1, \Delta^{1-}+1) \right] \\
&\leq V(W^z, Z^d, E+W^e, \Delta^{0+}+1, \Delta^{1+}+1) \\
&\quad - V(W^z, Z^d, E+W^e, \Delta^{0-}+1, \Delta^{1-}+1). \quad (61)
\end{aligned}$$

In the following Lemma, we demonstrate the last inequality which concludes the proof of Theorem 1. \square

APPENDIX B LEMMA 1

Lemma 1. Suppose that $s_E^+ = [Z, Z^d, E, \Delta^{0+}, \Delta^{0+}]^T$, $s_E^- = [Z, Z^d, E, \Delta^{0-}, \Delta^{0-}]^T$, $s_{E-1}^+ = [Z, Z^d, E-1, \Delta^{0+}, \Delta^{0+}]^T$, and $s_{E-1}^- = [Z, Z^d, E-1, \Delta^{0-}, \Delta^{0-}]^T$ are four states such that $E \in \{1, 2, 3, \dots\}$ and $(\Delta^{0+}, \Delta^{1+}) \geq (\Delta^{0-}, \Delta^{1-})$; then the value function satisfies the following inequality:

$$(1-P_s) [V(s_{E-1}^+) - V(s_{E-1}^-)] \leq V(s_E^+) - V(s_E^-). \quad (62)$$

Proof. We employ the Value Iteration Algorithm (VIA) to prove the lemma. In each iteration at time step k , the value function is updated as follows:

$$\begin{aligned}
V_k(s) &= \min_{a \in \{0,1\}} \left\{ \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) [g(s, a) + \gamma V_{k-1}(\tilde{s})] \right\} \\
&= g(s) + \min_{a \in \{0,1\}} \left\{ \gamma \sum_{\tilde{s} \in S} P(\tilde{s}|s, a) V_{k-1}(\tilde{s}) \right\}. \quad (63)
\end{aligned}$$

VIA converges to the value function of the Bellman's equation irrespective of the initial value assigned to $V_0(s)$, i.e., $\lim_{k \rightarrow \infty} V_k(s) = V(s) \forall s \in S$. Therefore, it suffices to establish the following:

$$(1-P_s) [V_k(s_{E-1}^+) - V_k(s_{E-1}^-)] \leq V_k(s_E^+) - V_k(s_E^-), \quad \forall k = 0, 1, 2, \dots \quad (64)$$

We utilize mathematical induction to proceed the proof. Assuming $V_0(s) = 0$ for all $s \in S$, (64) holds true for $k = 0$. Now, with the same assumption extending up to $k > 0$, we prove its validity for $k + 1$, i.e.:

$$\begin{aligned}
(1-P_s) [V_{k+1}(s_{E-1}^+) - V_{k+1}(s_{E-1}^-)] &\leq V_{k+1}(s_E^+) - V_{k+1}(s_E^-) \\
\Leftrightarrow (1-P_s) [V_{k+1}(s_{E-1}^+) - V_{k+1}(s_{E-1}^-)] & \\
- [V_{k+1}(s_E^+) - V_{k+1}(s_E^-)] &\leq 0. \quad (65)
\end{aligned}$$

Let us define $V_{k+1}^0(s)$ and $V_{k+1}^1(s)$ as follows:

$$V_{k+1}^0(s) = g(s) + \gamma \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=0) V_k(\tilde{s}) = g(s) \quad (66)$$

$$+ \gamma \sum_{[W^e, W^z]} V_k(W^z, Z^d, E+W^e, \Delta^{0+}+1, \Delta^{1+}+1) P[W^e] P[W^z|Z],$$

$$\begin{aligned}
V_{k+1}^1(s) &= g(s) + \gamma \sum_{\tilde{s} \in S} P(\tilde{s}|s, a=1) V_k(\tilde{s}) = g(s) \\
&+ \gamma (1-P_s) \sum_{[W^e, W^z]} V_k(W^z, Z^d, E+W^e-1, \Delta^{0+}+1, \Delta^{1+}+1) P[W^e] P[W^z|Z] \\
&+ \gamma P_s \sum_{[W^e, W^z]} V_k(W^z, Z, E+W^e-1, 0, 1) P[W^e] P[W^z|Z], \quad (67)
\end{aligned}$$

then we have $V_{k+1}(s) = \min \{V_{k+1}^0(s), V_{k+1}^1(s)\}$, according to VIA iteration (63) at time slot $k + 1$. Thus, (65) can be rewritten as follows:

$$\begin{aligned}
(1-P_s) &\left[\min \{V_{k+1}^0(s_{E-1}^+), V_{k+1}^1(s_{E-1}^+)\} \right. \\
&\quad \left. - \min \{V_{k+1}^0(s_{E-1}^-), V_{k+1}^1(s_{E-1}^-)\} \right] \\
&- \left[\min \{V_{k+1}^0(s_E^+), V_{k+1}^1(s_E^+)\} \right. \\
&\quad \left. - \min \{V_{k+1}^0(s_E^-), V_{k+1}^1(s_E^-)\} \right] \leq 0. \quad (68)
\end{aligned}$$

Now, we consider four cases.

Case 1. $V_{k+1}^0(s_{E-1}^-) \leq V_{k+1}^1(s_{E-1}^-)$ and $V_{k+1}^0(s_E^+) \leq V_{k+1}^1(s_E^+)$. In this case, equation (68) is simplified to:

$$\begin{aligned}
(1-P_s) &[\min \{V_{k+1}^0(s_{E-1}^+), V_{k+1}^1(s_{E-1}^+)\} - V_{k+1}^0(s_{E-1}^-)] \\
&- [V_{k+1}^0(s_E^+) - \min \{V_{k+1}^0(s_E^-), V_{k+1}^1(s_E^-)\}] \leq 0. \quad (69)
\end{aligned}$$

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

$$\begin{aligned}
(1-P_s) &[V_{k+1}^0(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^-)] \\
&\quad \underbrace{\leq 0}_{\leq 0} \\
&+ \underbrace{(1-P_s) \min \{0, V_{k+1}^1(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^+)\}}_{\leq 0} \\
&- [V_{k+1}^0(s_E^+) - V_{k+1}^0(s_E^-)] \\
&+ \underbrace{\min \{0, V_{k+1}^1(s_E^-) - V_{k+1}^0(s_E^-)\}}_{\leq 0} \leq 0, \quad (70)
\end{aligned}$$

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

$$\begin{aligned}
(1-P_s) &[V_{k+1}^0(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^-)] \\
&- [V_{k+1}^0(s_E^+) - V_{k+1}^0(s_E^-)] \leq 0. \quad (71)
\end{aligned}$$

According to (66), we have:

$$\begin{aligned}
 & (1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] \\
 & + \gamma \sum_{\tilde{s} \in S} \left[P(\tilde{s}|s_{E-1}^+, a=0) - P(\tilde{s}|s_{E-1}^-, a=0) \right] V_k(\tilde{s}) \\
 & - \left[g(s_E^+) - g(s_E^-) \right] \\
 & + \gamma \sum_{\tilde{s} \in S} \left[P(\tilde{s}|s_E^+, a=0) - P(\tilde{s}|s_E^-, a=0) \right] V_k(\tilde{s}) \leq 0 \\
 & \Leftrightarrow (1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] - \left[g(s_E^+) - g(s_E^-) \right] \\
 & + \gamma \sum_{\tilde{s} \in S} \left\{ (1 - P_s) \left[P(\tilde{s}|s_{E-1}^+, a=0) - P(\tilde{s}|s_{E-1}^-, a=0) \right] \right. \\
 & \left. - \left[P(\tilde{s}|s_E^+, a=0) - P(\tilde{s}|s_E^-, a=0) \right] \right\} V_k(\tilde{s}) \leq 0. \quad (72)
 \end{aligned}$$

We know that $g(s_{E-1}^+) = g(s_E^+) = (1 - Z)\Delta^{0+} + Z(\Delta^{1+})^2$, $g(s_{E-1}^-) = g(s_E^-) = (1 - Z)\Delta^{0-} + Z(\Delta^{1-})^2$. Additionally, we have $g(s_{E-1}^+) = g(s_E^+) \geq g(s_{E-1}^-) = g(s_E^-)$ since $\Delta^{0+} \geq \Delta^{0-}$ and $\Delta^{1+} \geq \Delta^{1-}$. Therefore, $(1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] - \left[g(s_E^+) - g(s_E^-) \right] = -P_s \left[g(s_E^+) - g(s_E^-) \right] \leq 0$. Now, we prove that the summation in (72) is also negative. Simplifying this summation based on equation (66) results in the following expression:

$$\begin{aligned}
 & \sum_{[W^e, W^z]} \left\{ (1 - P_s) \left[V_k(W^z, Z^d, E + W^e - 1, \Delta^{0+} + 1, \Delta^{1+} + 1) \right. \right. \\
 & \left. - V_k(W^z, Z^d, E + W^e - 1, \Delta^{0-} + 1, \Delta^{1-} + 1) \right] \\
 & - \left[V_k(W^z, Z^d, E + W^e, \Delta^{0+} + 1, \Delta^{1+} + 1) \right. \\
 & \left. - V_k(W^z, Z^d, E + W^e, \Delta^{0-} + 1, \Delta^{1-} + 1) \right] \Big\} P[W^e] P[W^z|Z] \leq 0. \quad (73)
 \end{aligned}$$

Let us define $\tilde{s}_E^+ = \tilde{s}_E^+(W^e, W^z) = [W^z, Z^d, E + W^e, \Delta^{0+} + 1, \Delta^{1+} + 1]^T$ and $\tilde{s}_E^- = \tilde{s}_E^-(W^e, W^z) = [W^z, Z^d, E + W^e, \Delta^{0-} + 1, \Delta^{1-} + 1]^T$, then (73) can be rewritten:

$$\begin{aligned}
 & \sum_{[W^e, W^z]} \left\{ (1 - P_s) \left[V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-) \right] \right. \\
 & \left. - \left[V_k(\tilde{s}_E^+) - V_k(\tilde{s}_E^-) \right] \right\} P[W^e] P[W^z|Z] \leq 0. \quad (74)
 \end{aligned}$$

In accordance with the assumption stated in equation (64), $(1 - P_s) \left[V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-) \right] - \left[V_k(\tilde{s}_E^+) - V_k(\tilde{s}_E^-) \right] \leq 0$. As a result, the summation (73) is negative (non-positive). This observation concludes the proof for case 1.

Case 2. $V_{k+1}^0(s_{E-1}^-) \leq V_{k+1}^1(s_{E-1}^-)$ and $V_{k+1}^0(s_E^+) > V_{k+1}^1(s_E^+)$. In this case, equation (68) is reduced to:

$$\begin{aligned}
 & (1 - P_s) \left[\min \{ V_{k+1}^0(s_{E-1}^+), V_{k+1}^1(s_{E-1}^+) \} - V_{k+1}^0(s_{E-1}^-) \right] \\
 & - \left[V_{k+1}^1(s_E^+) - \min \{ V_{k+1}^0(s_E^-), V_{k+1}^1(s_E^-) \} \right] \leq 0 \\
 & \Leftrightarrow (1 - P_s) \left[V_{k+1}^0(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^-) \right] \\
 & + \overbrace{(1 - P_s) \min \{ 0, V_{k+1}^1(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^+) \}}^{\leq 0} \\
 & - \left[V_{k+1}^1(s_E^+) - V_{k+1}^1(s_E^-) \right] + \underbrace{\min \{ V_{k+1}^0(s_E^-) - V_{k+1}^1(s_E^-), 0 \}}_{\leq 0} \leq 0. \quad (75)
 \end{aligned}$$

It is adequate to show that:

$$\begin{aligned}
 & (1 - P_s) \left[V_{k+1}^0(s_{E-1}^+) - V_{k+1}^0(s_{E-1}^-) \right] \\
 & - \left[V_{k+1}^1(s_E^+) - V_{k+1}^1(s_E^-) \right] \leq 0. \quad (76)
 \end{aligned}$$

According to (66) and (67), we have:

$$\begin{aligned}
 & (1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] \\
 & + \gamma \sum_{\tilde{s} \in S} \left[P(\tilde{s}|s_{E-1}^+, a=0) - P(\tilde{s}|s_{E-1}^-, a=0) \right] V_k(\tilde{s}) \\
 & - \left[g(s_E^+) - g(s_E^-) \right] \\
 & + \gamma \sum_{\tilde{s} \in S} \left[P(\tilde{s}|s_E^+, a=1) - P(\tilde{s}|s_E^-, a=1) \right] V_k(\tilde{s}) \leq 0 \\
 & \Leftrightarrow \overbrace{(1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] - \left[g(s_E^+) - g(s_E^-) \right]}^{=-P_s[g(s_E^+) - g(s_E^-)] \leq 0} \\
 & + \gamma \sum_{\tilde{s} \in S} \left\{ (1 - P_s) \left[P(\tilde{s}|s_{E-1}^+, a=0) - P(\tilde{s}|s_{E-1}^-, a=0) \right] \right. \\
 & \left. - \left[P(\tilde{s}|s_E^+, a=1) - P(\tilde{s}|s_E^-, a=1) \right] \right\} V_k(\tilde{s}) \leq 0. \quad (77)
 \end{aligned}$$

We prove that the summation in (77) is also non-positive. Rewriting this sum using the equations (66) and (67) yields the subsequent equation:

$$\begin{aligned}
 & \sum_{[W^e, W^z]} (1 - P_s) \left\{ \left[V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-) \right] \right. \\
 & \left. - \left[V_k(\tilde{s}_E^+) - V_k(\tilde{s}_E^-) \right] \right\} P[W^e] P[W^z|Z] = 0, \quad (78)
 \end{aligned}$$

and it concludes the proof for case 2.

Case 3. $V_{k+1}^0(s_{E-1}^-) > V_{k+1}^1(s_{E-1}^-)$ and $V_{k+1}^0(s_E^+) \leq V_{k+1}^1(s_E^+)$. In this case, equation (68) is simplified to:

$$\begin{aligned}
 & (1 - P_s) \left[\min \{ V_{k+1}^0(s_{E-1}^+), V_{k+1}^1(s_{E-1}^+) \} - V_{k+1}^1(s_{E-1}^-) \right] \\
 & - \left[V_{k+1}^0(s_E^+) - \min \{ V_{k+1}^0(s_E^-), V_{k+1}^1(s_E^-) \} \right] \leq 0 \\
 & \Leftrightarrow (1 - P_s) \left[V_{k+1}^1(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^-) \right] \\
 & + \overbrace{(1 - P_s) \min \{ V_{k+1}^0(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^+), 0 \}}^{\leq 0} \\
 & - \left[V_{k+1}^0(s_E^+) - V_{k+1}^0(s_E^-) \right] + \underbrace{\min \{ 0, V_{k+1}^1(s_E^-) - V_{k+1}^0(s_E^-) \}}_{\leq 0} \leq 0, \quad (79)
 \end{aligned}$$

It is adequate to demonstrate that:

$$\begin{aligned}
 & (1 - P_s) \left[V_{k+1}^1(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^-) \right] \\
 & - \left[V_{k+1}^0(s_E^+) - V_{k+1}^0(s_E^-) \right] \leq 0. \quad (80)
 \end{aligned}$$

¹It is noteworthy that this case does not occur when $E = 1$, as the action $a = 0$ is optimal, and $V_{k+1}^0(s_{E-1}^-) = V_{k+1}^1(s_{E-1}^-)$.

$$\begin{aligned}
& \sum_{[W^e, W^z]} \left\{ (1 - P_s)^2 [V_k(\tilde{s}_{E-2}^+) - V_k(\tilde{s}_{E-2}^-)] - [V_k(\tilde{s}_E^+) - V_k(\tilde{s}_E^-)] \right\} P[W^e] P[W^z|Z] \\
&= \sum_{[W^e, W^z]} \left\{ \underbrace{(1 - P_s)^2 [V_k(\tilde{s}_{E-2}^+) - V_k(\tilde{s}_{E-2}^-)] - (1 - P_s) [V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-)]}_{(a)} \right. \\
&\quad \left. + \underbrace{(1 - P_s) [V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-)] - [V_k(\tilde{s}_E^+) - V_k(\tilde{s}_E^-)]}_{(b)} \right\} P[W^e] P[W^z|Z], \tag{82}
\end{aligned}$$

According to (66) and (67):

$$\begin{aligned}
& (1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] \\
&+ \gamma \sum_{\tilde{s} \in S} [P(\tilde{s}|s_{E-1}^+, a=1) - P(\tilde{s}|s_{E-1}^-, a=1)] V_k(\tilde{s}) \\
&- \left[g(s_E^+) - g(s_E^-) \right] \\
&+ \gamma \sum_{\tilde{s} \in S} [P(\tilde{s}|s_E^+, a=0) - P(\tilde{s}|s_E^-, a=0)] V_k(\tilde{s}) \leq 0 \\
&\quad = -P_s [g(s_E^+) - g(s_E^-)] \leq 0 \\
&\Leftrightarrow \underbrace{(1 - P_s) [g(s_{E-1}^+) - g(s_{E-1}^-)] - [g(s_E^+) - g(s_E^-)]}_{(a)} \\
&+ \gamma \sum_{\tilde{s} \in S} \left\{ (1 - P_s) [P(\tilde{s}|s_{E-1}^+, a=1) - P(\tilde{s}|s_{E-1}^-, a=1)] \right. \\
&\quad \left. - [P(\tilde{s}|s_E^+, a=0) - P(\tilde{s}|s_E^-, a=0)] \right\} V_k(\tilde{s}) \leq 0. \tag{81}
\end{aligned}$$

We prove that the sum in (81) is likewise non-positive. By simplifying this sum using equations (66) and (67), we arrive at the expression (82), where both (a) and (b) are negative according to (64), thus concluding the proof for case 3.

$$\begin{aligned}
(a) &= (1 - P_s) \left\{ (1 - P_s) [V_k(\tilde{s}_{E-2}^+) - V_k(\tilde{s}_{E-2}^-)] \right. \\
&\quad \left. - [V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-)] \right\} \\
&\leq (1 - P_s) [V_k(\tilde{s}_{E-2}^+) - V_k(\tilde{s}_{E-2}^-)] - [V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-)] \\
&\leq 0. \tag{83}
\end{aligned}$$

Case 4. $V_{k+1}^0(s_{E-1}^-) > V_{k+1}^1(s_{E-1}^-)$ and $V_{k+1}^0(s_E^+) > V_{k+1}^1(s_E^+)$. In this case, equation (68) is reduced to:

$$\begin{aligned}
& (1 - P_s) [\min \{V_{k+1}^0(s_{E-1}^+), V_{k+1}^1(s_{E-1}^+)\} - V_{k+1}^1(s_{E-1}^-)] \\
&- [V_{k+1}^1(s_E^+) - \min \{V_{k+1}^0(s_E^-), V_{k+1}^1(s_E^-)\}] \leq 0 \\
&\Leftrightarrow (1 - P_s) [V_{k+1}^1(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^-)] \\
&\quad \leq 0 \\
&+ \underbrace{(1 - P_s) \min \{V_{k+1}^0(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^+), 0\}}_{\leq 0} \\
&- [V_{k+1}^1(s_E^+) - V_{k+1}^1(s_E^-)] + \underbrace{\min \{V_{k+1}^0(s_E^-) - V_{k+1}^1(s_E^-), 0\}}_{\leq 0} \leq 0. \tag{84}
\end{aligned}$$

It suffices to demonstrate that:

$$\begin{aligned}
& (1 - P_s) [V_{k+1}^1(s_{E-1}^+) - V_{k+1}^1(s_{E-1}^-)] \\
&- [V_{k+1}^1(s_E^+) - V_{k+1}^1(s_E^-)] \leq 0. \tag{85}
\end{aligned}$$

²It is noteworthy that this case does not occur when $E = 1$, as the action $a = 0$ is optimal, and $V_{k+1}^0(s_{E-1}^-) = V_{k+1}^1(s_{E-1}^-)$.

According to (67):

$$\begin{aligned}
& (1 - P_s) \left[g(s_{E-1}^+) - g(s_{E-1}^-) \right] \\
&+ \gamma \sum_{\tilde{s} \in S} [P(\tilde{s}|s_{E-1}^+, a=1) - P(\tilde{s}|s_{E-1}^-, a=1)] V_k(\tilde{s}) \\
&- \left[g(s_E^+) - g(s_E^-) \right] \\
&+ \gamma \sum_{\tilde{s} \in S} [P(\tilde{s}|s_E^+, a=1) - P(\tilde{s}|s_E^-, a=1)] V_k(\tilde{s}) \leq 0 \\
&\quad = -P_s [g(s_E^+) - g(s_E^-)] \leq 0 \\
&\Leftrightarrow \underbrace{(1 - P_s) [g(s_{E-1}^+) - g(s_{E-1}^-)] - [g(s_E^+) - g(s_E^-)]}_{(a)} \\
&+ \gamma \sum_{\tilde{s} \in S} \left\{ (1 - P_s) [P(\tilde{s}|s_{E-1}^+, a=1) - P(\tilde{s}|s_{E-1}^-, a=1)] \right. \\
&\quad \left. - [P(\tilde{s}|s_E^+, a=1) - P(\tilde{s}|s_E^-, a=1)] \right\} V_k(\tilde{s}) \leq 0. \tag{86}
\end{aligned}$$

We demonstrate that the sum in (86) is also non-positive. Simplifying the summation using the equation (67) yields the subsequent expression:

$$\begin{aligned}
& \sum_{[W^e, W^z]} \left\{ (1 - P_s)^2 [V_k(\tilde{s}_{E-2}^+) - V_k(\tilde{s}_{E-2}^-)] \right. \\
&\quad \left. - (1 - P_s) [V_k(\tilde{s}_{E-1}^+) - V_k(\tilde{s}_{E-1}^-)] \right\} P[W^e] P[W^z|Z] \leq 0, \tag{87}
\end{aligned}$$

and the proof for case 4 and Lemma 1 is completed. \square