# ABSTRACT

Title of Dissertation: AGE OF INCORRECT INFORMATION: A NEW PERFORMANCE METRIC IN SEMANTIC COMMUNICATIONS

Yutao Chen
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Anthony Ephremides
Department of Electrical and Computer Engineering

With the increasing popularity of smart devices and the rapid development of networking and communication technologies, cyber-physical system applications have been widely deployed and are receiving increasing attention. Some examples of these systems include vehicle networks, where vehicles collect real-time external information through their on-board sensors and cameras to generate a reliable description of the surroundings; intelligent transportation systems, where real-time monitoring of road conditions and traffic congestion is essential; and natural or man-made disaster prevention and management, where real-time monitoring of omens and disaster propagation is crucial. A common feature of these systems is the high requirement for the timeliness of the acquired information, which has led to the development of optimization frameworks aimed at capturing information freshness. Age of Information (AoI) is a prime example, but it has the drawback of only considering information freshness and ignoring the importance of content. As a result, the Age of Incorrect Information (AoII) has been developed to capture both the

freshness and content of information. In this dissertation, we study the fundamental nature and optimization of AoII in numerous systems.

With the proliferation of smart devices, energy consumption has become a major concern. In the first part, we focus on the characteristics and performance of AoII under limited resources. In particular, we propose an efficient algorithm to obtain the AoII-optimal policy under resource-constrained conditions and compute the performance of the optimal policies.

The massive connectivity of communication systems has made scheduling a hot research topic. In the second part, we analyze and optimize the performance of AoII in the scheduling problem. We present the Whittle's index policy for AoII, whose superior performance has been recognized in many other problems. However, it also has limitations. Therefore, we propose a new scheduling policy, the indexed priority policy, which has comparable performance to the Whittle's index policy but has broader applicability.

With the unprecedented increase in the amount and types of data to be transmitted and the impact of external factors such as urban construction, data transmission will experience numerous uncertainties. Therefore, in the third part, we study the characteristics and optimization of AoII in an environment with random delays. Specifically, in the first half, we consider the case where the communication channel suffers from a random delay. In the second half, we build on the first half and consider the case where the transmitter has preemption capability. For both halves, we precisely compute the performance of some canonical policies and theoretically find the optimal policies, which lay the foundation for further generalization and application of AoII.

# AGE OF INCORRECT INFORMATION: A NEW PERFORMANCE METRIC IN SEMANTIC COMMUNICATIONS

by

Yutao Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Anthony Ephremides, Chair/Advisor
Professor Sennur Ulukus
Professor Prakash Narayan
Professor Behtash Babadi
Professor Ashok Agrawala

# Dedication

This dissertation is dedicated to my family for their unwavering and boundless support.

# Acknowledgments

Foremost, I would like to thank my advisor, Dr. Ephremides, whose support and guidance have been the cornerstone of my academic journey. As a beacon of scholarly excellence, he has exemplified the traits of a good researcher and has always been a direction and role model for my future endeavors. I still remember how nervous I was when I first met him, and how overwhelmed I was to be a student of such a prestigious scholar in the academic world. However, his amiable personality dispelled my worries and strengthened my resolve to emulate him in the future. I am always grateful for his mentorship, which has left an indelible mark on my life.

I am also grateful to all the scholars who have helped me over the years. Whether our paths have crossed or not, their advice is one of the things that motivates me to work hard and move forward. I also thank all the members of my defense committee, whose constructive feedback has helped me to refine my dissertation.

Next, I must express my heartfelt appreciation to my family, despite the distance that separates us. Without the support and understanding of my family, it would have been difficult for me to pursue my personal goals with determination. Their resolute backing in every decision I made, even when the path was unclear, has been a vital source of strength and determination. Once again, I would like to express my highest respect and gratitude to my family.

Finally, I would like to thank my friends, both in my home country and here. They have been invaluable companions on this journey of life and have helped me in countless ways, both

academically and in life. They have shared in my joys and sorrows, and their unwavering camaraderie has been an invaluable asset that I cherish. To them, I offer my most profound thanks and deep respect.

I would be remiss if I failed to acknowledge the mysterious force of luck that has been a fortuitous presence in my life. Its timely intervention during critical moments has been a crucial element in shaping who I am today. To luck, I extend my deepest appreciation, and I sincerely hope that its influence will continue to grace my life.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

## 1.1 Background and Motivation

With the development of communication technology, communication systems have become ubiquitous in all aspects of our daily lives, not only for transmitting text, voice, and images [1–4]. This pervasiveness of communication systems and their expanding communication purposes require us to set higher performance requirements. However, as the requirements increase, the traditional optimization frameworks seem increasingly ineffectual. To illustrate, let us consider the status update system depicted in Figure 1.1. In this system, the dispatcher controls the


Figure 1.1: An illustration of the status update system.

transmission of updates over a communication channel to a remote destination to ensure the remote destination has the best real-time knowledge of the dynamic source. The performance of the system depends on the chosen optimization framework. In traditionally optimization frameworks, we choose to optimize metrics such as throughput and delay. However, these metrics are objective-blind, i.e., they do not take into account the communication goals of the system. They assume that optimizing these metrics will always result in the best system performance. Unfortunately, the fact is that using these traditional metrics can result in reduced performance

and wasted resources.

To see such dissatisfaction, we again consider a time-sensitive application with the system model given by Figure 1.1, where the dispatcher is now an elementary queue, and the dynamic source submits data at an adjustable rate. We note that the most critical task in time-sensitive applications is to guarantee the timeliness of the information at the receiving end. We know that the most common optimization framework in throughput is to maximize it. To do this, we need to increase the data generation rate to the point where system resources are fully utilized. Increasing the data generation rate this way may sound appealing because resources are well spent. However, this can cause problems in time-sensitive applications because increasing the data generation rate can cause the data to wait too long in the queue, which significantly ages the data. So, blindly maximizing the throughput is not the best option. At the same time, delay causes the same dissatisfaction for the opposite reason. In delay minimization frameworks, we minimize the time it takes for data to arrive at its destination from the start of transmission. Such requirements are usually achieved by carefully adjusting the data generation rate so that the data spends the minimum amount of time in the queue. However, we know that delays occur only when data is being transmitted. Therefore, to reduce the average delay, the system can skip transmissions if the expected delay is too large. Reducing the delay by skipping the transmission will result in too little data being transmitted, which will harm the timeliness of the information at the receiving end.

One of the reasons these optimization frameworks fall short in time-sensitive applications is that these metrics treat all transmitted updates equally, ignoring the fact that not all updates provide the same level of importance to the receiver for communication purposes. For example, in time-sensitive applications, an update after a long idle is more important than an update after

a period of sustained transmissions. Hence, adopting traditional metrics alone is not enough. To optimize communication performance, it is critical to consider other factors, such as the freshness and content of the information being transmitted. By taking these factors into account, we can achieve a more efficient communication system that better meets the specific communication purposes of the data exchange.

## 1.2 Semantic Communication

Awareness of these problems and the importance of solving them has led to new ways of designing and evaluating communication networks. One example is semantic communication. According to [5], semantic communication is defined as *"the provisioning of the right and significant piece of information to the right point of computation (or actuation) at the right time"*. However, traditional network optimization frameworks do not incorporate the semantics of information, which can be highly inefficient given the unprecedented growth in data exchange. Therefore, the first step in addressing this problem is to define the semantics of information. In [5], the authors claim that the semantics of information is *"defined not necessarily as the meaning of the messages, but as their significance, possibly within a real-time constraint, relative to the purpose of the data exchange"*. Hence, in semantics communications, updates are treated differently based on their significance relative to the communication purpose. To achieve this vision, unlike the metrics considered in conventional optimization frameworks, the metrics in semantic communications should incorporate one or more semantic measures. The paper outlines three examples of semantic measures, the first being *Freshness*, which captures the timeliness of the information. This measure is crucial in applications that require real-time information, such as

vehicular networks. The second measure is *Relevance*, which measures the change in a process since the last sample. This measure can alter the sampling strategies for remote estimation or tracking with low-power devices. The third measure is *Value*, which evaluates the benefit of having a particular sample compared to the cost of transmitting it. This measure is particularly applicable to control networks, where incorporating the value of information can help achieve the same level of performance with reduced data traffic. Overall, semantic communication marks a departure from the traditional approaches of processing and transmitting the data by incorporating the semantics of information.

### 1.2.1 Information Freshness

One of the most successful efforts in incorporating the semantics of information is the introduction of the Age of Information (AoI) in [6]. AoI captures the freshness of information by tracking the time elapsed since the generation of the last received update. Let $V(t)$ be the most recent generation time of updates received up to time $t$. Then, AoI at time $t$ is defined by

$$\Delta_{AoI}(t) = t - V(t).$$

We notice that the age increases when no new updates are received, and the delivery of new updates can bring down the age. Therefore, AoI is small when the receiver has relatively timely information. Note that, under AoI, not all updates are equal. For example, when the update can significantly reduce the age at the receiver side, it becomes important and worth the extra resources to transmit.

After its introduction, AoI has attracted extensive attention and research [7–9]. In general,

the AoI analysis can be divided into the following cases.

- *Time-average Age*: The average AoI over a time period $T$ can be calculated by

$$\langle \Delta \rangle_T = \frac{1}{T} \int_0^T \Delta_{AoI}(t) dt.$$

  In this case, there are two commonly used analysis tools. The first is based on a graphical method that divides the zigzag dynamics into several trapezoids. The other leverages the notion of stochastic hybrid system (SHS), a mathematical framework used to model complex systems that exhibit both continuous and discrete behavior.

- *Peak Age [10]*: The peak age $A_n$ is the age at the instant before the delivery of the $n$th update. The average peak age is defined by

$$\Delta^{(p)} = \lim_{T \to \infty} \frac{1}{N(T)} \sum_{n=1}^{N(T)} A_n,$$

  where $N(T)$ is the number of peaks in the time period $T$. Peak age is more tractable than time-average AoI.

- *Functions of Age [11]*: The age penalty function is denoted by $p(\Delta_{AoI}(t))$ where $p : [0; \infty) \mapsto \mathbb{R}$ is non-decreasing. One can specify the age penalty function based on the applications.

Equipped with the above analysis approaches, the majority of the work on AoI can be roughly divided into the following categories.

- *AoI in queuing systems* [6, 10, 12–17]: In this case, AoI is studied in a system with one or

more queues and sources submit updates according to stochastic processes.

- *AoI under resource constraints* [18–22]: In this case, AoI is examined in systems where the capability of the transmitter is constrained by external factors such as limited or unstable power supplies. Hence, the transmitter should choose wisely how to utilize the limited resources to maximize the freshness of the information.

- *AoI in remote estimation* [11, 23–25]: The core task of remote estimation is to reconstruct the real-time information of the transmitting end at the receiving end. Consequently, the timeliness of the information is important. Hence, the relationship between estimation error and AoI has also become a hot research direction.

- *AoI in wireless networks* [26–30]: In this case, AoI is studied in systems where the data is distributed over wireless networks. The literature mainly focuses on AoI analysis and optimization under conventional protocols. At the same time, age-based protocols are also proposed and analyzed [31–35].

Although AoI captures well the timeliness of information, the limitation is that AoI ignores the information content of the transmitted updates. Therefore, it falls short in the context of remote estimation. For example, we want to estimate a rapidly changing event remotely. A small AoI does not necessarily mean that the receiver has accurate information about the event. Likewise, when the event changes slowly, the receiver may not need very timely information to estimate the event relatively accurately. It is shown in [23] that optimizing the AoI does not necessarily lead to the minimum estimation error. Thus, considering only the freshness of the information may lead to non-optimal performances in some applications. At the same time, we recall that error-based metrics, such as real-time error, are typical for remote estimation

tasks. Consequently, researchers seek to establish a more sophisticated framework that combines freshness and error-based metrics, which brings us to the Age of Incorrect Information (AoII) introduced in [8].

### 1.2.2 Age of Incorrect Information

We consider a basic transmitter-receiver pair in a slotted-time system, where the transmitter observes a dynamic source and transmits updates over a communication channel to the remote receiver. In light of the shortcomings discussed in the previous subsection, a more sophisticated performance metric should capture the following two aspects of the system.

The first aspect is the purpose of the update transmission. Let us take remote estimation as an example for a more intuitive understanding. Generally, in remote estimation problems, the purpose of the update transmission is to allow the receiver to have enough information to reconstruct the information of the remote source in real time. Thus, as long as the receiver has enough information for its reconstruction, we can assume that the purpose of the update transmission is achieved, and thus no further penalties need to be paid. We note that AoI does not always capture this. For example, we want to monitor the water level in a reservoir. The water level changes slowly. Hence, old information may still contain the correct information about the water level. Then, we do not need to waste resources to keep the information at the receiver as fresh as possible. For a more formal elaboration, let $X_t$ and $\hat{X}_t$ be the information content of the dynamic source and the information content of the receiver at time slot $t$, respectively. Then, we define the information penalty function as $g(X_t, \hat{X}_t)$ where $g(\cdot) : \mathcal{S} \times \mathcal{S}' \mapsto [0, +\infty)$. $\mathcal{S}$ and $\mathcal{S}'$ are the state space of $X_t$ and $\hat{X}_t$, respectively. Then, we can choose $g(X_t, \hat{X}_t)$ so that it quantifies the

information inadequacy on the receiver side. Note that we define $g(X_t, \hat{X}_t) \triangleq 0$ as the case where the receiver has enough information to achieve the communication purpose. In practice, we can choose different information penalty functions according to the sensitivity of different systems to information inadequacy. For example, $g(X_t, \hat{X}_t) = |X_t - \hat{X}_t|$, $g(X_t, \hat{X}_t) = (X_t - \hat{X}_t)^2$, $g(X_t, \hat{X}_t) = \mathbb{1}\{|X_t - \hat{X}_t| > c\}$, and so on.

The second aspect is the impact of inadequate information on system performance over time. We notice that, for many real-world applications, the impact of information inadequacy on system performance accumulates over time. In other words, an old, persistent inadequacy may have more severe consequences than a new, short-term one. However, this is rarely captured by traditional performance metrics. For example, when we use real-time error as the performance metric, we focus on the estimation error of the moment and ignore its history, which can lead to huge costs because small errors that persist for a long time can cause as much damage to the system as a short-lived significant error. A real-world example of this is machine temperature monitoring, where the damage caused by overheating accumulates rapidly over time. Therefore, we introduce the time penalty function to capture the time effect. To this end, we first define

$$U_t = \max\{k \leq t : g(X_k, \hat{X}_k) = 0\}.$$

Hence, $t - U_t$ captures the time elapsed since the last time the receiver had enough information to achieve the communication purpose. Because of the different definitions of $U_t$ and $V_t$, there is a fundamental difference between AoI and $t - U_t$. Recall that AoI will continue to increase when no new update is received, and the successful delivery of a new update can bring down AoI. However, this is not true for $t - U_t$. Note that the old information may still convey enough information

for the receiver to accomplish the communication purpose, and the new update may still be insufficient due to the communication delay. Hence, the delivery of updates does not necessarily lead to a decrease or increase in the value of $t-U_t$. Then, on top of $U_t$, we define the time penalty function as $f(t-U_t)$ where $f(\cdot):[0,+\infty)\mapsto[0,+\infty)$ can be any non-decreasing function. Then, we choose $f(t-U_t)$ so that it can capture the aging process of the inadequate information. In the application, we are free to choose which time penalty function to adopt based on the characteristics of the system under consideration. For example, $f(t)=t^2$, $f(t)=\log(1+t)$, $f(t)=e^t$, and so on. Then, combined with the information penalty function introduced earlier, we are ready to present the formal definition of AoII.

In a slotted-time system, the AoII is defined by

$$\Delta_{AoII}(X_t,\hat{X}_t,t)=\sum_{k=U_t+1}^{t}\Big(g(X_k,\hat{X}_k)F(k-U_t)\Big),$$

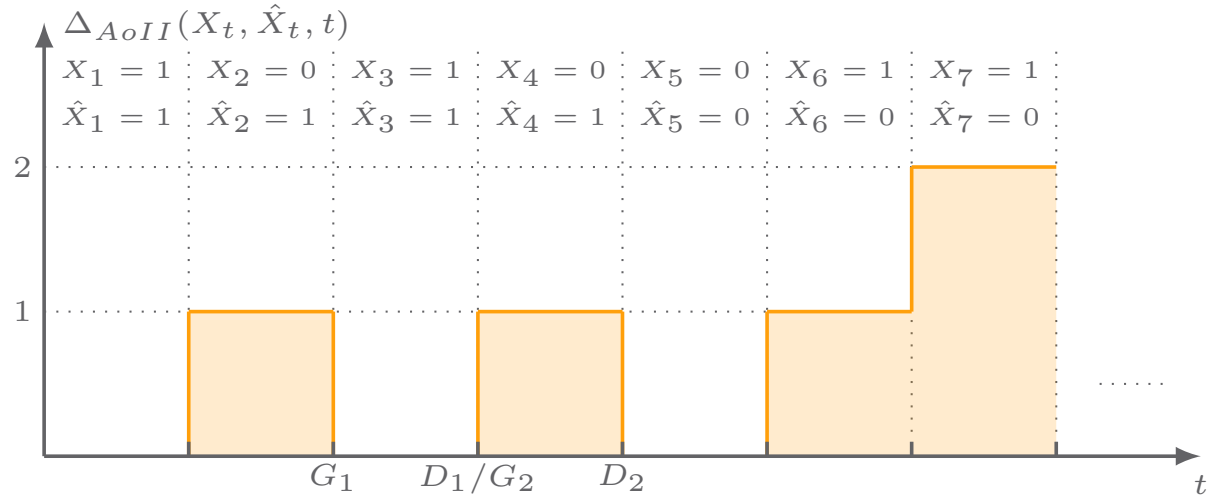
where $F(t)\triangleq f(t)-f(t-1)$ captures the increment of the time penalty. We notice that AoII increases if the receiver does not have enough information to achieve the communication purpose, and the increase is amplified by the degree of inadequacy. Hence, AoII captures both the inadequacy of the information and the aging process of the inadequate information. To better understand the metric, we consider the case where we want to estimate a two-state dynamic source remotely. Hence, $X_t\in\{0,1\}$ and $\hat{X}_t\in\{0,1\}$. Moreover, we choose $g(X_t,\hat{X}_t)=|X_t-\hat{X}_t|$ and $f(t)=t$. Then, AoII can be simplified as $\Delta_{AoII}(X_t,\hat{X}_t,t)=t-U_t$. We note that for this particular choice of penalty functions, AoII is simply the time elapsed since the last time the receiver's estimate was correct. A sample path of $\Delta_{AoII}(X_t,\hat{X}_t,t)$ in this case is depicted in Figure 1.2a. We also consider a more sophisticated setting, where the source process

has eight states. More precisely, $X_t \in \{0, 1, ..., 7\}$ and $\hat{X}_t \in \{0, 1, ..., 7\}$. Meanwhile, we choose $f(t) = t^2$ and $g(X_t, \hat{X}_t) = \mathbb{1}\{|X_t - \hat{X}_t| \geq 2\}$. A sample path of AoII with this setting is shown in Figure 1.2b.

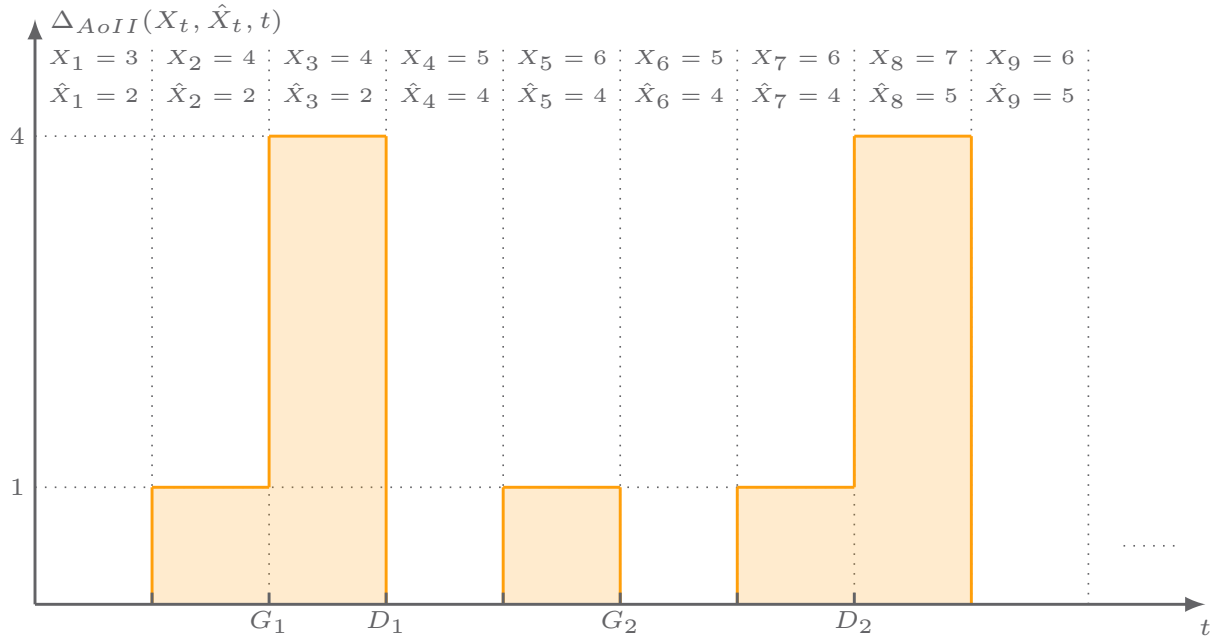
Unlike traditional metrics such as throughput and delay, AoII treats each update differently depending on how well the information it carries can achieve the communication purpose and the current AoII of the system. For example, we consider the case where an update carries the enough information about the source, and the current AoII of the system is high. In this case, it is worth spending extra resources to ensure the timely and accurate delivery of the update to the receiving end. At the same time, AoII also differs from AoI in that AoI ignores the information content of the transmitted updates. In AoII, the information content carried by the update will strongly influence the evolution of AoII.

## 1.3 Research Trends

In [36], the authors consider a transmitter-receiver pair in a slotted time system. In the setting, the updates are transmitted over an unreliable channel. The minimization of AoII for the transmitter with and without power constraints is studied. The power constraints limit the average number of transmissions a transmitter can initiate. When the transmitter is power-constrained, the results show that the optimal policy is a mixture of two threshold policies, where the threshold policy is the one under which the transmitter transmits the update only when the AoII exceeds the threshold. Then, the authors extend the results to the case of the generic time penalty function in [37]. Similar results are obtained in this case. Moreover, three real-life applications are studied to highlight the performance advantages of AoII over AoI and real-time estimation error.




(a) For a two-state source with $g(X_t, \hat{X}_t) = |X_t - \hat{X}_t|$ and $f(t) = t$.




(b) For an eight-state source with $g(X_t, \hat{X}_t) = \mathbb{1}\{|X_t - \hat{X}_t| \geq 2\}$ and $f(t) = t^2$.

Figure 1.2: Sample paths of $\Delta_{AoII}(X_t, \hat{X}_t, t)$. In the figure, $X_t$ and $\hat{X}_t$ are the state of the dynamic source and the receiver's estimate at time slot $t$, respectively. $G_i$ and $D_i$ are the sampling time and delivery time of the $i$-th update, respectively. Note that the sampling decisions in the graph are random. We assume that the sampling occurs at the beginning of a time slot and the update arrives at the end of a time slot.

The authors of [38] consider a problem where we need to monitor symmetric binary sources over a communication channel with random packet delivery times. More precisely, the source decides whether or not to sample at the beginning of each time slot. The sample enters the channel, and after a random number of time slots delay, it is received error-free at the monitor. The authors consider three performance metrics: AoII, AoI, and real-time error. The authors formulate the optimal sampling problem using the Markov decision process (MDP) and apply a simplified relative value iteration (RVI) algorithm to obtain the optimal policy for each metric. Furthermore, the performance of the obtained optimal policy is compared with two canonical sampling policies: the sample-at-change policy and the zero-wait policy, through simulation results. Under the sample-at-change policy, it samples the source whenever the source changes state. Under the zero-wait policy, it samples the source whenever possible. The numerical results indicate that for various delay distributions and AoII penalty functions, the optimal policies for the real-time error and AoII coincide with the sample-at-change policy. In contrast, the optimal policy for AoI is a threshold policy. In most cases, the threshold is zero, which is equivalent to the zero-wait policy.

The authors of [39] studied the AoII minimization problem in the context of scheduling. They consider a system where the central scheduler monitors multiple Markovian sources and needs to update multiple users. The central scheduler can only update a subset of the users in each time slot. Hence, the central scheduler must choose which users to update to minimize the AoII. A fundamental assumption made in the paper is that the central scheduler cannot know the states of the sources prior to transmission. In this case, the central scheduler cannot know the exact value of AoII in each time slot. To overcome this problem, the authors introduce the belief value, which can be interpreted as the probability that the state at the receiver side is

correct. Then, the Whittle's index policy is developed, and the performance is compared with the Whittle's index policy developed when AoI is adopted.

In the real world, we usually do not know the statistical model of the dynamic process we want to observe, or we do not know the parameters of the statistical model. Minimizing AoII, in this case, is usually difficult, but very important. Therefore, the authors of [40] consider the problem of minimizing AoII without knowing the parameters of a Markovian process. The relationship between the estimation error and AoII is studied in [41]. Moreover, a variant of AoII, the Age of Incorrect Estimates (AoIE), is proposed in [42]. AoIE is defined as the product of an estimation error and the time elapsed since the last time at which the receiver had a sufficiently correct estimate. In the paper, the authors consider a slotted-time system where a source generates status updates about an auto-regressive Markov process and sends the updates to a remote receiver. Then, the receiver will make estimations about the realization of the process based on the previously received updates using the least-square estimation strategy. The paper considers the optimal sampling problem, and the results show that the optimal policy is a threshold-type policy. Another variant called Age of Outdated Information ($\text{Ao}^2\text{I}$) is introduced in [43]. In this paper, $\text{Ao}^2\text{I}$ quantifies the elapsed time between the present and the first time when the stored information at the destination becomes outdated compared to its source. Then, $\text{Ao}^2\text{I}$ is studied in the context of scheduling, and the authors derive a theoretical lower bound for the minimum $\text{Ao}^2\text{I}$ and propose a low-complexity online scheduler.

## 1.4 Outline and Contribution

In Chapter 2, we investigate the problem of minimizing AoII under certain constraints. Specifically, we consider the AoII in a system where a transmitter sends updates about a multi-state Markovian source to a remote receiver over an unreliable channel. The communication goal is to minimize AoII subject to a power constraint. We cast the problem into a constrained Markov decision process (CMDP) and prove that the optimal policy is a combination of two deterministic threshold policies. Afterward, by leveraging the notion of RVI and the structural properties of threshold policies, we propose an efficient algorithm to find the threshold policies as well as the mixing coefficient.

In Chapter 3, we investigate the properties and optimization of AoII in scheduling problems. More precisely, we study a slotted-time system where a base station needs to update multiple users simultaneously. Due to the limited resources, only part of the users can be updated in each time slot. We consider the problem of minimizing AoII when imperfect channel state information (CSI) is available. Leveraging the notion of MDP, we obtain the structural properties of the optimal policy. By introducing a relaxed version of the original problem, we develop the Whittle's index policy under a simple condition. However, indexability is required to ensure the existence of the Whittle's index. To avoid indexability, we develop a new scheduling policy called the indexed priority policy based on the optimal policy for the relaxed problem.

In Chapter 4 and Chapter 5, we investigate the effect of delay on the performance and optimization of AoII. Specifically, we consider a transmitter-receiver pair in a slotted-time system. The transmitter observes a dynamic source and sends updates to a remote receiver over an error-free communication channel that suffers a random delay. The receiver estimates the state of the

dynamic source using the received updates. In Chapter 4, we assume that when the channel is busy, the transmitter can do nothing but wait for the channel to become idle. Then, we investigate the problem of optimizing the transmitter's action in each time slot to minimize AoII. We first characterize the optimization problem using the MDP and investigate the performance of the threshold policy, under which the transmitter transmits updates only when the transmission is allowed and the AoII exceeds the threshold $\tau$. By delving into the characteristics of the system evolution, we precisely compute the expected AoII achieved by the threshold policy using the Markov chain. Then, we prove that the optimal policy exists and provide a computable RVI algorithm to estimate the optimal policy. Furthermore, by leveraging the policy improvement theorem, we theoretically prove that, under an easily verifiable condition, the optimal policy is the threshold policy with $\tau = 1$. In Chapter 5, we consider a similar problem, but the transmitter has the ability to preempt. In this scenario, we prove the existence of the optimal policy and provide a feasible value iteration algorithm to approximate the optimal policy. We also analyze the system characteristics under two canonical delay distributions and theoretically obtain the corresponding optimal policies using the policy improvement theorem.

# Chapter 2: Freshness under Limited Resources

## 2.1 Overview

With the massive deployment of communication devices, the energy consumption of these devices has become a critical design consideration. Typically, these devices can only access limited resources to perform their tasks because they must share a centralized energy source with other devices or obtain their own energy from the outside world to operate for extended periods of time without human intervention. This leads to the question of how to optimize system performance with limited resources. In this chapter, we study the optimization problem in the presence of power constraints. The existing literature [36, 37] only considers simple communication models and penalty functions. Thus, the performance of more general AoII in a more complicated system is still unclear. In particular, we study the system where the source is modeled by a multi-state Markov chain. Meanwhile, the considered AoII adopts non-binary information penalty function. Similar constraints are considered in [18], with the goal of minimizing the AoI. Our problem is more complicated because AoI ignores the information content of the transmitted updates. Remote estimation under resource constraints is studied in [44–47]. However, they focus mainly on minimizing the estimation error but ignore the effect of time since, in many real-world applications, persistent errors will cause more damage to the system than short-lived errors.

The main contributions of this chapter can be summarized as follows. 1) We study the minimization of AoII under power constraints. 2) We consider the AoII that adopts non-binary information penalty function and model the source using a multi-state Markov chain. 3) We rigorously prove the structural properties of the optimal policy and propose an efficient algorithm to obtain the optimal policy.

This chapter summarizes the work in [48], and the rest is organized as follows. Section 2.2 elaborates on the modeling of the system and discusses the system dynamic under the chosen AoII. In section 2.3, we present a step-by-step analysis of the optimization problem and detail the proposed algorithm. Lastly, in Section 2.4, numerical results are laid out.

## 2.2 System Overview

### 2.2.1 System Model

We consider a slotted-time system in which a transmitter sends updates about a process to a remote receiver through an unreliable channel. The transmitted update will not be corrupted during the transmission but the transmission will not necessarily succeed. When the transmission fails, the update will be discarded and it will not affect the transmitter's decision at the next time slot. We denote the channel realization as $r_t$ where $r_t = 1$ if the transmission succeeds and $r_t = 0$ otherwise. We assume $r_t$ is independent and identically distributed over the time slots. We define $Pr(r_t = 1) = p_s$ and $Pr(r_t = 0) = 1 - p_s = p_f$. We notice that, in many status-update systems, the size of the update is very small so that the transmission time for an update is much smaller than the time unit used to measure the dynamic of the process. Thus, when a transmission attempt succeeds, the update is assumed to be received instantly by the receiver.

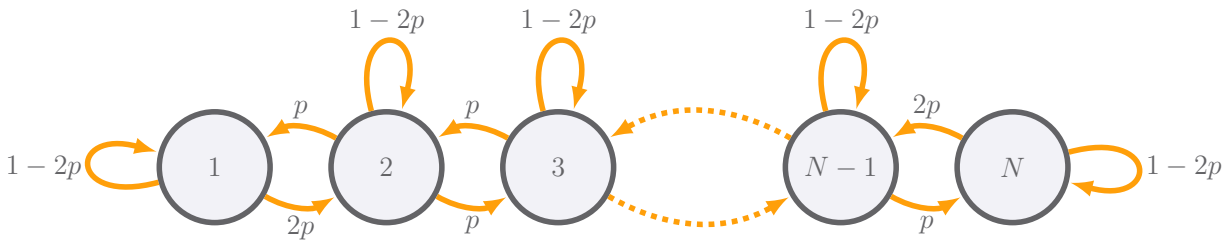



Figure 2.1: Illustrations of the Markovian source.

This assumption will provide us with analytical benefits, and a similar assumption is also made in [49]. The source process $\{X_t\}_{t\in\mathbb{N}}$ is modeled by an N-state Markov chain where transmissions only happen between adjacent states with probability $2p$ and themselves with probability $1-2p$. An illustration of the Markovian source is shown in Figure 2.1. The transmitter is capable of generating update $X_t$ by sampling the process at any time on its own will. However, the sampling opportunities only occur at the beginning of each time slot. We assume the transmitter is also capable of making transmission attempt in the same time slot the sampling happens. Every time the transmission succeeds, the receiver will use the received update as its new estimate $\hat{X}_t$. The receiver will send an $ACK/NACK$ packet to inform the transmitter whether it has received a new update. We suppose that the $ACK/NACK$ packets will be delivered reliably, and the transmission time is negligible as the packets are very small in general. Therefore, if $ACK$ is received, the transmitter knows that the transmission succeeded, and the receiver's estimate changed to the transmitted update. If $NACK$ is received, the transmitter knows that the receiver did not receive the new update, and the receiver's estimate did not change. Hence, we can assume that the transmitter always knows the receiver's estimate.

The system adopts the AoII as the performance measure. Since the dynamic source has $N$ states, we let $X_t \in \{1, ..., N\}$ and $\hat{X}_t \in \{1, ..., N\}$. Then, we choose $g(X_t, \hat{X}_t) = |X_t - \hat{X}_t|$, and $f(t) = t$. Consequently, $F(t) = 1$ by its definition and $d_t \triangleq g(X_t, \hat{X}_t) \in \{0, 1, ..., N-1\}$.

### 2.2.2 System Dynamic

Now, we tackle down the system dynamic which can be fully captured by the dynamic of the pair $(d_t, \Delta_t)$. $\Delta_t$ is short for $\Delta_{AoII}(X_t, \hat{X}_t, t)$. Thus, it is essential to characterize the relationship between $(d_{t+1}, \Delta_{t+1})$ and $(d_t, \Delta_t)$. We notice that the relationship depends on the transmitter's action and its result. Therefore, we define $a_t \in \{0, 1\}$ as the transmitter's action at time $t$ where $a_t = 1$ if the transmitter makes the transmission attempt and $a_t = 0$ otherwise. Then, we can divide our discussion into the following three cases.

Case 1: $a_t = 0$. In this case, no new update is received by the receiver. Thus, the estimate $\hat{X}_{t+1}$ will be nothing but $\hat{X}_t$, and $X_t$ will evolve following the Markov chain shown in Figure 2.1. When the state of the source process does not change which happens with probability $1 - 2p$, we have $X_{t+1} = X_t$. Then, $d_{t+1} = d_t$. When the state of the source process changes, $d_{t+1}$ depends on the value of $d_t$. Thus, we further distinguish between the following cases:

- When $d_t = 0$, according to the Markovian source reported in Figure 2.1, $d_{t+1} = 1$ with probability $2p$.
- When $1 \leq d_t \leq N - 2$, to simplify our analysis, we assume, for any $X_t \in \{1, 2, ..., N\}$, $Pr(X_{t+1} = X_t - 1 \mid X_t) = Pr(X_{t+1} = X_t + 1 \mid X_t) = p$. Then, when $X_t > \hat{X}_t$, $d_{t+1} = |X_t \pm 1 - \hat{X}_t| = |d_t \pm 1| = d_t \pm 1$. When $X_t < \hat{X}_t$, $d_{t+1} = |X_t \pm 1 - \hat{X}_t| = | - d_t \pm 1| = d_t \mp 1$. Combining together, $d_{t+1} = d_t \pm 1$ with equal probability $p$.
- When $d_t = N - 1$, $X_t$ must be either 1 or $N$ and $\hat{X}_t$ must be either $N$ or 1, respectively. Combining with the Markovian source reported in Figure 2.1, $d_{t+1} = N-2$ with probability $2p$.

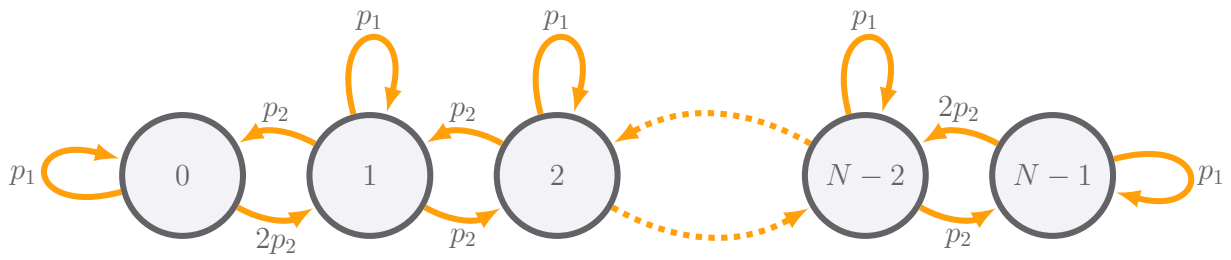



Figure 2.2: Illustrations of the evolution of $d$.

Let us denote by $P_{d,d'}$ the transition probability from $d$ to $d'$. Then, the results can be summarized as follows.

$$\begin{cases} P_{d,d} = 1 - 2p & 0 \leq d \leq N-1, \\ P_{d,d+1} = P_{d,d-1} = p & 1 \leq d \leq N-2, \\ P_{0,1} = P_{N-1,N-2} = 2p. \end{cases} \tag{2.1}$$

Such dynamic can be characterized by the Markov chain shown in Figure 2.2 with $p_1 = 1 - 2p$ and $p_2 = p$. Meanwhile, the value of $\Delta_{t+1}$ can be captured by the following two cases.

- When $d_{t+1} = 0$, the receiver's estimate at time $t+1$ is correct. By definition, $U_{t+1} = t+1$. Hence, $\Delta_{t+1} = 0$.
- When $d_{t+1} \neq 0$, the receiver's estimate at time $t+1$ is incorrect. In this case, $U_{t+1} = U_t$ by definition. Therefore, $\Delta_{t+1} = \Delta_t + d_{t+1}$.

To sum up,

$$\Delta_{t+1} = \mathbb{1}\{d_{t+1} \neq 0\}(\Delta_t + d_{t+1}),$$

where $\mathbb{1}\{A\}$ is an indicator function that takes value 1 when $A$ is true and 0 otherwise.

Case 2: $a_t = 1$ but $r_t = 0$. In this case, we notice that no new update is received by the receiver. Following the same trajectory as in Case 1, we can conclude that the dynamic of $d_t$ can

be characterized by the Markov chain shown in Figure 2.2 with $p_1 = p_f(1-2p)$ and $p_2 = p_f p$. $\Delta_{t+1}$ is fully dictated by $\Delta_t$ and $d_{t+1}$ as detailed in Case 1.

Case 3: $a_t = 1$ and $r_t = 1$. In this case, the receiver receives the update instantly. Thus, $U_{t+1} = t$. Then, we can conclude that $\Delta_{t+1} = d_{t+1}$. Since the update is received instantly, we have $d_{t+1} \in \{0, 1\}$. More precisely,

$$Pr[(d_{t+1}, \Delta_{t+1}) = (0,0) | (d_t, \Delta_t), a_t = 1] = p_s(1-2p).$$

$$Pr[(d_{t+1}, \Delta_{t+1}) = (1,1) | (d_t, \Delta_t), a_t = 1] = 2p_s p.$$

Combining the above three cases, we can fully capture the evolution of $(d_t, \Delta_t)$.

### 2.2.3 Problem Formulation

We consider the problem where there exists a unit power consumption along with each transmission attempt regardless of the result. At the same time, the transmitter has a power budget $\alpha < 1$. We define $\phi = (a_0, a_1, ...)$ as a sequence of actions the transmitter takes and denote all the feasible series of actions as $\Phi$. Then, our problem can be formulated as

$$\begin{aligned} \underset{\phi \in \Phi}{\arg\min} \quad & \bar{\Delta}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \Delta_t \right) \\ \text{subject to} \quad & \bar{R}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\phi \left( \sum_{t=0}^{T-1} a_t \right) \leq \alpha. \end{aligned} \tag{2.2}$$

As (2.2) shows, the system is resource-constrained. Thus, we realize a necessity to require the transmission attempts to help minimize AoII. More precisely, we require $Pr[(0,0) \mid (d_t, \Delta_t), a_t =$

$1] \geq Pr[(0,0) \mid (d_t, \Delta_t), a_t = 0]$ for any $(d_t, \Delta_t)$. Leveraging the system dynamic in Section 2.2.2, we conclude that it is sufficient to require $p \in [0, \frac{1}{3}]$. Therefore, we only consider the case of $p \in [0, \frac{1}{3}]$ throughout the rest of this chapter. A similar assumption is also made in [37].

We notice that solving problem (2.2) is equivalent to solving a constrained Markov decision process (CMDP). To this end, we adopt the Lagrangian approach.

## 2.3 Problem Optimization

### 2.3.1 Lagrangian Approach

First of all, we write the constrained optimization problem (2.2) into its Lagrangian form.

$$\mathcal{L}(\phi, \lambda) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_\phi \left[ \sum_{t=0}^{T-1} (\Delta_t + \lambda a_t) \right] - \lambda \alpha,$$

where $\lambda$ is the Lagrange multiplier. Then, the corresponding dual function will be

$$\mathcal{G}(\lambda) = \min_{\phi \in \Phi} \mathcal{L}(\phi, \lambda). \tag{2.3}$$

According to the results in [50], the optimal policy for the constrained problem (2.2) can be characterized by the optimal policies for the minimization problem (2.3) under certain $\lambda$. Thus, we start with solving problem (2.3) for any given $\lambda \geq 0$. As $\lambda\alpha$ is independent of policy, we can ignore it which leads to the following optimization problem.

$$\underset{\phi \in \Phi}{\text{minimize}} \quad \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_\phi \left[ \sum_{t=0}^{T-1} (\Delta_t + \lambda a_t) \right]. \tag{2.4}$$

The above problem can be cast into an infinite horizon with average cost Markov decision process (MDP) $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \mathcal{P}, \mathcal{C})$, where

- $\mathcal{X}$ denotes the state space: the state is $x = (d, \Delta)$. We define $x_d = d$ and $x_\Delta = \Delta$. We will use $x$ and $(d, \Delta)$ to represent the state interchangeably for the rest of this chapter.
- $\mathcal{A}$ denotes the action space: the two feasible actions are making the transmission attempt $(a = 1)$ and staying idle $(a = 0)$. The action space is independent of the state and the time. More precisely, $a \in \mathcal{A}(x, t) = \mathcal{A} = \{0, 1\}$.
- $\mathcal{P}$ denotes the state transition probabilities: we define $P_{x,x'}(a)$ as the probability that action $a$ at state $x$ will lead to state $x'$. The values of $P_{x,x'}(a)$ can be obtained easily from Section 2.2.2.
- $\mathcal{C}$ denotes the instant cost: when the system is at state $x$ and action $a$ is chosen, the instant cost is $C(x, a) = x_\Delta + \lambda a$.

### 2.3.2 Structural Results

In this section, we provide the key structural properties of the optimal policy for $\mathcal{M}$, which plays a vital role in the analysis later on. The optimal policy for $\mathcal{M}$ is captured by its value function $V(x)$, which can be obtained by solving the Bellman equation. In the infinite horizon with average cost MDP, the Bellman equation is defined as

$$\theta + V(x) = x_\Delta + \min_{a \in \{0,1\}} \left\{ \lambda a + \sum_{x' \in \mathcal{X}} P_{x,x'}(a) V(x') \right\}, \tag{2.5}$$

where $\theta$ is the minimal value of (2.4). A canonical procedure to solve the Bellman equation is applying the relative value iteration (RVI) [51]. To this end, we denote by $V_\nu(\cdot)$ the estimated value function at iteration $\nu$ of RVI. We initialize $V_0(x) = x_\Delta$. Then, the estimated value function is updated in the following way.

$$V_{\nu+1}(x) = Q_{\nu+1}(x) - Q_{\nu+1}(x^{ref}), \tag{2.6}$$

where $x^{ref}$ is the reference state. $Q_{\nu+1}(x)$ is the interim value function and is calculated by applying the right-hand side of (2.5). More precisely,

$$Q_{\nu+1}(x) = x_\Delta + \min_{a\in\{0,1\}} \left\{ \lambda a + \sum_{x'\in\mathcal{X}} P_{x,x'}(a) V_\nu(x') \right\}. \tag{2.7}$$

RVI is guaranteed to converge to $V(\cdot)$ when $\nu \to +\infty$ regardless of the initialization [51]. However, it requires infinitely many iterations to achieve the exact solution. To conquer the impracticality, we leverage the special properties of our system and provide the structural property of $V_\nu(\cdot)$, which turns out to be enough to characterize the structure properties of the optimal policy for $\mathcal{M}$. We start with the following lemma.

**Lemma 1.** *The estimated value function $V_\nu(x)$ is increasing in both $x_d$ and $x_\Delta$ at any iteration $\nu$.*

*Proof.* Leveraging the iterative nature of RVI, we use induction to prove the desired results. The complete proof can be found in Appendix A.1. □

We refer state $x$ as active if the optimal action at $x$ is making the transmission attempt and inactive otherwise. Then, leveraging Lemma 1, we provide the key structural properties of the

optimal policy for $\mathcal{M}$.

**Proposition 1.** *The optimal policy for $\mathcal{M}$ under any $\lambda \geq 0$ is a threshold policy which possesses the following properties.*

- *State $(0,0)$ will never be active.*
- *For states $x$ with fixed $x_d \neq 0$, the optimal action $a^*$ will switch from $a^* = 0$ to $a^* = 1$ as $x_\Delta$ increases and the switching point (i.e. threshold) is non-increasing in $x_d$.*

*Proof.* The optimal action at state $x$ is captured by the sign of $\delta V(x) \triangleq V^1(x) - V^0(x)$ where $V^a(x)$ is the value function resulting from taking action $a$ at state $x$. Then, we characterize the sign of $\delta V(x)$ using Lemma 1. The complete proof can be found in Appendix A.2. □

We define the threshold for the states with fixed $d \neq 0$ as the smallest $\Delta$ such that $a^* = 1$. We notice that state $(0,0)$ will never be active if optimal policy is adopted. Hence, we can characterize the optimal policy using the thresholds. In the following, an optimal policy is denoted by a vector $\boldsymbol{n}_\lambda$ where $(\boldsymbol{n}_\lambda)_i$ is the threshold for the states with $d = i$. The subscript $\lambda$ indicates the dependency between the optimal policy and $\lambda$.

### 2.3.3 Finite-State MDP Approximation

In the sequel, we tackle down the problem of finding the optimal policy for $\mathcal{M}$. Direct application of RVI becomes impractical as we need to estimate infinitely many value functions at each iteration. To overcome this difficulty, we use approximating sequence method (ASM) [52] and rigorously show the convergence of this approximation. To this end, we construct another

$\mathcal{M}^{(m)} = (\mathcal{X}^{(m)}, \mathcal{A}, \mathcal{P}^{(m)}, \mathcal{C})$ by truncating the value of $\Delta$. More precisely, we impose

$$\mathcal{X}^{(m)}: \begin{cases} x_d^{(m)} \in \{0, 1, ..., N-1\}, \\ x_\Delta^{(m)} \in \{0, 1, ..., m\}, \end{cases}$$

where $m$ is the predetermined maximal value of $\Delta$. The transition probabilities from $x \in \mathcal{X}^{(m)}$ to $z \in \mathcal{X} - \mathcal{X}^{(m)}$ (called excess probabilities) are redistributed to the states $x' \in \mathcal{X}^{(m)}$ in the following way.

$$P_{x,x'}^{(m)}(a) = P_{x,x'}(a) + \sum_{z \in \mathcal{X} - \mathcal{X}^{(m)}} P_{x,z}(a) q_z(x'),$$

where $q_z(x')$ is the probability of distributing state $z$ to state $x'$ and $\sum_{x' \in \mathcal{X}^{(m)}} q_z(x') = 1$. We choose $q_z(x') = \mathbb{1}_{\{x'_d = z_d\}} \times \mathbb{1}_{\{x'_\Delta = m\}}$. So, the transition probabilities $P_{x,x'}^{(m)}(a)$ satisfy the following.

$$P_{x,x'}^{(m)}(a) = \begin{cases} P_{x,x'}(a) & x'_\Delta < m, \\ P_{x,x'}(a) + \sum_{G(z,x')} P_{x,z}(a) & x'_\Delta = m, \end{cases}$$

where $G(z, x') = \{z : z_d = x'_d, z_\Delta > m\}$. The action space $\mathcal{A}$ and the instant cost $\mathcal{C}$ are the same as defined in $\mathcal{M}$.

**Theorem 1.** *The sequence of optimal policies for $\mathcal{M}^{(m)}$ will converge to the optimal policy for $\mathcal{M}$ as $m \to \infty$.*

*Proof.* We show that our system verifies the two assumptions given in [52]. Then, the convergence is guaranteed according to the results in [52]. The complete proof can be found in Appendix A.3.

□

For a given truncation parameter $m$, the state space $\mathcal{X}^{(m)}$ is finite with size $|\mathcal{X}^{(m)}| \propto m$.

When $m$ is huge, the basic RVI will be inefficient since the minimum operator in (2.7) requires calculations for both feasible actions at every state. In the following, we propose an improved RVI which avoids minimum operators at certain states. To this end, we claim that the properties in Proposition 1 are also possessed by the optimal policies for $\mathcal{M}^{(m)}$ at any iteration $\nu$ of RVI. The proof is omitted since it is very similar to what we did in Section 2.3.2. Utilizing Proposition 1, we can conclude that, for any state $x^{(m)}$, if there exists an active state $y^{(m)}$ such that $y_\Delta^{(m)} \leq x_\Delta^{(m)}$ and $y_d^{(m)} \leq x_d^{(m)}$, then $x^{(m)}$ must also be active. The update step at each iteration of the improved RVI can be summarized as follows. For each $x^{(m)} \in \mathcal{X}^{(m)}$,

- if $y^{(m)}$ exists, we can determine the optimal action at $x^{(m)}$ immediately, and the minimum operator is avoided.
- if $y^{(m)}$ does not exist, the optimal action at $x^{(m)}$ is determined by applying the minimum operator.

In this way, we avoid $\sum_{i=1}^{N}(m - n_i)$ minimum operators at each iteration of RVI where $n_i = (\boldsymbol{n}^t)_i$ and $\boldsymbol{n}^t$ is the optimal policy at iteration $\nu$ of RVI. In almost all cases, we have $n_i \ll m$. The pseudocode is given in Algorithm 1. In the algorithm, $V_\nu$ converges when the maximum difference between the results of two consecutive iterations is less than $\epsilon$.

### 2.3.4 Expected Transmission Rate

In this section, we calculate the expected transmission rate $\bar{R}_{\boldsymbol{n}}$ under given threshold policy $\boldsymbol{n}$. It enables us to develop an efficient algorithm for finding the optimal policy for (2.2). As we can see in (2.2), $\bar{R}_{\boldsymbol{n}}$ is nothing but the expected average number of transmission attempts made. Thus, it can be fully captured by the stationary distribution of the discrete-time Markov chain

**Algorithm 1** Improved Relative Value Iteration

**procedure** RELATIVEVALUEITERATION($\mathcal{M}$)
  Initialize $V_0(x) = x_\Delta$; $\nu = 0$
  Choose $x^{ref} \in \mathcal{X}$ arbitrarily
  **while** $V_\nu$ is not converged **do**
    **for** $x \in \mathcal{X}$ **do**
      **if** $\exists$ active state $y$ s.t. $y_d \leq x_d$ and $y_\Delta \leq x_\Delta$ **then**
        $a^*(x) = 1$
        $Q_{\nu+1}(x) = C(x,1) + \sum_{x'} P_{xx'}(1) \cdot V_\nu(x')$
      **else**
        **for** $a \in \mathcal{A}$ **do**
          $H_{x,a} = C(x,a) + \sum_{x'} P_{xx'}(a) \cdot V_\nu(x')$
        $a^*(x) = \arg\min_a \{H_{x,a}\}$
        $Q_{\nu+1}(x) = H_{x,a^*}$
    **for** $x \in \mathcal{X}$ **do**
      $V_{\nu+1}(x) = Q_{\nu+1}(x) - Q_{\nu+1}(x^{ref})$
    $\nu = \nu + 1$
  **return** $\boldsymbol{n} \leftarrow a^*(x)$

(DTMC) induced by $\boldsymbol{n}$. More precisely,

$$\bar{R}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \sum_{\Delta=n_d}^{+\infty} \pi_d(\Delta),$$

where $\pi_d(\Delta)$ is the steady-state probability of state $(d, \Delta)$ and $n_d = (\boldsymbol{n})_d$. To obtain the stationary distribution, we utilize the balance equation associated with the induced DTMC which takes the following form

$$\pi(x) = \sum_{x' \in \mathcal{X}} P_{x',x}(a^{\boldsymbol{n}}_{x'}) \pi(x'), \tag{2.8}$$

where $a^{\boldsymbol{n}}_{x'}$ is the action suggested by policy $\boldsymbol{n}$ at state $x'$. The problem arises since the state space of the induced DTMC is infinite. To overcome the difficulty, we present the following proposition.

**Proposition 2.** *The expected transmission rate under threshold policy $\boldsymbol{n}$ is*

$$\bar{R}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \left( \sum_{\Delta=n_d}^{\tau-1} \pi_d(\Delta) + \Pi_d(\tau) \right),$$

*where $\tau = max\{\boldsymbol{n}\}$. $\pi_d(\Delta)$'s and $\Pi_d(\tau)$'s are the solution to the following finite system of linear equations.*

$$\begin{aligned} \pi_0(0) =& (1-2p)\pi_0(0) + p\sum_{\Delta=1}^{\tau-1}(1-p_s a_{1,\Delta})\pi_1(\Delta) + p_f p \Pi_1(\tau) + \\ & p_s(1-2p)\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right), \end{aligned} \tag{2.9}$$

$$\pi_1(1) = 2p\pi_0(0) + 2p_s p \sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right), \tag{2.10}$$

$$\sum_{d=1}^{N-1}\left(\sum_{\Delta=l_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right) + \pi_0(0) = 1, \tag{2.11}$$

*and for each $1 \le d \le N-1$,*

$$\pi_d(\Delta) = 0, \quad 0 < \Delta < l_d, \tag{2.12}$$

$$\pi_d(\Delta) = \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d})\pi_{d'}(\Delta - d), \quad \max\{2, l_d\} \le \Delta \le \tau - 1, \tag{2.13}$$

$$\Pi_d(\tau) = \sum_{d'=1}^{N-1} P_{d',d}\left(\sum_{\Delta=\tau-d}^{\tau-1}(1-p_s a_{d',\Delta})\pi_{d'}(\Delta) + p_f \Pi_{d'}(\tau)\right), \tag{2.14}$$

*where $l_d = \frac{d^2+d}{2}$ and $a_{d,\Delta}$ is the action suggested by $\boldsymbol{n}$ at state $(d, \Delta)$.*

*Proof.* We cast the induced infinite-state Markov chain to an equivalent finite-state Markov chain with size depending on the policy. Then, $\pi_d(\Delta)$'s and $\Pi_d(\tau)$'s are the steady-state probabilities

of the finite-state Markov chain. The complete proof can be found in Appendix A.4. □

**Remark 1.** *We can verify that* (2.9) *is a linear combination of the other equations in the system of linear equations. Therefore, we can exclude* (2.9) *in practice.*

The system of linear equations can be reformulated into the matrix form $\boldsymbol{A}\boldsymbol{\pi} = \boldsymbol{b}$ where $\boldsymbol{\pi}$ is the unknowns and $\boldsymbol{A}$, $\boldsymbol{b}$ can be obtained easily from Proposition 2. However, solving a finite system of linear equations can still be problematic, especially when the system is huge. For our problem, the size of $\boldsymbol{A}$ is $\mathcal{O}((N-1)\tau)$, and we notice that $\boldsymbol{A}$ is sparse.

In general, solving a large system of linear equations of size $\mathcal{O}(n)$ requires $\mathcal{O}(n^2)$ storage and $\mathcal{O}(n^3)$ floating-point arithmetic operations when $\boldsymbol{A}$ is dense. In the case of sparse $\boldsymbol{A}$, the computational cost will be less. The sparse matrix algorithms are designed to solve equations in time and space proportional to $\mathcal{O}(n) + \mathcal{O}(cn)$ where $c$ is the average number of non-zero entries in each column. Although there are cases where this linear target cannot be met, the complexity of sparse linear algebra is far less than that in dense case [53]. Generally speaking, the complexity depends on the sparsity of $\boldsymbol{A}$. By exploiting the zero entries, we can often reduce the storage and computational requirements to $\mathcal{O}(cn)$ and $\mathcal{O}(cn^2)$, respectively.

The calculation of $\bar{R}_{\boldsymbol{n}}$ is constantly needed, and it requires a significant amount of time when the thresholds in $\boldsymbol{n}$ are huge. Hence, we provide an efficient alternative that can approximate the expected transmission rate in this case.

**Corollary 1.** *When the thresholds in $\boldsymbol{n}$ are huge, the expected transmission rate under policy $\boldsymbol{n}$ can be approximated as*

$$\bar{R}_{\boldsymbol{n}} \approx \sum_{d=1}^{N-1} \left( \sum_{\Delta=n_d}^{\tau-1} \pi_d(\Delta) + \Pi_d(\tau) \right),$$

*where $\tau = max\{\boldsymbol{n}\}$. $\pi_d(\Delta)$'s and $\Pi_d(\tau)$'s are the solution to the following finite system of linear*

*equations.*

$$\begin{aligned}\pi_0(0) =&(1-2p)\pi_0(0) + p\Pi_1(\eta) + p_f p\Pi_1(\tau) + p\sum_{\Delta=\eta+1}^{\tau-1}(1-p_s a_{1,\Delta})\pi_1(\Delta)+\\ &p_s(1-2p)\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta)+\Pi_d(\tau)\right),\end{aligned} \tag{2.15}$$

$$\pi_1(1) = 2p\pi_0(0) + 2p_s p\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta)+\Pi_d(\tau)\right), \tag{2.16}$$

$$\sum_{d=1}^{N-1}\left(\Pi_d(\eta)+\sum_{\Delta=\eta+1}^{\tau-1}\pi_d(\Delta)+\Pi_d(\tau)\right)+\pi_0(0)=1, \tag{2.17}$$

$$\Pi_1(\eta)-\pi_1(1)+\pi_1(\eta+1)=\sum_{d'=1}^{N-1}P_{d',1}\Pi_{d'}(\eta), \tag{2.18}$$

$$\Pi_d(\eta)+\sum_{\Delta=\eta+1}^{\eta+d}\pi_d(\Delta)=\sum_{d'=1}^{N-1}P_{d',d}\Pi_{d'}(\eta),\quad 2\le d\le N-1, \tag{2.19}$$

*and for each* $1\le d\le N-1$*:*

$$\pi_d(\Delta)=\rho\sum_{d'=1}^{N-1}P_{d',d}(1-p_s a_{d',\Delta-d})\sigma_{d'}(\Delta-d),\quad \eta+1\le\Delta\le\eta+d, \tag{2.20}$$

$$\pi_d(\Delta)=\sum_{d'=1}^{N-1}P_{d',d}(1-p_s a_{d',\Delta-d})\pi_{d'}(\Delta-d),\quad \eta+d+1\le\Delta\le\tau-1, \tag{2.21}$$

$$\Pi_d(\tau)=\sum_{d'=1}^{N-1}P_{d',d}\left(\sum_{\Delta=\tau-d}^{\tau-1}(1-p_s a_{d',\Delta})\pi_{d'}(\Delta)+p_f\Pi_{d'}(\tau)\right), \tag{2.22}$$

*where* $a_{d,\Delta}$ *is the action suggested by the threshold policy* $\boldsymbol{n}$ *at state* $(d,\Delta)$*,* $\eta=\min\{\boldsymbol{n}\}-1$ *and* $\rho=\frac{\pi_0(0)}{\sigma_0(0)}$*.* $\sigma_d(\Delta)$ *is the stationary distribution associated with the Markov chain induced by another threshold policy* $\boldsymbol{n}'=[\eta',...,\eta']$ *where* $\eta'=\eta+1$*.*

*Proof.* When the thresholds in $\boldsymbol{n}$ are huge, the expected transmission rate will be insignificant. Combining with (2.10), we have $\pi_1(1) \approx 2p\pi_0(0)$. Consequently, we can show that, for any state $(d, \Delta)$, $\pi_d(\Delta) \approx c^{\boldsymbol{n}}_{d,\Delta}\pi_0(0)$ where $c^{\boldsymbol{n}}_{d,\Delta}$ is a scalar that depends on the policy and the state. At the same time, we notice that, for any two threshold policies $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, the suggested actions at states with $\Delta < \min\{[\boldsymbol{n}_1, \boldsymbol{n}_2]\}$ are the same. We denote by $G(\boldsymbol{n}_1, \boldsymbol{n}_2)$ the set of these states. Then, we can prove that, for $(d, \Delta) \in G(\boldsymbol{n}_1, \boldsymbol{n}_2)$,

$$\frac{\pi^1_d(\Delta)}{\pi^2_d(\Delta)} \approx \frac{c_{d,\Delta}\pi^1_0(0)}{c_{d,\Delta}\pi^2_0(0)} = \frac{\pi^1_0(0)}{\pi^2_0(0)}, \tag{2.23}$$

where $\pi^1_d(\Delta)$ and $\pi^2_d(\Delta)$ are the stationary distributions when $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are adopted, respectively. Based on (2.23), we can obtain the two systems of linear equations. The complete proof is in Appendix A.5. □

**Remark 2.** *For the same reason as in Remark 1, we can exclude* (2.15) *in practice.*

In Corollary 1, instead of solving a large system of linear equations of size $\mathcal{O}((N-1)\tau)$, we approximate $\bar{R}_\lambda$ by solving two systems of linear equations of size $\mathcal{O}((N-1)(\eta+1))$ and $\mathcal{O}((N-1)(\tau-\eta+1))$, respectively. It is worth noting that when $\tau \approx \eta$ or $\tau \gg \eta$, the complexity reduction of Corollary 1 is limited. For other cases, Corollary 1 can significantly reduce the complexity and the resulting error is negligible. The methodology presented in Proposition 2 can also be applied to the calculation of the expected AoII $\bar{\Delta}_{\boldsymbol{n}}$. More precisely, we can use the following corollary.

**Corollary 2.** *The expected AoII under threshold policy $\boldsymbol{n}$ is*

$$\bar{\Delta}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \left( \sum_{\Delta=l_d}^{\tau-1} \omega_d(\Delta) + \Omega_d(\tau) \right),$$

*where $\tau = max\{\boldsymbol{n}\}$ and $l_d = \frac{d^2+d}{2}$. $\omega_d(\Delta)$'s and $\Omega_d(\tau)$'s are the solution to the following finite system of linear equations. For each $1 \leq d \leq N-1$:*

$$\omega_0(0) = 0;\ \omega_d(\Delta) = 0, \quad \Delta < l_d, \tag{2.24}$$

$$\omega_d(\Delta) = \Delta \pi_d(\Delta), \quad l_d \leq \Delta \leq \tau - 1, \tag{2.25}$$

$$\Omega_d(\tau) = \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau-d}^{\tau-1} (1 - p_s a_{d',\Delta}) \omega_{d'}(\Delta) + p_f \Omega_{d'}(\tau) \right) + d\Pi_d(\tau), \tag{2.26}$$

*where $a_{d,\Delta}$ is the action suggested by the threshold policy $\boldsymbol{n}$ at state $(d, \Delta)$. $\pi_d(\Delta)$'s and $\Pi_d(\tau)$'s can be obtained using Proposition 2 with the same threshold policy $\boldsymbol{n}$.*

*Proof.* We define $\omega_d(\Delta) \triangleq \Delta \pi_d(\Delta)$ and $\Omega_d(\tau) \triangleq \sum_{\Delta=\tau}^{+\infty} \omega_d(\Delta)$. Then, we combine the states with $\Delta \geq \tau$ as did in Proposition 2. After some rearrangements, we can obtain the system of linear equations shown above. The complete proof is in Appendix A.6. □

### 2.3.5 Optimal Policy

Till this point, we are able to find the optimal policy for problem (2.4). However, our goal is to find the optimal policy for the constrained problem (2.2). Based on the work in [50], the optimal policy for problem (2.2) can be expressed as a mixture of two deterministic policies that are both optimal for problem (2.4) with $\lambda = \lambda^*$. More precisely, the optimal policy can be

summarized in the following theorem.

**Theorem 2.** *The optimal policy for the constrained problem* (2.2) *can be expressed as a mixture of two deterministic policies* $\boldsymbol{n}_{\lambda_+^*}$ *and* $\boldsymbol{n}_{\lambda_-^*}$ *that are both optimal for problem* (2.4) *with* $\lambda = \lambda^* \triangleq \inf\{\lambda > 0 : \bar{R}_\lambda \leq \alpha\}$. $\bar{R}_\lambda$ *is the expected transmission rate resulting from policy* $\boldsymbol{n}_\lambda$. *More precisely, if we choose*

$$\mu = \frac{\alpha - \bar{R}_{\lambda_+^*}}{\bar{R}_{\lambda_-^*} - \bar{R}_{\lambda_+^*}}, \tag{2.27}$$

*the mixed policy* $\boldsymbol{n}_{\lambda^*}$, *which selects* $\boldsymbol{n}_{\lambda_-^*}$ *with probability* $\mu$ *and* $\boldsymbol{n}_{\lambda_+^*}$ *with probability* $1-\mu$ *each time the system reaches state* $(0,0)$, *is optimal for the constrained problem* (2.2) *and the constraint in* (2.2) *is met with equality.*

*Proof.* We verify that our system satisfies all the assumptions given in [50]. Then, combining the characteristics of our system and the results in [50], we obtain the optimal policy. The complete proof is in Appendix A.7. □

Next, we describe an efficient algorithm to obtain the optimal policy for the constrained problem (2.2). The core of obtaining the optimal policy is to find $\lambda^*$. We recall that, for any given $\lambda$, the deterministic policy $\boldsymbol{n}_\lambda$ is obtained by applying the improved RVI and the resulting $\bar{R}_\lambda$, which is non-increasing in $\lambda$ [50], is calculated using Proposition 2. Hence, $\bar{R}_\lambda$ can be regarded as a non-increasing function of $\lambda$, and we can use Bisection search with tolerance $\xi$ to find $\lambda^*$. Then, $\lambda_+^*$ and $\lambda_-^*$ can be the boundaries of the final interval. More precisely, we initialize $\lambda_- = 0$ and $\lambda_+ = 1$. Then, the procedure can be summarized as follows.

- As long as $\bar{R}_{\lambda_+} \geq \alpha$, we set $\lambda_- = \lambda_+$ and $\lambda_+ = 2\lambda_+$. Then, we end up with an interval $I = [\lambda_-, \lambda_+]$.

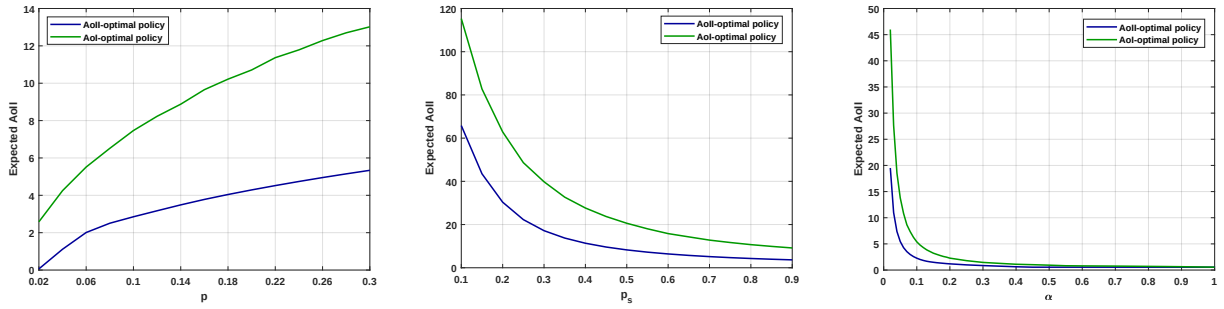
**Algorithm 2** Bisection Search

**procedure** BISECTIONSEARCH($\mathcal{M}^{(m)}(\lambda)$, $\alpha$)
  Initialize $\lambda_- = 0$; $\lambda_+ = 1$
  $\boldsymbol{n}_{\lambda_+} = RVI(\mathcal{M}^{(m)}(\lambda_+), \epsilon)$ using Algorithm 1
  $\bar{R}_{\lambda_+} = \bar{R}(\boldsymbol{n}_{\lambda_+})$ using Proposition 2
  **while** $\bar{R}_{\lambda_+} \geq \alpha$ **do**
    $\lambda_- = \lambda_+$; $\lambda_+ = 2\lambda_+$
    $\boldsymbol{n}_{\lambda_+} = RVI(\mathcal{M}^{(m)}(\lambda_+), \epsilon)$ using Algorithm 1
    $\bar{R}_{\lambda_+} = \bar{R}(\boldsymbol{n}_{\lambda_+})$ using Proposition 2
  **while** $\lambda_+ - \lambda_- \geq \xi$ **do**
    $\lambda = \frac{\lambda_+ + \lambda_-}{2}$
    $\boldsymbol{n}_{\lambda} = RVI(\mathcal{M}^{(m)}(\lambda), \epsilon)$ using Algorithm 1
    $\bar{R}_{\lambda} = \bar{R}(\boldsymbol{n}_{\lambda})$ using Proposition 2
    **if** $\bar{R}_{\lambda} \geq \alpha$ **then**
      $\lambda_- = \lambda$
    **else**
      $\lambda_+ = \lambda$
  **return** $(\lambda_+^*, \lambda_-^*) \leftarrow (\lambda_+, \lambda_-)$

- We apply Bisection search on the interval $I$ until the length of $I$ is less than the tolerance $\xi$. Then, the algorithm returns $\lambda_+^*$ and $\lambda_-^*$.

The pseudocode is given in Algorithm 2. Finally, the mixing coefficient $\mu$ is calculated using (2.27) and the resulting expected AoII is calculated using Corollary 2. The algorithm is efficient for the following reasons.

- We obtain $\boldsymbol{n}_{\lambda}$ using the improved RVI which avoids minimum operators at certain states.
- When calculating $\bar{R}_{\lambda}$, we cast the induced infinite-state Markov chain to a finite-state Markov chain.
- We find $\lambda_+^*$ and $\lambda_-^*$ using Bisection search which has a logarithmic complexity.




(a) The expected AoII in function of $p$.

(b) The expected AoII in function of $p_s$.

(c) The expected AoII in function of $\alpha$.

Figure 2.3: Illustrations of AoII-optimal policy and AoI-optimal policy. The truncation parameter in ASM $m = 800$ and the tolerance in Bisection search $\xi = 0.01$. RVI converges when the maximum difference between the results of two consecutive iterations is less than $\epsilon = 0.01$.

## 2.4 Numerical Results

In this section, we provide numerical results that accent the effect of system parameters on the performance of AoII-optimal policy. We also compare the AoII-optimal policy with the AoI-optimal policy derived in [18].

Effect of $p$: We compare the performances of AoII-optimal policies under different values of $p$. To this end, we fix $N = 7$ and $p_s = 0.8$. We also set $\alpha = 0.06$. We vary the value of $p$ and plot the corresponding results. As we can see in Figure 2.3a, the expected AoII is increasing in $p$. To explain this trend, we notice that as $p$ increases, the source process will be more inclined to change state at the next time slot. Then, those successfully transmitted updates will more likely be obsolete at the next time slot. As the power budget $\alpha$ is fixed which dictates the transmission rate, the expected AoII will increase as $p$ increases.

We also show, in Table 2.1, the deterministic policies $\boldsymbol{n}_{\lambda_+^*}$, $\boldsymbol{n}_{\lambda_-^*}$ and the corresponding mixing coefficient $\mu$ for some values of $p$. [1] The optimal policy is the mixture of $\boldsymbol{n}_{\lambda_+^*}$ and $\boldsymbol{n}_{\lambda_-^*}$

[1] In the table, " / " indicates the threshold where the two policies differ.

Table 2.1: Optimal thresholds for different $p$

| | Mixing Coef. | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $n_5$ | $n_6$ |
|---|---|---|---|---|---|---|---|
| $p = 0.1$ | $\mu = 0.7176$ | 15 | 6/7 | 1 | 1 | 1 | 1 |
| $p = 0.2$ | $\mu = 0.0331$ | 37 | 16 | 8/9 | 1 | 1 | 1 |
| $p = 0.3$ | $\mu = 0.1178$ | 69 | 25/26 | 15 | 1 | 1 | 1 |

Table 2.2: Optimal thresholds for different $p_s$

| | Mixing Coef. | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $n_5$ | $n_6$ |
|---|---|---|---|---|---|---|---|
| $p_s = 0.2$ | $\mu = 0.6712$ | 556 | 228 | 140 | 96 | 70/71 | 60 |
| $p_s = 0.4$ | $\mu = 0.3260$ | 151 | 62 | 36/37 | 24 | 17 | 1 |
| $p_s = 0.6$ | $\mu = 0.4089$ | 67 | 27/28 | 16 | 1 | 1 | 1 |
| $p_s = 0.8$ | $\mu = 0.0331$ | 37 | 16 | 8/9 | 1 | 1 | 1 |

with mixing coefficient $\mu$ as described in Theorem 2. We can see that the thresholds are, in general, increasing in $p$. The reason behind this is as follows. When $p$ is small, the successfully transmitted updates are more likely still accurate in the next few time slots. In another word, the transmission is more "efficient". We refer a transmission as "efficient" if it reduces the age to the greatest extent. This allows the transmitter to make transmission attempts when the age is relatively low without violating the power constraint.

Effect of $p_s$: In this scenario, we fix $p = 0.2$ and investigate the effect of channel reliability $p_s$ on the performance of AoII-optimal policy. We still consider the case of $N = 7$ and $\alpha = 0.06$. The corresponding results are shown in Figure 2.3b. As $p_s$ increases, the expected AoII will decrease. The reason is as follows. As $p_s$ increases, the transmitted updates will more likely be successful. Consequently, the transmission will be more "efficient". As the power budget is fixed, the expected AoII will decrease as $p_s$ increases. We also present some selected thresholds in Table 2.2. As we see in the table, the thresholds are, in general, decreasing in $p_s$. We recall

that as $p_s$ increases, the transmission will be more "efficient". Thus, the transmitter can make transmission attempts when the age is relatively low while keeping the transmission rate not exceeding the power budget.

Effect of $\alpha$: Then, we analyze the performances of AoII-optimal policies under different values of $\alpha$. We adopt $N = 7$ and $p_s = 0.8$. We also set $p = 0.2$. The expected AoII achieved by the AoII-optimal policies are plotted in Figure 2.3c. As we see, the expected AoII decreases as $\alpha$ increases. The reason is simple. As the power budget increases, more transmission attempts are allowed. We recall that we impose a transmission attempt to always help reduce the age. Keeping this in mind, we can conclude that the expected AoII is decreasing in $\alpha$. It is worth noting that as $\alpha$ increases, the expected AoII will stop decreasing before $\alpha = 1$. To explain this, we recall that the transmitter will never make transmission attempt at state $(0, 0)$ if an optimal policy is adopted. Thus, as $\alpha$ becomes large, the transmitter will have enough budget to make transmission attempts at any states other than $(0, 0)$. Then, the transmission rate is saturated, and the expected AoII will not decrease further.

Comparison with AoI-optimal policy: Lastly, we compare the AoII-optimal policy with the AoI-optimal policy. From Figure 2.3, we can see that the performance gap expands with the increase in $p$ and the decrease in $p_s$ and $\alpha$. The reason behind it lies in the value of transmission attempts. As $p$ increases, the source process becomes more inclined to change states. Therefore, transmission attempts are more needed to bring correct information to the receiver. As $p_s$ decreases, the number of successful transmissions will decrease, and the AoII will build up faster. In this case, transmission attempts are more valuable because they will greatly reduce AoII once they

succeed. As $\alpha$ decreases, transmission attempts will be more valuable as fewer attempts are allowed. Due to the different definitions of age, there are often cases where AoI is large, but AoII is small. In these cases, the AoI-optimal policy will waste valuable transmission attempts. Therefore, the increase in the value of transmission attempts will lead to an expansion of the performance gap.

## 2.5 Conclusion

In this chapter, we consider a system where the source process is modeled by an N-state Markov chain. The AoII that adopts non-binary information penalty function is used. We study the problem of minimizing the AoII subject to a power constraint. By casting the problem into a CMDP, we can prove that the optimal policy is a mixture of two deterministic threshold policies. Then, an efficient algorithm is proposed to find such policies and the mixing coefficient. Lastly, numerical results are provided to illustrate the performance of the AoII-optimal policy and compare it with the AoI-optimal policy.

# Chapter 3: Scheduling for Freshness

## 3.1 Overview

With the development of communication and sensor technology, the number of users that a base station need to handle is constantly increasing. However, due to resource limitations, the base station can only process a subset of the users simultaneously. This leads us to study how the base station should reasonably schedule these resources so that all users can get optimal service under a given evaluation system. Therefore, in this chapter, we study the AoII minimization problem in the context of scheduling. We consider a system where a base station needs to update a part of all available users at the same time. Meanwhile, we consider the generic time penalty function and study the minimization problem in the presence of imperfect channel state information (CSI). Due to the presence of CSI, the Whittle's index policy becomes infeasible in general. Therefore, we introduce another scheduling policy that is more versatile and has comparable performance to Whittle's index policy. The scheduling problem with AoI as the performance measure is studied under various system settings in [54–58]. The problem studied in this chapter is different and more complicated because AoII considers the aging process of inconsistent information rather than the aging process of updates. Meanwhile, none of them consider the case where CSI is available. The problem of optimizing information freshness in the presence of CSI is studied in [59, 60]. However, they focus on the single-user system and

mainly discuss cases where CSI is perfect. The scheduling problems with the goal of minimizing an error-based performance measure are considered in [61–63]. Our problem is fundamentally different because AoII also considers the time effect. Moreover, we consider the system where a base station observes multiple sources simultaneously and has to send updates to multiple destinations.

The main contributions of this chapter can be summarized as follows. 1) We study the problem of minimizing AoII in a multi-user system where imperfect CSI is available. Meanwhile, the time penalty function is generic. 2) We derive the structural properties of the optimal policy for the considered problem. 3) We establish the indexability of the considered problem under a simple condition and develop the Whittle's index policy. 4) We obtain the optimal policy for a relaxed version of the original problem. By exploring the characteristics of the relaxed problem, we provide an efficient algorithm to obtain the optimal policy. 5) Based on the optimal policy for the relaxed problem, we develop the indexed priority policy, which is free of indexability and has comparable performance to Whittle's index policy.

This chapter summarizes the work in [64], and the rest is organized as follows. Section 3.2 introduces the system model and formulates the primal problem. In Section 3.3, we explore the structural properties of the optimal policy for the primal problem. Under a simple condition, we develop the Whittle's index policy in Section 3.4. Section 3.5 presents the optimal policy for a relaxed version of the primal problem. On this basis, we develop the indexed priority policy in Section 3.6. Finally, in Section 3.7, the numerical results are laid out.

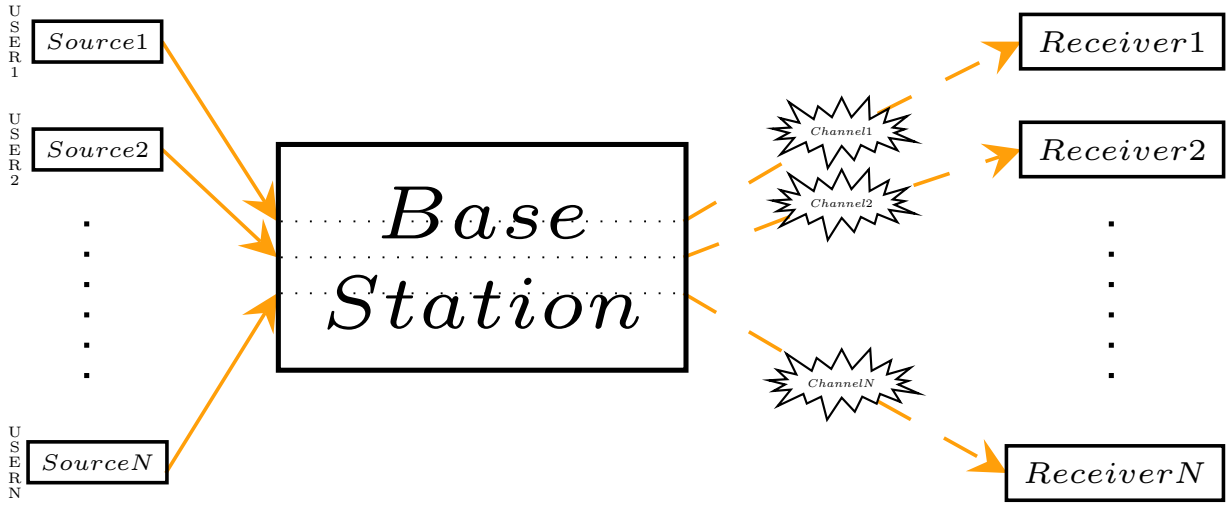



Figure 3.1: The structure of the system model.

## 3.2 System Overview

### 3.2.1 System Model

We consider a slotted-time system with $N$ users and one base station. Each user has a source process, a channel, and a receiver. We assume all the users share the same structure, but the parameters are different. The structure of the communication model is provided in Figure 3.1. For user $i$, the source process is modeled by a two-state Markov chain where transitions happen between the two states with probability $p_i > 0$ and self-transitions happen with probability $1-p_i$. At any time slot $t$, the state of the source process $X_{i,t} \in \{0,1\}$ will be reported to the base station as an update, and the base station will decide whether to transmit this update through the corresponding channel. The channel is unreliable, but the estimate of the CSI is available at the beginning of each time slot. Let $r_{i,t} \in \{0,1\}$ be the CSI at time $t$. We assume that $r_{i,t}$ is independent across time and user indices. $r_{i,t} = 1$ if and only if the transmission attempt at time $t$ will succeed and $r_{i,t} = 0$ otherwise. Then, we denote by $\hat{r}_{i,t} \in \{0,1\}$ the estimate of $r_{i,t}$. We

assume that $\hat{r}_{i,t}$ is an independent Bernoulli random variable with parameter $\gamma_i$, i.e., $\hat{r}_{i,t} = 1$ with probability $\gamma_i \in [0, 1]$ and $\hat{r}_{i,t} = 0$ with probability $1 - \gamma_i$. However, the estimate is imperfect. We assume that the error depends only on the user and its estimate. More precisely, we define the probability of error as $p_{e,i}^{\hat{r}_i} \triangleq Pr[r_i \neq \hat{r}_i \mid \hat{r}_i]$. We assume $p_{e,i}^{\hat{r}_i} < 0.5$ because we can flip the estimate if $p_{e,i}^{\hat{r}_i} > 0.5$. We are not interested in the case of $p_{e,i}^{\hat{r}_i} = 0.5$ since $\hat{r}_{i,t}$ is useless in this case. Although the channel is unreliable, each transmission attempt takes exactly one time slot regardless of the result, and the successfully transmitted update will not be corrupted. Every time an update is received, the receiver will use it as the new estimate $\hat{X}_{i,t} \in \{0, 1\}$. The receiver will send an $ACK/NACK$ packet to inform the base station of its reception of the new update. Since an $ACK/NACK$ packet is generally very small and simple, we assume that it is transmitted reliably and received instantaneously. Then, if $ACK$ is received, the base station knows that the receiver's estimate changed to the transmitted update. If $NACK$ is received, the base station knows that the receiver's estimate did not change. Therefore, the base station always knows the estimate at the receiver side.

At the beginning of each time slot, the base station receives updates from each source and the estimates of CSI from each channel. The old updates and estimates are discarded upon the arrival of new ones. Then, the base station decides which updates to transmit, and the decision is independent of the transmission history. Due to the limited resources, at most $M < N$ updates are allowed per transmission attempt. We consider a base station that always transmits $M$ updates.

All the users adopt AoII as a performance metric, but the choices of penalty functions may vary. We consider the case where the users adopt the same information penalty function $g(X_t, \hat{X}_t) = |X_t - \hat{X}_t|$ but possibly different time penalty functions. To ease the analysis, we require $f(t)$ to be unbounded. Combined with the monotonicity requirement of the time penalty

function, we require $f(t_1) \leq f(t_2)$ if $t_1 < t_2$ and $\lim_{t\to+\infty} f(t) = +\infty$. Without a loss of generality, we assume $f(0) = 0$. Meanwhile, as the source is modeled by a two-state Markov chain, $g(X_t, \hat{X}_t) \in \{0, 1\}$. Hence, AoII can be written as

$$\Delta_{AoII}(X_t, \hat{X}_t, t) = \sum_{k=U_t+1}^{t} F(k - U_t) = f(s_t),$$

where $U_t$ is defined in Section 1.2.2 and $s_t \triangleq t - U_t$. Therefore, the evolution of $s_t$ is sufficient to characterize the evolution of AoII. To this end, we distinguish between the following cases.

- When the receiver's estimate is correct at time $t + 1$, we have $U_{t+1} = t + 1$. Then, by definition, $s_{t+1} = 0$.
- When the receiver's estimate is incorrect at time $t + 1$, we have $U_{t+1} = U_t$. Then, by definition, $s_{t+1} = t + 1 - U_t = s_t + 1$.

To sum up, we get

$$s_{t+1} = \mathbb{1}\{U_{t+1} \neq t + 1\}(s_t + 1). \tag{3.1}$$

In the remainder of this chapter, we use $f_i(\cdot)$ to denote the time penalty function user $i$ adopts.

### 3.2.2 System Dynamic

In this section, we tackle the system dynamic. We notice that the status of user $i$ can be captured by the pair $x_{i,t} \triangleq (s_{i,t}, \hat{r}_{i,t})$. In the following, we will interchangeably use $x_{i,t}$ and $(s_{i,t}, \hat{r}_{i,t})$. Then, the system dynamic can be fully characterized by the dynamic of $\boldsymbol{x}_t \triangleq (x_{1,t}, ..., x_{N,t})$. Hence, it suffices to characterize the value of $\boldsymbol{x}_{t+1}$ given $\boldsymbol{x}_t$ and the base station's action. To this end, we denote, by $\boldsymbol{a}_t = (a_{1,t}, ..., a_{N,t})$, the base station's action at time $t$. $a_{i,t} = 1$

if the base station transmits the update from user $i$ at time $t$ and $a_{i,t} = 0$ otherwise. We notice that given action $\boldsymbol{a}_t$, users are independent and the action taken on user $i$ will only affect itself. Consequently

$$Pr(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t, \boldsymbol{a}_t) = \prod_{i=1}^{N} Pr(x_{i,t+1} \mid x_{i,t}, \boldsymbol{a}_t) = \prod_{i=1}^{N} Pr(x_{i,t+1} \mid x_{i,t}, a_{i,t}).$$

Combined with the fact that all the users share the same structure, it is sufficient to study the dynamic of a single user. In the following discussions, we drop the user-dependent subscript $i$. We recall that $\hat{r}_{t+1}$ is an independent Bernoulli random variable. Then, we have

$$Pr(x_{t+1} \mid x_t, a_t) = P(\hat{r}_{t+1}) \times Pr(s_{t+1} \mid x_t, a_t).$$

By definition, $P(\hat{r}_{t+1} = 1) = \gamma$ and $P(\hat{r}_{t+1} = 0) = 1 - \gamma$. Then, we only need to tackle the value of $Pr(s_{t+1} \mid x_t, a_t)$. To this end, we distinguish between the following cases

- When $x_t = (0, \hat{r}_t)$, the estimate at time $t$ is correct (i.e., $\hat{X}_t = X_t$). Hence, for the receiver, $X_t$ carries no new information about the source process. In other words, $\hat{X}_{t+1} = \hat{X}_t$ regardless of whether an update is transmitted at time $t$. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Since the source is binary, we obtain $U_{t+1} = U_t$ if $X_{t+1} \neq X_t$, which happens with probability $p$ and $U_{t+1} = t + 1$ otherwise. According to (3.1), we obtain

$$Pr(1 \mid (0, \hat{r}_t), a_t) = p.$$

$$Pr(0 \mid (0, \hat{r}_t), a_t) = 1 - p.$$

- When $a_t = 0$ and $x_t = (s_t, \hat{r}_t)$, where $s_t > 0$, the channel will not be used and no new update will be received by the receiver,and so, $\hat{X}_{t+1} = \hat{X}_t$. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Since $X_t \neq \hat{X}_t$ and the source is binary, we have $U_{t+1} = U_t$ if $X_{t+1} = X_t$, which happens with probability $1 - p$ and $U_{t+1} = t + 1$ otherwise. According to (3.1), we obtain

$$Pr(s_t + 1 \mid (s_t, \hat{r}_t), a_t = 0) = 1 - p.$$

$$Pr(0 \mid (s_t, \hat{r}_t), a_t = 0) = p.$$

- When $a_t = 1$ and $x_t = (s_t, 1)$ where $s_t > 0$, the transmission attempt will succeed with probability $1-p_e^1$ and fail with probability $p_e^1$. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Then, when the transmission attempt succeeds (i.e., $\hat{X}_{t+1} = X_t$), $U_{t+1} = U_t$ if $X_{t+1} \neq X_t$ and $U_{t+1} = t + 1$ otherwise. When the transmission attempt fails (i.e., $\hat{X}_{t+1} = \hat{X}_t \neq X_t$), we have $U_{t+1} = U_t$ if $X_{t+1} = X_t$ and $U_{t+1} = t + 1$ otherwise. Combining (3.1) with the dynamic of the source process we obtain

$$Pr(s_t + 1 \mid (s_t, 1), a_t = 1) = p_e^1(1 - p) + (1 - p_e^1)p \triangleq \alpha.$$

$$Pr(0 \mid (s_t, 1), a_t = 1) = p_e^1 p + (1 - p_e^1)(1 - p) = 1 - \alpha.$$

- When $a_t = 1$ and $x_t = (s_t, 0)$, where $s_t > 0$, following the same line, we obtain

$$Pr(s_t + 1 \mid (s_t, 0), a_t = 1) = p_e^0 p + (1 - p_e^0)(1 - p) \triangleq \beta.$$

$$Pr(0 \mid (s_t, 0), a_t = 1) = p_e^0(1-p) + (1-p_e^0)p = 1-\beta.$$

Combines together, we obtain the value of $Pr(s_{t+1} \mid x_t, a_t)$ in all cases. As only $M$ out of $N$ updates are allowed per transmission attempt, we realize a necessity to require transmission attempts always help minimize AoII. It is equivalent to impose $Pr(s_{t+1} > s_t \mid (s_t, \hat{r}_t), a_t = 0) > Pr(s_{t+1} > s_t \mid (s_t, \hat{r}_t), a_t = 1)$ for any $(s_t, \hat{r}_t)$. Leveraging the results above, it is sufficient to require $p < 0.5$. As all the users share the same structure, we assume, for the rest of this chapter, that $0 < p_i < 0.5$ for $1 \leq i \leq N$.

### 3.2.3 Problem Formulation

The communication goal is to minimize the expected AoII. Therefore, the problem can be formulated as the following

$$\begin{aligned} \underset{\phi \in \Phi}{\arg\min} \quad & \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N} f_i(s_{i,t}) \right) \\ \text{subject to} \quad & \sum_{i=1}^{N} a_{i,t} = M \quad \forall t, \end{aligned} \tag{3.2}$$

where $\Phi$ is the set of all causal policies. We refer to the constrained minimization problem reported in (3.2) as the primal problem (PP). We notice that the PP is a restless multi-armed bandit (RMAB) Problem. The optimal policy for this type of problem is far from reachable since it is PSPACE-hard in general [65]. However, we can still derive the structural properties of the optimal policy. These structural properties can be used as a guide for the development of scheduling policies and can indicate the good performance of the developed scheduling policies.

## 3.3 Structural Properties of the Optimal Policy

In this section, we investigate the structural properties of the optimal policy for PP. We first define an infinite horizon with an average cost Markov decision process (MDP) $\mathcal{M}_N(w, M) = (\mathcal{X}_N, \mathcal{A}_N(M), \mathcal{P}_N, \mathcal{C}_N(w))$, where

- $\mathcal{X}_N$ denotes the state space. The state is $\boldsymbol{x} = (x_1, ..., x_N)$ where $x_i = (s_i, \hat{r}_i)$.
- $\mathcal{A}_N(M)$ denotes the action space. The feasible action is $\boldsymbol{a} = (a_1, ..., a_N)$ where $a_i \in \{0, 1\}$ and $\sum_{i=1}^{N} a_i = M$. Note that the feasible actions are independent of the state and the time.
- $\mathcal{P}_N$ denotes the state transition probabilities. We define $P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a})$ as the probability that action $\boldsymbol{a}$ at state $\boldsymbol{x}$ will lead to state $\boldsymbol{x}'$. It is calculated by

$$P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a}) = \prod_{i=1}^{N} P(\hat{r}_i') P_{s_i,s_i'}(a_i, \hat{r}_i),$$

  where $P_{s_i,s_i'}(a_i, \hat{r}_i)$ is the transition probability from $s_i$ to $s_i'$ when the estimate of CSI is $\hat{r}_i$ and action $a_i$ is taken. The values of $P_{s_i,s_i'}(a_i, \hat{r}_i)$ can be obtained easily from the results in Section 3.2.2.
- $\mathcal{C}_N(w)$ denotes the instant cost. When the system is at state $\boldsymbol{x}$ and action $\boldsymbol{a}$ is taken, the instant cost is $C(\boldsymbol{x}, \boldsymbol{a}) \triangleq \sum_{i=1}^{N} C(x_i, a_i) \triangleq \sum_{i=1}^{N} \big(f_i(s_i) + w a_i\big)$.

We notice that PP can be cast into $\mathcal{M}_N(0, M)$. Since $w = 0$, the instant cost is independent of action $\boldsymbol{a}$. Therefore, we abbreviate $C(\boldsymbol{x}, \boldsymbol{a})$ as $C(\boldsymbol{x})$. To simplify the analysis, we consider the case of $M = 1$. Equivalently, we investigate the structural properties of the optimal policy for $\mathcal{M}_N(0, 1)$.

**Remark 3.** *For the case of $M > 1$, we can apply the same methodology. However, as $M$ increases, the action space will grow quickly, resulting in the need to consider more feasible actions in each step of the proof. Hence, to better demonstrate the methodology, we only consider the case of $M = 1$ in this chapter.*

It is well known that the optimal policy for $\mathcal{M}_N(0,1)$ can be characterized by the value function. We denote the value function of state $\boldsymbol{x}$ as $V(\boldsymbol{x})$. A canonical procedure to calculate $V(\boldsymbol{x})$ is applying the value iteration algorithm (VIA). To this end, we define $V_\nu(\cdot)$ as the estimated value function at iteration $\nu$ of VIA and initialize $V_0(\cdot) = 0$. Then, VIA updates the estimated value functions in the following way

$$V_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \min_{\boldsymbol{a}\in\mathcal{A}_N(1)} \left\{ \sum_{\boldsymbol{x}'\in\mathcal{X}_N} P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a}) V_\nu(\boldsymbol{x}') \right\}, \tag{3.3}$$

where $\theta$ is the optimal value of $\mathcal{M}_N(0,1)$. VIA is guaranteed to converge to the value function [51]. More precisely, $V_\nu(\cdot) = V(\cdot)$ when $\nu \to +\infty$. However, the exact value function is impossible to get since we need infinite iterations and the state space is infinite. Instead, we provide two structural properties of the value function.

**Lemma 2.** *For $\mathcal{M}_N(0,1)$, $V(\boldsymbol{x})$ is non-decreasing in $s_i$ for $1 \le i \le N$.*

*Proof.* Leveraging the iterative nature of VIA, we use mathematical induction to prove the desired results. The complete proof can be found in Appendix B.1. □

Before introducing the next structural property, we make the following definition.

**Definition 1** (Statistically identical)**.** *Two users are said to be statistically identical if the user-dependent parameters and the adopted time penalty functions are the same.*

For the users that are statistically identical, we can prove the following

**Lemma 3.** *For $\mathcal{M}_N(0,1)$, if users $j$ and $k$ are statistically identical, $V(\boldsymbol{x}) = V\left(\mathscr{P}(\boldsymbol{x})\right)$ where $\mathscr{P}(\boldsymbol{x})$ is state $\boldsymbol{x}$ with $x_j$ and $x_k$ exchanged.*

*Proof.* Leveraging the iterative nature of VIA, we use mathematical induction to prove the desired results. At each iteration, we show that for each feasible action at state $\boldsymbol{x}$, we can find an equivalent action at state $\mathscr{P}(\boldsymbol{x})$. Two actions are equivalent if they lead to the same value function. The complete proof can be found in Appendix B.2. □

Equipped with the above lemmas, we proceed with characterizing the structural properties of the optimal policy. We recall that the optimal action at each state can be characterized by the value function. Hence, we denote, by $V^j(\boldsymbol{x})$, the value function resulting from choosing user $j$ to update at state $\boldsymbol{x}$. Then, $V^j(\boldsymbol{x})$ can be calculated by

$$V^j(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'-x'_j} \left\{ \left( \prod_{i \neq j} P_{x_i,x'_i}(0) \right) \sum_{\hat{r}'_j} \left[ P(\hat{r}'_j) \left( \sum_{s'_j} P_{s_j,s'_j}(1,\hat{r}_j) V(\boldsymbol{x}') \right) \right] \right\}.$$

If $V^j(\boldsymbol{x}) < V^k(\boldsymbol{x})$ for all $k \neq j$, it is optimal to transmit the update from user $j$. When $V^j(\boldsymbol{x}) = V^k(\boldsymbol{x})$, the two choices are equally desirable. In the following, we will characterize the properties of $\delta^{j,k}(\boldsymbol{x}) \triangleq V^j(\boldsymbol{x}) - V^k(\boldsymbol{x})$ for any $j$ and $k$.

**Theorem 3.** *For $\mathcal{M}_N(0,1)$, $\delta^{j,k}(\boldsymbol{x})$ has the following properties*

1. *$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $\hat{r}_k = p^0_{e,k} = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$.*
2. *$\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$ and is non-decreasing in $\hat{r}_k$ when $s_j, s_k > 0$. At the same time, $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j, k$.*

3. $\delta^{j,k}(\boldsymbol{x}) \leq 0$ *if* $s_k = 0$. *The equality holds when* $s_j = 0$ *or* $\hat{r}_j = p^0_{e,j} = 0$.

4. $\delta^{j,k}(\boldsymbol{x})$ *is non-increasing in* $s_j$ *if* $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ *and is non-decreasing in* $s_k$ *if* $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ *when* $s_j, s_k > 0$. *We define* $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ *and* $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ *for* $1 \leq i \leq N$.

5. $\delta^{j,k}(\boldsymbol{x}) \leq 0$ *if* $s_j \geq s_k$, $\hat{r}_j \geq \hat{r}_k$, *and users* $j$ *and* $k$ *are statistically identical.*

*Proof.* The proof can be found in Appendix B.3. □

We notice that $\Gamma_i^{\hat{r}_i}$ can be written as

$$\Gamma_i^{\hat{r}_i} = \frac{Pr(s_i + 1 \mid (s_i, \hat{r}_i), a_i = 1)}{Pr(s_i + 1 \mid (s_i, \hat{r}_i), a_i = 0)} < 1,$$

where $s_i$ can be any positive integer. Consequently, $\Gamma_i^{\hat{r}_i}$ is independent of any $s_i > 0$ and indicates the decrease in the probability of increasing $s_i$ caused by action $a_i = 1$. When $\Gamma_i^{\hat{r}_i}$ is large, action $a_i = 1$ will achieve a small decrease in the probability of increasing $s_i$. In the following, we provide an intuitive interpretation of why the monotonicity in Property 4 of Theorem 3 depends on $\Gamma_i^{\hat{r}_i}$. We take the case of $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ as an example and assume that there are only users $j$ and $k$ in the system. Then, according to Section 3.2.2, the dynamic of $s_j$ and $s_k$ can be divided into the following three cases

- Neither $s_j$ nor $s_k$ increases. In this case, both $s_j$ and $s_k$ become zero.
- Either $s_j$ or $s_k$ increases and the other becomes zero. We denote by $P_j^k$ the probability that only $s_k$ increases when $a_j = 1$. The notation for other cases is defined analogously. The probabilities can be obtained easily using the results in Section 3.2.2.

- Both $s_j$ and $s_k$ increase. We denote by $P_j$ the probability that both $s_j$ and $s_k$ increase when $a_j = 1$. $P_k$ is defined analogously. The probabilities can be obtained easily using the results in Section 3.2.2.

We notice that $\delta^{j,k}(\boldsymbol{x})$ implies the tendency of the base station to choose between the two users. The larger $\delta^{j,k}(\boldsymbol{x})$ is, the more the base station tends to choose user $k$. Thus, we investigate the base station's propensity to choose user $k$ when $s_k$ increases but $s_j$ stays the same. We ignore the case where the resulting $s_k$ is zero since it is independent of the increase in $s_k$. With this in mind, we first notice that $P_k^k \leq P_j^k$. Meanwhile, we can easily verify that $\frac{P_j}{P_k} = \frac{\Gamma_j^{\hat{r}_j}}{\Gamma_k^{\hat{r}_k}}$. When $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$, we have $P_j \leq P_k$. Then, there exists a subtle trade-off. More precisely, choosing user $k$ will result in $P_k^k \leq P_j^k$, but at the cost of $P_k \geq P_j$. Hence, in this case, the propensity of the base station is hard to determine. Following the same line, we can show that choosing user $j$ will lead to $P_j^j \leq P_k^j$ and $P_j \leq P_k$. Thus, there exists no such trade-off when we investigate the base station's propensity to choose user $j$ as $s_j$ increases but $s_k$ stays the same.

Leveraging Theorem 3, we can provide some specific structural properties of the optimal policy.

**Corollary 3.** *When $M = 1$, the optimal policy for PP must satisfy the following*

1. *The user $i$ with $\hat{r}_i = p_{e,i}^0 = 0$ or $s_i = 0$ will not be chosen unless it is to break the tie.*
2. *When user $j$ is chosen at state $\boldsymbol{x}_1$, then for state $\boldsymbol{x}_2$, such that $\hat{r}_{1,j} \leq \hat{r}_{2,j}$ and $s_{1,i} = s_{2,i}$ for $1 \leq i \leq N$, the optimal choice must be in the set $G = \{j\} \cup \{k : \hat{r}_{1,k} < \hat{r}_{2,k}\}$.*
3. *When $N = 2$, we consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which differ only in the value of $s_j$. Specifically, $s_{1,j} \leq s_{2,j}$. If user $j$ is chosen at state $\boldsymbol{x}_1$ and $\Gamma_j^{\hat{r}_{1,j}} \leq \Gamma_k^{\hat{r}_{1,k}}$, the optimal choice at state $\boldsymbol{x}_2$ will also be user $j$.*

4. *When $N = 2$, we consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which differ only in the value of $s_k$. Specifically, $s_{1,k} \geq s_{2,k}$. If user $j$ is chosen at state $\boldsymbol{x}_1$ and $\Gamma_j^{\hat{r}_{1,j}} \geq \Gamma_k^{\hat{r}_{1,k}}$, the optimal choice at state $\boldsymbol{x}_2$ will also be user $j$.*

5. *When all users are statistically identical, the optimal choice at any time slot must be either the user with $x = (s_{max,1}, 1)$ where $s_{max,1} \triangleq \max_{s_i}\{(s_i, 1)\}$ or the user with $x = (s_{max,0}, 0)$ where $s_{max,0} \triangleq \max_{s_i}\{(s_i, 0)\}$. Moreover,*

   - *If $s_{max,1} \geq s_{max,0}$, it is optimal to choose the user with $x = (s_{max,1}, 1)$.*
   - *If $s_{max,1} < s_{max,0}$, the optimal choice will switch from the user with $x = (s_{max,0}, 0)$ to the user with $x = (s_{max,1}, 1)$ when $s_{max,1}$ increases from $0$ to $s_{max,0}$ solely.*

*Proof.* The first property follows directly from Property 1 and Property 3 of Theorem 3. For the second property, leveraging Property 2 of Theorem 3, we have $\delta^{j,k}(\boldsymbol{x}_2) \leq \delta^{j,k}(\boldsymbol{x}_1) \leq 0$ if $\hat{r}_{1,j} \leq \hat{r}_{2,j}$, $\hat{r}_{1,k} \geq \hat{r}_{2,k}$, and $s_{1,i} = s_{2,i}$ for $1 \leq i \leq N$. Thus, the optimal choice will not be user $k$ in this case. Then, we can conclude that the optimal choice must be in the set $G = \{j\} \cup \{k : \hat{r}_{1,k} < \hat{r}_{2,k}\}$.

For the third property, we have proved in Property 4 of Theorem 3 that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $s_j$ if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$. Hence, $\delta^{j,k}(\boldsymbol{x}_2) \leq \delta^{j,k}(\boldsymbol{x}_1) \leq 0$. As we consider the case of $N = 2$, the optimal choice at state $\boldsymbol{x}_2$ will also be user $j$. The fourth property can be shown in a similar way by noticing that $\delta^{j,k}(\boldsymbol{x})$ is non-decreasing in $s_k$ when $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$.

For the last property, we recall from Property 5 of Theorem 3 that it is always better to choose the user with a larger $s$ if they are statistically identical and have the same $\hat{r}$. Thus, we can conclude that the optimal choice must be either the user with $x = (s_{max,1}, 1)$ or the user with $x = (s_{max,0}, 0)$. Without a loss of generality, we assume $x_j = (s_{max,1}, 1)$ and $x_k = (s_{max,0}, 0)$.

Now, we distinguish between the following cases

- According to Property 5 of Theorem 3, we can conclude that it is optimal to choose user $j$ when $s_{max,1} \geq s_{max,0}$.
- To determine the optimal choice in the case of $s_{max,1} < s_{max,0}$, we recall that the optimal choice will be user $k$ (i.e., $\delta^{j,k}(\boldsymbol{x}) \geq 0$) if $s_j = 0$ and will be user $j$ (i.e., $\delta^{j,k}(\boldsymbol{x}) \leq 0$) if $s_j = s_k$. At the same time, Property 4 of Theorem 3 tells us that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $s_j$ when users $j$ and $k$ are statistically identical. Therefore, we can conclude that the optimal choice will switch from user $k$ to user $j$ when $s_j$ increases from 0 to $s_k$ solely.

□

## 3.4 Whittle's Index Policy

Whittle's index policy is a well-known low-complexity heuristic that demonstrates a strong performance in many problems that belong to RMAB [66–68]. In this section, we develop Whittle's index policy for PP. We first present the general procedures we adopt to obtain Whittle's index.

- We first formulate a relaxed version of PP and apply the Lagrangian approach.
- Then, we decouple the problem of minimizing the Lagrangian function into N decoupled problems, each of which only considers a single user. By casting the decoupled problem into an MDP, we investigate the structural properties and performance of the optimal policy.
- Leveraging the results above and under a simple condition, we establish the indexability of the decoupled problem.

- Finally, we obtain the expression of Whittle's index by solving the Bellman equation.

### 3.4.1 Relaxed Problem

The first step in obtaining Whittle's index is to formulate the relaxed problem (RP). More precisely, instead of requiring the limit on the number of updates allowed per transmission attempt to be met in each time slot, we relax the constraint such that the limit is not violated in an average sense. Then, RP can be formulated as

$$\begin{aligned} \underset{\phi \in \Phi}{\arg\min} \quad & \bar{\Delta}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N} f_i(s_{i,t}) \right) \\ \text{subject to} \quad & \bar{\rho}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N} a_{i,t} \right) \leq M. \end{aligned} \tag{3.4}$$

As RP is specified, we apply the Lagrangian approach. First of all, we write RP into its Lagrangian form.

$$\mathcal{L}(\lambda, \phi) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N} (f_i(s_{i,t}) + \lambda a_{i,t}) \right) - \lambda M,$$

where $\lambda \geq 0$ is the Lagrange multiplier. Then, we investigate the problem of minimizing the Lagrangian function. Since $\lambda M$ is independent of policies, we can ignore it. More precisely, we consider the following minimization problem

$$\underset{\phi \in \Phi}{\text{minimize}} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N} (f_i(s_{i,t}) + \lambda a_{i,t}) \right). \tag{3.5}$$

### 3.4.2 Decoupled Model

In this section, we formulate the decoupled problem and investigate its optimal policy. The decoupled model associated with each user follows the system model with $N = 1$. Since all the users share the same structure, we drop the user-dependent subscript $i$ for simplicity. Then, the decoupled problem can be formulated as

$$\underset{\phi \in \Phi'}{\text{minimize}} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_{\phi}\left(\sum_{t=0}^{T-1}(f(s_t)+\lambda a_t)\right), \tag{3.6}$$

where $\Phi'$ is the set of all causal policies when $N = 1$. We notice that problem (3.6) can be cast into the MDP $\mathcal{M}_1(\lambda, -1)$. We define $M = -1$ when there is no restriction on the number of updates allowed per transmission attempt.

We first investigate the structural properties of the optimal policy for $\mathcal{M}_1(\lambda, -1)$ when $\lambda$ is a given non-negative constant. We start with characterizing the corresponding value function $V(x)$.

**Corollary 4.** *For $\mathcal{M}_1(\lambda, -1)$, $V(x)$ is non-decreasing in $s$.*

*Proof.* The proof follows the same steps as in the proof of Lemma 2. The complete proof can be found in Appendix B.4. □

Equipped with the above corollary, we can characterize the structural properties of the optimal policy for (3.6).

**Proposition 3.** *The optimal policy for the decoupled problem is a threshold policy with the following properties.*

- *The optimal policy can be fully captured by* $\boldsymbol{n} = (n_0, n_1)$*. More precisely, when the system is at state* $(s, \hat{r})$*, it is optimal to make a transmission attempt only when* $s \geq n_{\hat{r}}$*.*

- $n_0 \geq n_1 > 0$.

*Proof.* We define $\Delta V(x) \triangleq V^1(x) - V^0(x)$, where $V^a(x)$ is the value function resulting from taking action $a$ at state $x$. Then, the optimal action at state $x$ is $a = 1$ if $\Delta V(x) < 0$, and $a = 0$ is optimal otherwise. We use Corollary 4 to characterize the sign of $\Delta V(x)$. The complete proof can be found in Appendix B.5. □

In the following, we evaluate the performance of the threshold policy in Proposition 3. More precisely, we calculate the expected AoII $\bar{\Delta}_{\boldsymbol{n}}$ and the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ resulting from the adoption of threshold policy $\boldsymbol{n}$. We will see in the following that $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$ are essential for establishing the indexability and obtaining the expression of Whittle's index.

**Proposition 4.** *Under threshold policy* $\boldsymbol{n} = (n_0, n_1)$*,*

$$\bar{\Delta}_{\boldsymbol{n}} = \pi_0 p \left[ \sum_{k=1}^{n_1-1} f(k)(1-p)^{k-1} + (1-p)^{n_1-1} \left( \sum_{k=n_1}^{n_0-1} f(k) c_1^{k-n_1} + c_1^{n_0-n_1} \sum_{k=n_0}^{+\infty} f(k) c_2^{k-n_0} \right) \right],$$

$$\bar{\rho}_{\boldsymbol{n}} = \pi_0 p (1-p)^{n_1-1} \left[ \frac{\gamma}{1-c_1} + c_1^{n_0-n_1} \left( \frac{1}{1-c_2} - \frac{\gamma}{1-c_1} \right) \right],$$

*where*

$$\pi_0 = \frac{1}{2 + p(1-p)^{n_1-1} \left[ \frac{1}{1-c_1} - \frac{1}{p} + c_1^{n_0-n_1} \left( \frac{1}{1-c_2} - \frac{1}{1-c_1} \right) \right]},$$

$c_1 = (1-\gamma)(1-p) + \gamma\alpha$*, and* $c_2 = (1-\gamma)\beta + \gamma\alpha$.

*Proof.* We notice that the dynamic of AoII under the threshold policy can be fully captured by a discrete-time Markov chain (DTMC). Then, combined with the fact that $\hat{r}$ is an independent

Bernoulli random variable, we can obtain the desired results from the stationary distribution of the induced DTMC. The complete proof can be found in Appendix B.6. □

As $f(\cdot)$ can be any non-decreasing function, $\bar{\Delta}$ can grow indefinitely. Thus, it is necessary to require that there exists at least one threshold policy that causes a finite $\bar{\Delta}$. By noting that $1-p \geq c_1 \geq c_2$, we have

$$\begin{aligned}\bar{\Delta} &\geq \pi_0 p \left[\sum_{k=1}^{n_1-1} f(k)c_2^{k-1} + c_2^{n_1-1}\left(\sum_{k=n_1}^{n_0-1} f(k)c_2^{k-n_1} + c_2^{n_0-n_1}\sum_{k=n_0}^{+\infty} f(k)c_2^{k-n_0}\right)\right] \\ &= \pi_0 p \left(\sum_{k=1}^{+\infty} f(k)c_2^{k-1}\right).\end{aligned}$$

The equality is achieved when $n_0 = n_1 = 1$. Then, we can conclude that it is sufficient to require $\sum_{k=1}^{+\infty} f(k)c_2^{k-1} < +\infty$. This will be the underlying assumption throughout the rest of this chapter.

### 3.4.3 Indexability

In this section, we establish the indexability of the decoupled problem, which ensures the existence of Whittle's index. We start with the definition of indexability.

**Definition 2** (Indexability)**.** *The decoupled problem is indexable if the set of states in which $a = 0$ is the optimal action increases with $\lambda$, that is,*

$$\lambda' < \lambda \Longrightarrow D(\lambda') \subseteq D(\lambda),$$

*where $D(\lambda)$ is the set of states in which $a = 0$ is optimal when Lagrange multiplier $\lambda$ is adopted.*

The Lagrange multiplier $\lambda$ can be viewed as a cost associated with each transmission attempt. Intuitively, as $\lambda$ increases, the base station should stay idle (i.e., $a = 0$) for a longer time until $s$ becomes large enough to offset the cost. Although it is intuitively correct that the decoupled problem is indexable, the indexability is hard to establish as the optimal policy is characterized by two thresholds. Thus, Whittle's index does not necessarily exist. However, the indexability can be established when the following condition is satisfied

$$p_{e,i}^0 = 0 \quad for\ 1 \leq i \leq N. \tag{3.7}$$

**Remark 4.** *Equation* (3.7) *only requires the estimate* $\hat{r}_i$ *to be perfect when* $\hat{r}_i = 0$. *In the case of* $\hat{r}_i = 1$, *we still allow the estimate to be inaccurate.*

When (3.7) is satisfied, Propositions 3 and 4 reduce to the following

**Corollary 5.** *When* (3.7) *is satisfied, the optimal policy for the decoupled problem* (3.6) *is the threshold policy* $\boldsymbol{n} = (+\infty, n)$. *The corresponding* $\bar{\Delta}_{\boldsymbol{n}}$ *and* $\bar{\rho}_{\boldsymbol{n}}$ *are*

$$\bar{\Delta}_{\boldsymbol{n}} = \pi_0 p \left( \sum_{k=1}^{n-1} f(k)(1-p)^{k-1} + (1-p)^{n-1} \sum_{k=n}^{+\infty} f(k) c_1^{k-n} \right),$$

$$\bar{\rho}_{\boldsymbol{n}} = \pi_0 p (1-p)^{n-1} \left( \frac{\gamma}{1-c_1} \right),$$

*where*

$$\pi_0 = \frac{1}{2 + p(1-p)^{n-1} \left( \frac{1}{1-c_1} - \frac{1}{p} \right)}.$$

*Proof.* We continue with the same notations as in the proof of Propositions 3 and 4. It is sufficient to show that $n_0 = +\infty$. To this end, we consider the state $x = (s, 0)$. By following the same

steps as in the proof of Proposition 3, we have

$$\Delta V(s,0) = \lambda \geq 0.$$

Therefore, it is optimal to stay idle (i.e., $a = 0$) at state $x = (s, 0)$ for any $s \geq 0$. Equivalently, $n_0 = +\infty$. Then, the corresponding $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$ can be calculated as a special case of Proposition 4 where $n_0 = +\infty$, $n_1 = n$, and $p_e^0 = 0$. □

Leveraging Corollary 5, we can establish the indexability of the decoupled problem.

**Proposition 5.** *The decoupled problem is indexable when* (3.7) *is satisfied.*

*Proof.* According to [69, Proposition 2.2], we only need to verify that the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ is strictly decreasing in $n$. From Corollary 5, we have

$$\bar{\rho}_{\boldsymbol{n}} = \frac{\gamma\left(\frac{p}{1-c_1}\right)}{\frac{2}{(1-p)^{n-1}} + \left(\frac{p}{1-c_1} - 1\right)}.$$

As $\frac{1}{2} < 1 - p < 1$, we can easily verify that $\bar{\rho}_{\boldsymbol{n}}$ is strictly decreasing in $n$. Thus, the decoupled problem is indexable when (3.7) is satisfied. □

### 3.4.4 Whittle's Index Policy

In this section, we proceed with finding the expression of Whittle's index and defining Whittle's index policy. First of all, we give the definition of Whittle's index.

**Definition 3** (Whittle's index)**.** *When the decoupled problem is indexable, Whittle's index at state $x$ is defined as the infimum $\lambda$, such that both actions are equally desirable. Equivalently, Whittle's*

*index at state $x$ is defined as the infimum $\lambda$ such that $V^0(x) = V^1(x)$.*

Let us denote by $W_x$ the Whittle's index at state $x$. Then, the expression of Whittle's index is given by the following Proposition.

**Proposition 6.** *When (3.7) is satisfied, Whittle's index is*

$$W_x = \begin{cases} 0 & x = (0, \hat{r}) \text{ or } x = (s, 0), \\ \dfrac{(1-c_1)\sum\limits_{k=s+1}^{+\infty} f(k) c_1^{k-s-1} - \bar{\Delta}_s}{\dfrac{(1-c_1)(1-p) - \gamma(1-p-\alpha)}{c_1(1-p-\alpha)}} + \bar{\rho}_s & x = (s, 1), \end{cases}$$

*where $s > 0$ and $c_1 = (1-\gamma)(1-p) + \gamma\alpha$. $\bar{\Delta}_s$ and $\bar{\rho}_s$ are the expected AoII and the expected transmission rate when threshold policy $\boldsymbol{n} = (+\infty, s)$ is adopted, respectively. At the same time, $W_x$ is non-negative and is non-decreasing in $s$.*

*Proof.* Whittle's indexes at state $x = (0, \hat{r})$ and $x = (s, 0)$ are obtained easily from the proof of Proposition 3. For state $x = (s, 1)$, we first use backward induction to calculate the expressions of some value functions. Then, the expression of Whittle's index can be obtained from its definition. The complete proof can be found in Appendix B.7. □

**Definition 4** (Whittle's index policy)**.** *At any state $\boldsymbol{x} = (x_1, x_2, ..., x_N)$, the base station will transmit the updates from $M$ users with the largest $W_{x_i}$. The ties are broken arbitrarily. $W_{x_i}$ is calculated using Proposition 6 with the parameters of user $i$.*

**Remark 5.** *Whittle's index policy possesses the structural properties detailed in Corollary 3.*

- *The first two properties can be verified by noting that $W_{x_i} \geq 0$ and the equality holds when $\hat{r}_i = 0$ or $s_i = 0$. At the same time, $W_{x_i}$ is non-decreasing in $\hat{r}_i$.*

- *The third and fourth properties can be verified by noting that $W_{x_i}$ is non-decreasing in $s_i$.*

- *For the last property, we first notice that $W_{x_j} = W_{x_k}$ when users $j$ and $k$ are statistically identical and $x_j = x_k$. Then, the property can be verified by noting that $W_{x_i}$ is non-decreasing in both $s_i$ and $\hat{r}_i$.*

## 3.5 Optimal Policy for Relaxed Problem

In this section, we provide an efficient algorithm to obtain the optimal policy for RP, based on which we will develop another scheduling policy for PP in the next section that is free from indexability. At the same time, the performance of the optimal policy for RP forms a universal lower bound because the following ordering holds

$$\bar{\Delta}_{AoII}^{RP} \leq \bar{\Delta}_{AoII}^{PP},$$

where $\bar{\Delta}_{AoII}^{RP}$ and $\bar{\Delta}_{AoII}^{PP}$ are the minimal expected AoII of RP and PP, respectively.

**Remark 6.** *Note that the optimal policy for RP may not necessarily be a valid policy for PP, as the transmitter may transmit more than $M$ updates in one transmission attempt under RP-optimal policy.*

To solve RP, we follow the discussion in Section 3.4.1. More precisely, we take the Lagrangian approach and consider the problem reported in (3.5). We will see in the following discussion that the optimal policy for RP can be characterized by the optimal policies for problem (3.5). Therefore, we first cast problem (3.5) into the MDP $\mathcal{M}_N(\lambda, -1)$. However, the optimal policy for $\mathcal{M}_N(\lambda, -1)$ is difficult to obtain because the state space is infinite. Even though we

can make the state space finite by imposing an upper limit on the value of $s$, the state space and the action space grow exponentially with the number of users in the system. To overcome the difficulty, we investigate the optimal policy for $\mathcal{M}_1^i(\lambda, -1)$ where $1 \leq i \leq N$. The superscript $i$ means that the only user in the system is user $i$. We will show later that the optimal policy for $\mathcal{M}_N(\lambda, -1)$ can be fully characterized by the optimal policies for $\mathcal{M}_1^i(\lambda, -1)$ where $1 \leq i \leq N$.

### 3.5.1 Optimal Policy for Single User

In this section, we tackle the problem of finding the optimal policy for $\mathcal{M}_1^i(\lambda, -1)$. Since the users share the same structure, we ignore the superscript $i$ for simplicity. To find the optimal policy, we first use the approximating sequence method (ASM) introduced in [52] to make the state space finite. More precisely, we impose $s \leq m$ where $m$ is a predetermined upper limit. The state transition probabilities $P'_{s,s'}(a, \hat{r})$ are modified in the following way

$$P'_{s,s'}(a, \hat{r}) = \begin{cases} P_{s,s'}(a, \hat{r}) & if \;\; s' < m, \\ P_{s,s'}(a, \hat{r}) + \sum_{z>m} P_{s,z}(a, \hat{r}) & if \;\; s' = m. \end{cases} \tag{3.8}$$

The action space and the instant cost remain unchanged. Then, we can apply relative value iteration (RVI) with convergence criteria $\epsilon$ to obtain the optimal policy. We notice that $\mathcal{M}_1(\lambda, -1)$ coincides with the decoupled model studied in Section 3.4.2. Hence, we can utilize the threshold structure of the optimal policy to improve RVI. To this end, we class a state as active if the optimal action at this state is $a = 1$. Then, the threshold structure detailed in Proposition 3 tells us the following. For any state $x$, if there exists an active state $x_1$ with $s_1 \leq s$ and $\hat{r}_1 \leq \hat{r}$, then $x$ must also be active. Hence, we can determine the optimal action at state $x$ immediately

**Algorithm 3** Improved Relative Value Iteration

**procedure** RELATIVEVALUEITERATION($\mathcal{M}$,$\epsilon$)
  Initialize $V_0(x) = 0$; $\nu = 0$
  Choose $x^{ref} \in \mathcal{X}$ arbitrarily
  **while** $V_\nu$ is not converged **do**
    **for** $x = (s, \hat{r}) \in \mathcal{X}$ **do**
      **if** $\exists$ active state $(s_1, \hat{r}_1)$ s.t. $s_1 \leq s$ and $\hat{r}_1 \leq \hat{r}$ **then**
        $a^*(x) = 1$
        $Q_{\nu+1}(x) = C(x, 1) + \sum_{x'} P_{xx'}(1) V_\nu(x')$
      **else**
        **for** $a \in \mathcal{A}$ **do**
          $H_{x,a} = C(x, a) + \sum_{x'} P_{xx'}(a) V_\nu(x')$
        $a^*(x) = \arg\min_a \{H_{x,a}\}$
        $Q_{\nu+1}(x) = H_{x,a^*}$
    **for** $x \in \mathcal{X}$ **do**
      $V_{\nu+1}(x) = Q_{\nu+1}(x) - Q_{\nu+1}(x^{ref})$
    $\nu = \nu + 1$
  **return** $\boldsymbol{n} \leftarrow a^*(x)$

instead of comparing all feasible actions. In this way, we can reduce the running time of RVI. The pseudocode for the improved RVI can be found in Algorithm 3 where $V_\nu$ converges when the maximum difference between the results of two consecutive iterations is less than $\epsilon$. A similar technique is also presented in [48].

For $\mathcal{M}_1(\lambda, -1)$, when problem (3.7) is satisfied, Whittle's index exists and can be calculated efficiently using Proposition 6. Therefore, we can obtain the optimal policy using Whittle's index and further reduce the computational complexity. To this end, we denote by $\boldsymbol{n}_\lambda$ the optimal policy for $\mathcal{M}_1(\lambda, -1)$ and present the following proposition

**Proposition 7.** *When (3.7) is satisfied, the optimal policy for $\mathcal{M}_1(\lambda, -1)$ is $\boldsymbol{n}_\lambda = (+\infty, n)$ where $n$ is given by*

$$n = \begin{cases} 1 & \lambda = 0, \\ \max\{s \in \mathbb{N}_0 : W_s \leq \lambda\} + 1 & \lambda > 0. \end{cases}$$

*$W_s$ is the Whittle's index at state $(s, 1)$.*

*Proof.* We notice that $\mathcal{M}_1(\lambda, -1)$ coincides with the decoupled model studied in Section 3.4.2. Then, we show the optimal action for each state with $\hat{r} = 1$ using the definition of Whittle's index and the fact that the decoupled problem is indexable when (3.7) is satisfied. The complete proof can be found in Appendix B.8. □

In the following, we provide a randomized policy that is also optimal for $\mathcal{M}_1(\lambda, -1)$. We will see later that the randomized policy is the key to obtaining the optimal policy for RP.

**Theorem 4.** *There exist two deterministic policies $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$, which are both optimal for $\mathcal{M}_1(\lambda, -1)$. We consider the following randomized policy $\boldsymbol{n}_\lambda$: every time the system reaches state $(0, 0)$, the base station will make the choice between $\boldsymbol{n}_{\lambda_-}$ with probability $\mu$ and $\boldsymbol{n}_{\lambda_+}$ with probability $1 - \mu$. The chosen policy will be followed until the next choice. Then, the randomized policy $\boldsymbol{n}_\lambda$ is optimal for $\mathcal{M}_1(\lambda, -1)$ under any $\mu \in [0, 1]$.*

*Proof.* We show that our system verifies the assumptions given in [50]. Then, leveraging the characteristics of our system, we can obtain the optimal randomized policy. The complete proof can be found in Appendix B.9. □

In practice, we approximate $\lambda_+ \approx \lambda + \xi$ and $\lambda_- \approx \lambda - \xi$ where $\xi$ is a small perturbation. Then, the deterministic policies $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ can be obtained by following the discussion at the beginning of this subsection. Note that, in most cases, $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ are the same.

### 3.5.2 Optimal Policy for RP

In this section, we characterize the optimal policy for RP. Let us denote by $V(\boldsymbol{x})$ and $V^i(x_i)$ the value functions of $\mathcal{M}_N(\lambda, -1)$ and $\mathcal{M}_1^i(\lambda, -1)$, respectively. Then, we can prove the following

**Proposition 8.** *$V(\boldsymbol{x}) = \sum_{i=1}^N V^i(x_i)$ where $\boldsymbol{x} = (x_1, ..., x_N)$. In other words, the policy, under which each user adopts its own optimal policy, is optimal for $\mathcal{M}_N(\lambda, -1)$.*

*Proof.* We show $V(\boldsymbol{x}) = \sum_{i=1}^N V^i(x_i)$ by comparing the Bellman equations they must satisfy. The complete proof can be found in Appendix B.10. □

We denote the optimal policy for $\mathcal{M}_N(\lambda, -1)$ as $\phi_\lambda = [\boldsymbol{n}_{\lambda,1}, ..., \boldsymbol{n}_{\lambda,N}]$ where $\boldsymbol{n}_{\lambda,i}$ is the optimal policy for $\mathcal{M}_1^i(\lambda, -1)$. For simplicity, we define $\bar{\Delta}(\lambda)$ and $\bar{\rho}(\lambda)$ as the expected AoII and the expected transmission rate associated with $\phi_\lambda$, respectively. $\bar{\Delta}^i(\lambda)$ and $\bar{\rho}^i(\lambda)$ are defined analogously for user $i$ under policy $\boldsymbol{n}_{\lambda,i}$. We also define $\lambda^* \triangleq \inf\{\lambda > 0 : \bar{\rho}(\lambda) \leq M\}$. With Proposition 8 and the above definitions in mind, we proceed with constructing the optimal policy for RP.

**Theorem 5.** *The optimal policy for RP can be characterized by two deterministic policies $\phi_{\lambda_+^*} = [\boldsymbol{n}_{\lambda_+^*,1}, ..., \boldsymbol{n}_{\lambda_+^*,N}]$ and $\phi_{\lambda_-^*} = [\boldsymbol{n}_{\lambda_-^*,1}, ..., \boldsymbol{n}_{\lambda_-^*,N}]$ where $\boldsymbol{n}_{\lambda_+^*,i}$ and $\boldsymbol{n}_{\lambda_-^*,i}$ are both the optimal deterministic policies for $\mathcal{M}_1^i(\lambda^*, -1)$. Then, we mix $\phi_{\lambda_+^*}$ and $\phi_{\lambda_-^*}$ in the following way: for each user $i$, every time the user reaches state $(0, 0)$, the base station will make the choice between $\boldsymbol{n}_{\lambda_-^*,i}$ with probability $\mu_i$ and $\boldsymbol{n}_{\lambda_+^*,i}$ with probability $1 - \mu_i$. The chosen policy will be followed*

*by user $i$ until the next choice. Where $1 \leq i \leq N$, the $\mu_i$ is chosen in such a way as to satisfy*

$$\sum_{i=1}^{N} \bar{\rho}^i(\lambda^*) = \sum_{i=1}^{N} \left( \mu_i \bar{\rho}^i(\lambda_-^*) + (1-\mu_i)\bar{\rho}^i(\lambda_+^*) \right) = M. \tag{3.9}$$

*Then, the mixed policy, denoted by $\phi_{\lambda^*}$, is optimal for RP.*

*Proof.* According to [50, Lemma 3.10], a policy is optimal for RP if

1. It is optimal for $\mathcal{M}_N(\lambda^*, -1)$;

2. The resulting expected transmission rate is equal to $M$.

Then, we construct such a policy using Theorem 4 and Proposition 8. The complete proof can be found in Appendix B.11. □

Since we approximate $\lambda_+^* \approx \lambda^* + \xi$ and $\lambda_-^* \approx \lambda^* - \xi$ in practice, $\bar{\rho}^i(\lambda_+^*) \leq \bar{\rho}^i(\lambda_-^*)$ for all $i$ according to the monotonicity given by [50, Lemma 3.4]. Combining with the definition of $\lambda^*$, we must have $\bar{\rho}(\lambda_+^*) \leq M < \bar{\rho}(\lambda_-^*)$. Therefore, we can always find $\mu_i$'s that realize (3.9). In this chapter, we choose

$$\mu_i = \mu = \frac{M - \bar{\rho}(\lambda_+^*)}{\bar{\rho}(\lambda_-^*) - \bar{\rho}(\lambda_+^*)}, \quad for\ 1 \leq i \leq N. \tag{3.10}$$

Then, we describe the algorithm used to obtain the optimal policy for RP. As detailed in Theorem 5, it is essential to find $\lambda^*$. To this end, we recall that, for any user $i$ under given $\lambda$, the optimal deterministic policy $\boldsymbol{n}_{\lambda,i}$ can be obtained using the results in Section 3.5.1 and the resulting expected transmission rate $\bar{\rho}^i(\lambda)$ is given by Proposition 4. Since $\bar{\rho}^i(\lambda)$ is non-increasing in $\lambda$ for all $i$ according to [50, Lemma 3.4], $\bar{\rho}(\lambda) = \sum_{i=1}^{N} \bar{\rho}^i(\lambda)$ is also non-increasing in $\lambda$. Hence, we can regard $\bar{\rho}(\lambda)$ as a non-increasing function of $\lambda$. Then, according to the definition of $\lambda^*$, we can

**Algorithm 4** Bisection Search

**procedure** BISECTIONSEARCH($\mathcal{M}_N(\lambda, -1)$, $M$, $\xi$, $\epsilon$)
  Initialize $\lambda_- = 0$; $\lambda_+ = 1$
  $\phi_{\lambda_+} \leftarrow (\mathcal{M}_N(\lambda_+, -1), \epsilon)$ using Section 3.5.1 and Proposition 8
  $\bar{\rho}(\lambda_+) \leftarrow \phi_{\lambda_+}$ using Proposition 4
  **while** $\bar{\rho}(\lambda_+) \geq M$ **do**
    $\lambda_- = \lambda_+$; $\lambda_+ = 2\lambda_+$
    $\phi_{\lambda_+} \leftarrow (\mathcal{M}_N(\lambda_+, -1), \epsilon)$ using Section 3.5.1 and Proposition 8
    $\bar{\rho}(\lambda_+) \leftarrow \phi_{\lambda_+}$ using Proposition 4
  **while** $\lambda_+ - \lambda_- \geq 2\xi$ **do**
    $\lambda = \frac{\lambda_+ + \lambda_-}{2}$
    $\phi_\lambda \leftarrow (\mathcal{M}_N(\lambda, -1), \epsilon)$ using Section 3.5.1 and Proposition 8
    $\bar{\rho}(\lambda) \leftarrow \phi_\lambda$ using Proposition 4
    **if** $\bar{\rho}(\lambda) > M$ **then**
      $\lambda_- = \lambda$
    **else**
      $\lambda_+ = \lambda$
  **return** $(\lambda_+^*, \lambda_-^*) \leftarrow (\lambda_+, \lambda_-)$

use the Bisection search to obtain $\lambda^*$ efficiently. The main steps can be summarized as follows.

1. Initialize $\lambda_- = 0$ and $\lambda_+ = 1$.

2. Do $\lambda_- = \lambda_+$ and $\lambda_+ = 2\lambda_+$ until $\bar{\rho}(\lambda_+) < M$.

3. Run Bisection search on the interval $[\lambda_-, \lambda_+]$ until the tolerance $2\xi$ is met.

Then, $\lambda_-^*$ and $\lambda_+^*$ can simply be the boundaries of the final interval. The pseudocode for the Bisection search can be found in Algorithm 4. After obtaining $\lambda_-^*$ and $\lambda_+^*$, the optimal policy $\phi_{\lambda^*}$ is detailed in Theorem 5 and the mixing probabilities $\mu_i$'s are given by (3.10).

**Remark 7.** *We recall that the optimal deterministic policy for each user can be characterized by two positive thresholds (i.e., $n_0, n_1 > 0$). Consequently, under RP-optimal policy, the base station will never choose the user at state $(0, \hat{r})$. Then, when $M$ increases, the expected transmission rate achieved by RP-optimal policy will saturate before $M$ reaches $N$. When the expected*

*transmission rate saturates, the RP-optimal policy is* $\phi^* = [\boldsymbol{n}_1, ..., \boldsymbol{n}_N]$ *where* $\boldsymbol{n}_i = (1, 1)$ *for* $1 \leq i \leq N$. *The saturation happens when* $M$ *is larger than or equal to the expected transmission rate achieved by* $\phi^*$.

## 3.6 Indexed Priority Policy

Although the performance of Whittle's index policy is famously good, it requires indexability, which is usually difficult to establish. In this section, based on the primal-dual heuristic introduced in [70], we develop a policy that does not require indexability and has comparable performance to Whittle's index policy. We start with presenting the primal-dual heuristic.

### 3.6.1 Primal-Dual Heuristic

The heuristic is based on the optimal primal and dual solution pair to the linear program associated with RP. To introduce the linear program, we define $\pi_{x_i}^{a_i}(\phi) \geq 0$ as the expected time that user $i$ is at state $x_i$ and action $a_i$ is taken according to policy $\phi$. Then, for any $\phi$, $\pi_{x_i}^{a_i}(\phi)$ must satisfy the following problems

$$\pi_{x_i}^0(\phi) + \pi_{x_i}^1(\phi) = \sum_{x_i'} \sum_{a_i'} P_{x_i', x_i}(a_i') \pi_{x_i'}^{a_i'}(\phi), \quad \forall x_i, i.$$

$$\sum_{x_i} \sum_{a_i} \pi_{x_i}^{a_i}(\phi) = 1, \quad \forall i.$$

The objective function of RP can be rewritten as

$$\underset{\phi \in \Phi}{\text{minimize}} \quad \sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi_{x_i}^{a_i}(\phi),$$

where $C(x_i) = f_i(s_i)$ is the instant cost at state $x_i$. The constraint on the expected transmission rate can be rewritten as

$$\sum_{i=1}^{N} \sum_{x_i} \pi_{x_i}^1(\phi) \leq M.$$

Thus, the linear program associated with RP can be formulated as the following

$$\begin{aligned}
\underset{\pi_{x_i}^{a_i}}{\text{minimize}} \quad & \sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi_{x_i}^{a_i} \\
\text{subject to} \quad & \pi_{x_i}^0 + \pi_{x_i}^1 - \sum_{x_i'} \sum_{a_i'} P_{x_i', x_i}(a_i') \pi_{x_i'}^{a_i'} = 0, \quad \forall x_i, i, \\
& \sum_{x_i} \sum_{a_i} \pi_{x_i}^{a_i} = 1, \quad \forall i, \\
& \sum_{i=1}^{N} \sum_{x_i} \pi_{x_i}^1 \leq M, \\
& \pi_{x_i}^{a_i} \geq 0, \quad \forall x_i, a_i, i.
\end{aligned} \tag{3.11}$$

The corresponding dual problem is

$$\begin{aligned}
\underset{\sigma, \sigma_i, \sigma_{x_i}}{\text{maximize}} \quad & \sum_{i=1}^{N} \sigma_i - M\sigma \\
\text{subject to} \quad & \sigma_{x_i} + \sigma_i - \sum_{x_i'} P_{x_i, x_i'}(0) \sigma_{x_i'} \leq C(x_i), \quad \forall x_i, i, \\
& \sigma_{x_i} + \sigma_i - \sum_{x_i'} P_{x_i, x_i'}(1) \sigma_{x_i'} - \sigma \leq C(x_i), \quad \forall x_i, i, \\
& \sigma \geq 0.
\end{aligned} \tag{3.12}$$

Let $\{\bar{\pi}_{x_i}^{a_i}\}$ and $\{\bar{\sigma}, \bar{\sigma}_i, \bar{\sigma}_{x_i}\}$ be the optimal primal and dual solution pair to the problems reported in (3.11) and (3.12). We define

$$\bar{\psi}_{x_i}^0 = \sum_{x_i'} P_{x_i,x_i'}(0)\bar{\sigma}_{x_i'} + C(x_i) - \bar{\sigma}_i - \bar{\sigma}_{x_i} \geq 0,$$

$$\bar{\psi}_{x_i}^1 = \sum_{x_i'} P_{x_i,x_i'}(1)\bar{\sigma}_{x_i'} + \bar{\sigma} + C(x_i) - \bar{\sigma}_i - \bar{\sigma}_{x_i} \geq 0.$$

For any state $\boldsymbol{x} = (x_1, ..., x_N)$, let $h(\boldsymbol{x}) = \sum_{i=1}^N \mathbb{1}_{\{\bar{\pi}_{x_i}^1 > 0\}}$. Then, the heuristic operates in the following way

- If $h(\boldsymbol{x}) \geq M$, the base station will choose the $M$ users with the largest $\bar{\psi}_{x_i}^0$ among the $h(\boldsymbol{x})$ users.
- If $h(\boldsymbol{x}) < M$, these $h(\boldsymbol{x})$ users are chosen by the base station. The base station will choose $M - h(\boldsymbol{x})$ additional users with the smallest $\bar{\psi}_{x_i}^1$.

However, linear programming (LP) is a very general technique and does not appear to take advantage of the special structure of the problem. Although there are algorithms for solving rational LP that take time polynomial in the number of variables and constraints, they run very slowly in practice [71]. For our problem, we notice that the users have separate activity areas that are linked through a common resource constraint. Therefore, PP can be solved using Dantzig-Wolfe decomposition. Even so, the problem is still computationally demanding when the system scales up. We recall that we solved the exact problem efficiently using MDP-specific algorithms in Section 3.5. It is more efficient because of the following reasons

- According to Proposition 8, we can decompose the problem into $N$ subproblems.

- For each subproblem, the threshold structure of the optimal policy is utilized to reduce the running time of RVI.
- As we will see later, the developed policy can be obtained directly from the result of RVI in practice.

In the following, we will translate the results in Section 3.5 into the optimal primal and dual solution pair and propose the indexed priority policy.

### 3.6.2 Indexed Priority Policy

We first define the Lagrangian function associated with (3.11).

$$\mathcal{L}(\pi_{x_i}^{a_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}) = \left( \sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi_{x_i}^{a_i} \right) + \sum_{i, x_i} \sigma_{x_i} \left( \sum_{x_i'} \sum_{a_i'} P_{x_i', x_i}(a_i') \pi_{x_i'}^{a_i'} - \pi_{x_i}^0 - \pi_{x_i}^1 \right) + \sum_{i=1}^{N} \sigma_i \left( 1 - \sum_{x_i} \sum_{a_i} \pi_{x_i}^{a_i} \right) + \sigma \left( \sum_{i=1}^{N} \sum_{x_i} \pi_{x_i}^1 - M \right) - \sum_{i, x_i, a_i} \psi_{x_i}^{a_i} \pi_{x_i}^{a_i}.$$

Then, the corresponding Lagrangian dual function is

$$g(\sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}) = \inf_{\pi_{x_i}^{a_i}} \mathcal{L}(\pi_{x_i}^{a_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}).$$

Let $\pi_{x_i}$ be the expected time that user $i$ is at state $x_i$ caused by the adoption of $\phi_{\lambda^*}$, where $\phi_{\lambda^*}$ is the optimal policy detailed in Theorem 5. Then, we define $\{\pi_{x_i}^{a_i}\}$ as follows.

- For state $x_i$ where the randomization happens (randomization happens when the actions suggested by the two optimal deterministic policies are different), $\pi_{x_i}^0 = a_{\boldsymbol{n}_{\lambda_-^*, i}}(x_i)(1 - \mu_i)\pi_{x_i} + a_{\boldsymbol{n}_{\lambda_+^*, i}}(x_i)\mu_i \pi_{x_i}$ and $\pi_{x_i}^1 = \pi_{x_i} - \pi_{x_i}^0$ where $\mu_i$ is given by (3.10) and $a_{\boldsymbol{n}_{\lambda, i}}(x_i)$ is

the action suggested by $\boldsymbol{n}_{\lambda,i}$ at state $x_i$.

- For other values of $x_i$, we have $\pi^0_{x_i} = (1 - a_{\boldsymbol{n}_{\lambda^*,i}}(x_i))\pi_{x_i}$ and $\pi^1_{x_i} = \pi_{x_i} - \pi^0_{x_i}$.

We also define $\sigma = \lambda^*$, $\sigma_i = \theta_i$, and $\sigma_{x_i} = V^i(x_i)$ where $\lambda^*$ is specified in Section 3.5.2, $\theta_i$ is the optimal value of $\mathcal{M}^i_1(\lambda^*, -1)$, and $V^i(x_i)$ is the value function associated with $\mathcal{M}^i_1(\lambda^*, -1)$. Lastly, we define $\{\psi^{a_i}_{x_i}\}$ as follows.

$$\psi^0_{x_i} = \sum_{x'_i} P_{x_i,x'_i}(0)\sigma_{x'_i} + C(x_i) - \sigma_i - \sigma_{x_i}.$$

$$\psi^1_{x_i} = \sum_{x'_i} P_{x_i,x'_i}(1)\sigma_{x'_i} + \sigma + C(x_i) - \sigma_i - \sigma_{x_i}.$$

Then, we can prove the following proposition.

**Proposition 9.** *$\{\pi^{a_i}_{x_i}\}$ and $\{\sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i}\}$ are primal and dual solutions to (3.11), respectively.*

*Proof.* Since (3.11) is linear and strictly feasible, it is sufficient to show that $\{\sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i}\}$ and $\{\pi^{a_i}_{x_i}\}$ verify the KKT conditions, which can be expressed as the following four conditions.

1. Primal feasibility: the constraints in (3.11) are satisfied.
2. Dual feasibility: $\sigma \geq 0$ and $\psi^{a_i}_{x_i} \geq 0$ for all $x_i$, $a_i$, and $i$.
3. Complementary slackness: $\sigma\big(\sum_{i=1}^N \sum_{x_i} \pi^1_{x_i} - M\big) = 0$ and $\psi^{a_i}_{x_i}\pi^{a_i}_{x_i} = 0$ for all $x_i$, $a_i$, and $i$.
4. Stationarity: the gradient of $\mathcal{L}(\pi^{a_i}_{x_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i})$ with respect to $\{\pi^{a_i}_{x_i}\}$ vanishes.

Apparently, the first condition is satisfied by $\{\pi^{a_i}_{x_i}\}$. For the second condition, $\sigma \geq 0$ since $\sigma = \lambda^* \geq 0$ by definition. For $\psi^{a_i}_{x_i}$, we can verify that $\psi^{a_i}_{x_i} = V^{i,a_i}(x_i) - V^i(x_i)$ where $V^{i,a_i}(x_i)$

is the value function resulting from taking action $a_i$ at state $x_i$. Then, the non-negativity is guaranteed by the Bellman equation. For the third condition, the first term is zero because we choose the $\mu_i$'s given by (3.10). For the second term, we recall that $\psi_{x_i}^{a_i} = V^{i,a_i}(x_i) - V^i(x_i)$. According to the definition of $\pi_{x_i}^{a_i}$, we know $V^i(x_i) = V^{i,a_i}(x_i)$ if $\pi_{x_i}^{a_i} > 0$. Combined together, we can conclude that $\psi_{x_i}^{a_i} = 0$ when $\pi_{x_i}^{a_i} > 0$. Thus, the third condition is satisfied. For the last condition, setting the gradient equal to zero yields a system of linear equations. More precisely, for each $x_i$ and $1 \leq i \leq N$

$$\begin{cases} \sum_{x_i'} P_{x_i,x_i'}(0)\sigma_{x_i'} + C(x_i) = \sigma_{x_i} + \sigma_i + \psi_{x_i}^0. \\ \sum_{x_i'} P_{x_i,x_i'}(1)\sigma_{x_i'} + \sigma + C(x_i) = \sigma_{x_i} + \sigma_i + \psi_{x_i}^1. \end{cases}$$

Then, $\{\sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}\}$ verifies the system of linear equations by definition. Since all conditions are satisfied, we can conclude our proof. □

According to Proposition 9, we know that $\{\pi_{x_i}^{a_i}\}$ and $\{\sigma, \sigma_i, \sigma_{x_i}\}$ defined above are the optimal solutions to problems (3.11) and (3.12), respectively. As the optimal solutions are obtained, we can adopt the heuristic detailed in Section 3.6.1.

The heuristic can be expressed equivalently as an index policy. To this end, we define the index $I_{x_i}$ for state $x_i$ as

$$I_{x_i} \triangleq \bar{\psi}_{x_i}^0 - \bar{\psi}_{x_i}^1.$$

According to the complementary slackness, $I_{x_i}$ can be reduced to the following.

- For state $x_i$ such that $\bar{\pi}_{x_i}^1 > 0$ and $\bar{\pi}_{x_i}^0 = 0$, we have $\bar{\psi}_{x_i}^1 = 0$. Therefore, $I_{x_i} = \bar{\psi}_{x_i}^0 \geq 0$.

- For state $x_i$ such that $\bar{\pi}^1_{x_i} > 0$ and $\bar{\pi}^0_{x_i} > 0$, we have $\bar{\psi}^1_{x_i} = \bar{\psi}^0_{x_i} = 0$. Therefore, $I_{x_i} = 0$.

- For state $x_i$ such that $\bar{\pi}^1_{x_i} = 0$ and $\bar{\pi}^0_{x_i} > 0$, we have $\bar{\psi}^0_{x_i} = 0$. Therefore, $I_{x_i} = -\bar{\psi}^1_{x_i} \leq 0$.

We can show that $I_{x_i}$ possesses the following properties.

**Proposition 10.** *For $1 \leq i \leq N$, $I_{x_i} \geq -\lambda^*$ for any $x_i$. The equality holds when $\hat{r}_i = p^0_{e,i} = 0$ or $s_i = 0$. At the same time, $I_{x_i}$ is non-decreasing in both $s_i$ and $\hat{r}_i$.*

*Proof.* We notice that $I_{x_i}$ can be expressed as a function of $V^i(x_i)$ and $\lambda^*$. Also, $\mathcal{M}^i_1(\lambda^*, -1)$ coincides with the decoupled model studied in Section 3.4.2. Then, we can verify the properties of $I_{x_i}$ using the results in Section 3.4.2. The complete proof can be found in Appendix B.12. □

Comparing with the heuristic detailed in Section 3.6.1, we can define the indexed priority policy.

**Definition 5** (Indexed priority policy)**.** *At any state $\boldsymbol{x} = (x_1, x_2, ..., x_N)$, the base station will transmit the updates from $M$ users with the largest $I_{x_i}$. The ties are broken arbitrarily.*

**Remark 8.** *Indexed priority policy belongs to the class of priority policies introduced in [72]. These priority policies are asymptotically optimal when certain conditions are satisfied.*

**Remark 9.** *Indexed priority policy possesses the structural properties detailed in Corollary 3.*

- *The first two properties can be verified by noting that $I_{x_i} \geq -\lambda^*$ and the equality holds when $\hat{r}_i = p^0_{e,i} = 0$ or $s_i = 0$. At the same time, $I_{x_i}$ is non-decreasing in $\hat{r}_i$.*

- *The third and fourth properties can be verified by noting that $I_{x_i}$ is non-decreasing in $s_i$.*

- *For the last property, we first notice that $I_{x_j} = I_{x_k}$ when users $j$ and $k$ are statistically identical and $x_j = x_k$. Then, the property can be verified by noting that $I_{x_i}$ is non-decreasing in both $s_i$ and $\hat{r}_i$.*

We notice that $\theta_i$'s and $C(x_i)$'s are canceled out by the definition of $I_{x_i}$. Therefore, $I_{x_i}$ can be calculated using $\lambda^*$ and the value function of $\mathcal{M}_1^i(\lambda^*, -1)$. In practice, we can use either $\lambda_-^*$ or $\lambda_+^*$ to approximate $\lambda^*$, and the value function can be approximated by the result of the RVI detailed in Section 3.5.1. Since the state space is infinite, we only calculate a finite number of $V^i(x_i)$, the number of which depends on the truncation parameter $m$ of ASM. Meanwhile, the probabilities $P_{x_i,x_i'}(a_i)$ in $I_{x_i}$ are modified according to (3.8).

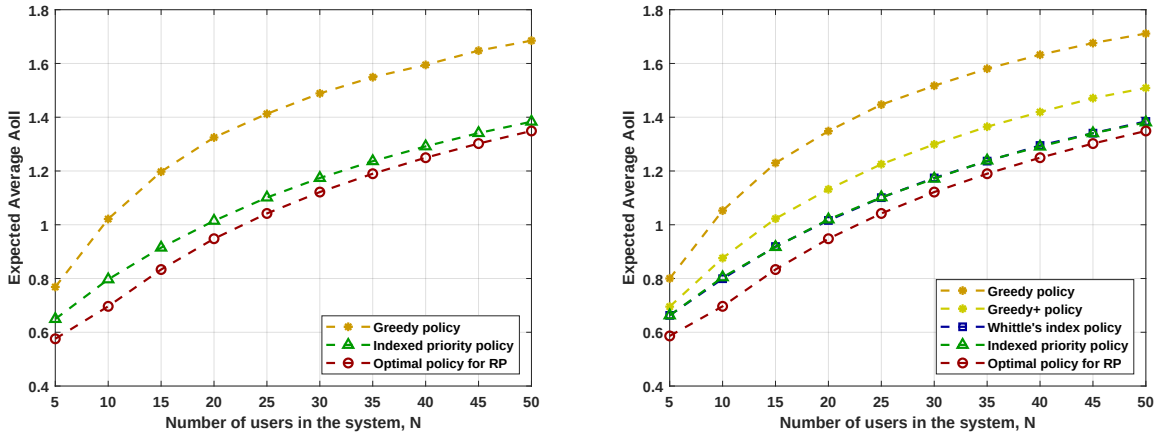
## 3.7 Numerical Results

In this section, we provide numerical results to showcase the performance of the developed scheduling policies. To eliminate the effect of $N$, we plot the expected average AoII. In particular, we provide the expected average AoII achieved by the indexed priority policy and Whittle's index policy when $M = 1$. The policies are calculated using the results detailed in Sections 3.4-3.6. When obtaining the indexed priority policy, we set the tolerance in the Bisection search to $\xi = 0.005$. Meanwhile, we choose the truncation parameter in ASM $m = 800$ and the convergence criteria in RVI $\epsilon = 0.01$. We notice that the calculation of Whittle's index involves an infinite sum. In practice, we approximate the result by replacing $+\infty$ with a large enough number $k_{max}$. Here, we choose $k_{max} = 800$. For both scheduling policies, the resulting expected average AoII is obtained via simulations. Each data point is the average of 15 runs with 15,000 time slots considered in each run.

We also compare the developed policies with the optimal policy for RP, which can be calculated by following the discussion in Section 3.5.2. We adopt the same choices of parameters as we used to obtain the developed policies. The corresponding performance is calculated using Proposition 4. Like before, the infinite sum is approximated by replacing $+\infty$ with $k_{max} = 800$. We also provide the expected average AoII achieved by the Greedy policy to show the performance advantages of the developed policies. When the Greedy policy is adopted, the base station always chooses the user with the largest AoII. The resulting expected average AoII is obtained via the same simulations as applied to the developed policies.

Figures 3.2 and 3.3 illustrate the performance when the source processes have different dynamics and when each user's communication goal is different, respectively. Figure 3.2a shows the performance when $p_i = 0.05 + \frac{0.4(i-1)}{N-1}$ for $1 \leq i \leq N$. For other parameters, the users make the same choices. More precisely, $f_i(s) = s$, $\gamma_i = 0.6$, and $p^0_{e,i} = p^1_{e,i} = 0.1$ for $1 \leq i \leq N$. Figure 3.3a provides the performance when $f_i(s) = s^{0.5+\frac{i-1}{N-1}}$ for $1 \leq i \leq N$. Same as before, the users make the same choices for other parameters. More precisely, $p_i = 0.3$, $\gamma_i = 0.6$, and $p^0_{e,i} = p^1_{e,i} = 0.1$ for $1 \leq i \leq N$. In Figures 3.2b and 3.3b, we force $p^0_{e,i} = 0$ for all users to ensure the existence of Whittle's index. Other choices remain the same as in Figures 3.2a and 3.3a. According to Corollary 3, the optimal policy will never choose the user with $\hat{r} = p^0_e = 0$ unless it is to break the tie. Therefore, in Figures 3.2b and 3.3b, we also consider the Greedy+ policy where the base station always chooses the user with the largest AoII among the users with $\hat{r} = 1$. The resulting expected average AoII is obtained via the same simulations as applied to the Greedy policy.

Figure 3.4 shows the performance in systems where the parameters for each user are generated uniformly and randomly within their ranges. In Figure 3.4a, we consider $N = 5$,




(a) When $p_e^0 = 0.1$.

(b) When $p_e^0 = 0$.

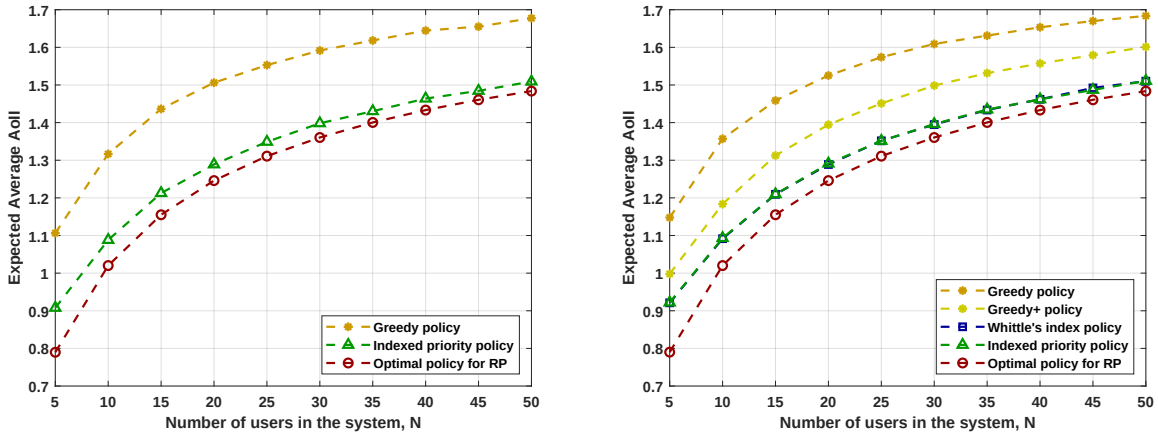
Figure 3.2: Performance when the source processes vary. We choose $p_i = 0.05 + \frac{0.4(i-1)}{N-1}$, $f_i(s) = s$, $\gamma_i = 0.6$, $p_{e,i}^0 = p_e^0$, and $p_{e,i}^1 = 0.1$ for $1 \leq i \leq N$.

$\gamma \in [0, 1]$, $p \in [0.05, 0.45]$, $p_e^{\hat{r}} \in [0, 0.45]$, and $f(s) = s^\tau$, where $\tau \in [0.5, 1.5]$. There are a total of 300 different choices and the results are sorted by the performance of RP-optimal policy in ascending order. Figure 3.4b adopts the same system settings except that we impose $p_{e,i}^0 = 0$ for $1 \leq i \leq N$ to ensure the feasibility of Whittle's index policy. Meanwhile, we ignore the Greedy policy since the Greedy+ policy achieves a better performance, as indicated by Figures 3.2b and 3.3b.

We can make the following observations from the figures.

- The Greedy+ policy yields a smaller expected average AoII than that achieved by the Greedy policy. Recall that we obtained the Greedy+ policy by applying the structural properties detailed in Corollary 3. Therefore, simple applications of the structural properties of the optimal policy can improve the performance of scheduling policies.

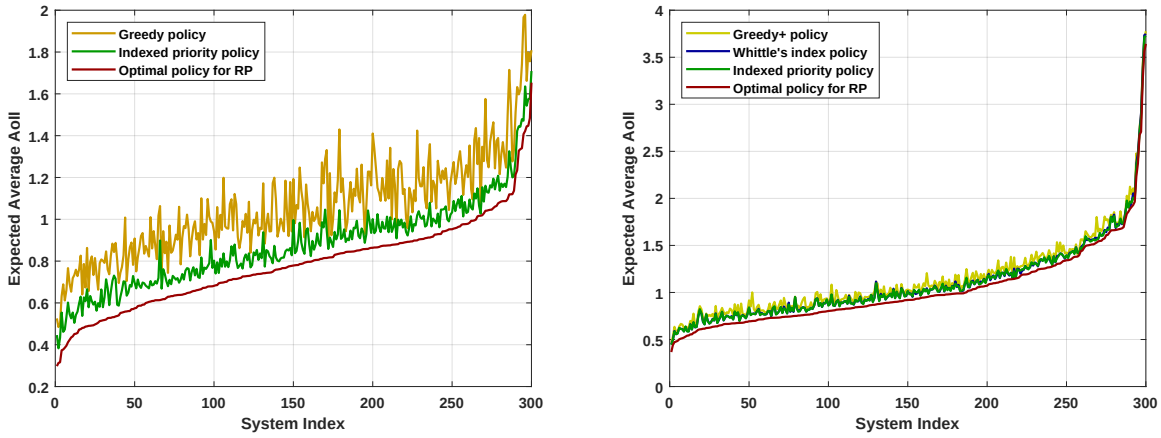
- The indexed priority policy has comparable performance to Whittle's index policy in all the system settings considered. The two policies have their own advantages. The indexed




(a) When $p_e^0 = 0.1$.

(b) When $p_e^0 = 0$.

Figure 3.3: Performance when the communication goals vary. We choose $f_i(s) = s^{0.5+\frac{i-1}{N-1}}$, $p_i = 0.3$, $\gamma_i = 0.6$, $p_{e,i}^0 = p_e^0$, and $p_{e,i}^1 = 0.1$ for $1 \leq i \leq N$.




(a) When $I = [0, 0.45]$.

(b) When $I = \{0\}$.

Figure 3.4: Performance in systems with random parameters when $N = 5$. The parameters for each user are chosen randomly within the following intervals: $\gamma \in [0, 1]$, $p \in [0.05, 0.45]$, $p_e^0 \in I$, $p_e^1 \in [0, 0.45]$, and $f(s) = s^\tau$ where $\tau \in [0.5, 1.5]$.

priority policy has a broader scope of application, while Whittle's index policy has a lower computational complexity.

- The performance of the indexed priority policy and Whittle's index policy is better than that of the Greedy/Greedy+ policies and is not far from the performance of the RP-optimal policy. Recall that the performance of the RP-optimal policy forms a universal lower bound

on the performance of all admissible policies for PP. Hence, we can conclude that both the indexed priority policy and Whittle's index policy achieve good performances.

## 3.8 Conclusions

In this chapter, we studied the problem of minimizing the Age of Incorrect Information in a slotted-time system where a base station needs to schedule $M$ users among $N$ available users. Meanwhile, the base station has access to imperfect CSI in each time slot. The problem is a RMAB problem which is SPACE-hard. However, by casting the problem into a MDP, we obtain the structural properties of the optimal policy. Then, we introduce a relaxed version of the original problem and investigate the decoupled model. Under a simple condition, we establish the indexability of the decoupled problem and obtain the expression of Whittle's index. On this basis, we developed Whittle's index policy. To get rid of the requirement for indexability, we developed the indexed priority policy based on the optimal policy for the RP. The characteristics of the RP are explored to make the calculation of its optimal policy more efficient. Finally, through numerical results, we show that simple applications of the structural properties can improve the performance of scheduling policies. Moreover, Whittle's index policy and the indexed priority policy achieve good and comparable performances.

# Chapter 4: Freshness against Generic Delay

## 4.1 Overview

With the rapid growth in the number of data exchanges and the unprecedented expansion of their application scenarios, the channel environment has become more complex, and the diversity of data has also increased. This leads to a large uncertainty in the data transmission time, which stimulates our interest in how to keep the information fresh in case of transmission delay. In this chapter, we consider the problem of minimizing AoII when the communication channel suffers a random delay. We provide a theoretical analysis of the problem, and the results apply to generic delay distributions. The system with a random delay communication channel has also been studied in the context of remote estimation and AoI [25, 73–75]. However, the problem considered in this chapter is very different, as AoII is a combination of age-based metric frameworks and error-based metric frameworks.

The main contributions of this chapter can be summarized as follows. 1) We investigate the AoII minimization problem in a system where the communication channel suffers a random delay and characterize the optimization problem using the Markov decision process. 2) We derive the analytical expression of the expected AoII achieved by the threshold policy, under which the transmitter initiates transmission only when the transmission is allowed, and AoII exceeds the threshold. 4) We prove the existence of the optimal policy and introduce a computable value

iteration algorithm to estimate the optimal policy. 5) We theoretically find the optimal policy using the policy improvement theorem.

This chapter summarizes the work in [76, 77], and the rest is organized as follows. We introduce the system model and the optimization problem in Section 4.2. Then, Section 4.3 characterizes the problem using the Markov decision process. In Section 4.4, we derive the analytical expression of the expected AoII achieved by the threshold policy. Then, we show the existence of the optimal policy, provide the value iteration algorithm to estimate it, and theoretically find the optimal policy using the policy improvement theorem in Section 4.5. Finally, Section 4.6 concludes the chapter with numerical results that highlight the performance of the optimal policy.

## 4.2 System Overview

### 4.2.1 System Model

We consider a slotted-time system in which a transmitter observes a dynamic source and needs to decide when to send status updates to a remote receiver so that the receiver can have a good knowledge of the current state of the dynamic source. The dynamic source is modeled by a two-state symmetric Markov chain with state transition probability $p$. The transmitter receives an update from the dynamic source at the beginning of each time slot. The update at time slot $k$ is denoted by $X_k$. The old update is discarded upon the arrival of a new one. Then, the transmitter decides whether to transmit the new update based on the current system status. When the channel is idle, the transmitter chooses between transmitting the new update and staying idle. When the channel is busy, the transmitter has no choice but to stay idle. The updates will be transmitted

over an error-free communication channel that suffers a random delay. In other words, the update will not be corrupted during the transmission, but each transmission will take a random amount of time $T \in \mathbb{N}^*$. We denote the probability mass function (PMF) by $p_t \triangleq Pr(T = t)$ and assume that $T$ is independent and identically distributed for each update. When a transmission finishes, the communication channel is immediately available for the subsequent transmission.

The receiver maintains an estimate of the current state of the dynamic source and modifies its estimate each time a new update is received. We denote by $\hat{X}_k$ the receiver's estimate at time slot $k$. According to [75], the best estimator when $p \leq \frac{1}{2}$ is the last received update. When $p > \frac{1}{2}$, the optimal estimator depends on the realization of transmission time. In this chapter, we only consider the case of $0 < p \leq \frac{1}{2}$. In this case, the receiver uses the last received update as the estimate. For the case of $p > \frac{1}{2}$, the results can be extended using the corresponding best estimator. The receiver uses $ACK/NACK$ packets to inform the transmitter of its reception of the new update. As is assumed in [36], the transmitter receives the $ACK/NACK$ packets reliably and instantaneously because the packets are generally very small compared to the size of the status updates. When $ACK$ is received, the transmitter knows that the receiver's estimate changes to the last sent update. When $NACK$ is received, the transmitter knows that the receiver's estimate does not change. In this way, the transmitter always knows the current estimate on the receiver side.

An illustration of the system model is shown in Figure 4.1. At the beginning of time slot $k$, the transmitter receives the update $X_k$ from the dynamic source. Then, the transmitter decides whether to transmit this update based on the system status. When the transmitter decides not to start transmission, it will stay idle. Otherwise, the transmitter will transmit the update through the communication channel, where the transmission of the update takes a random amount of time.

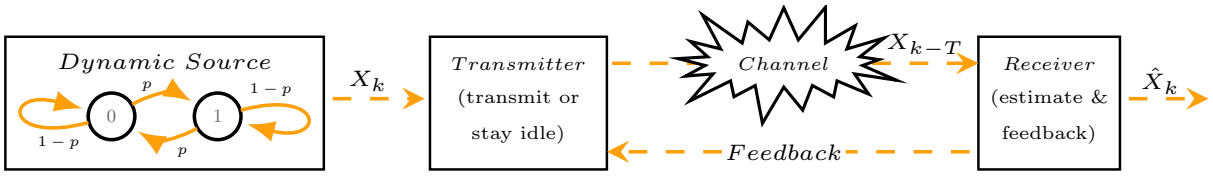



Figure 4.1: An illustration of the system model, where $X_k$ and $\hat{X}_k$ are the state of the dynamic source and the receiver's estimate at time slot $k$, respectively.

Thus, the update received by the receiver has a delay of several time slots (i.e., $X_{k-T}$). Then, the receiver will modify its estimation $\hat{X}_k$ based on the received update and send an $ACK$ packet to inform the transmitter of its reception of the update.

The system adopts AoII as the performance metric. Since the dynamic source has two states, we let $X_t \in \{0, 1\}$ and $\hat{X}_t \in \{0, 1\}$. Then, in this chapter, we choose $g(X_k, \hat{X}_k) = |X_k - \hat{X}_k|$ and $f(k) = k$. Hence, $F(k) = 1$ and $g(X_k, \hat{X}_k) \in \{0, 1\}$ as the dynamic source has two states. Then, AoII can be written as

$$\Delta_{AoII}(X_k, \hat{X}_k, k) = k - U_k \triangleq \Delta_k,$$

where $U_k$ is defined in Section 1.2.2. We can easily conclude from the simplified expression that, under the chosen penalty functions, AoII increases at the rate of 1 per time slot when the receiver's estimate is incorrect. Otherwise, AoII is 0. Next, we characterize the evolution of $\Delta_k$. To this end, we divide the evolution into the following cases.

- When $X_{k+1} = \hat{X}_{k+1}$, we have $U_{k+1} = k + 1$. Then, by definition, $\Delta_{k+1} = 0$.
- When $X_{k+1} \neq \hat{X}_{k+1}$, we have $U_{k+1} = U_k$. Then, by definition, $\Delta_{k+1} = k + 1 - U_k = \Delta_k + 1$.

Combining together, we have

$$\Delta_{k+1} = \mathbb{1}\{X_{k+1} \neq \hat{X}_{k+1}\}(\Delta_k + 1). \tag{4.1}$$

Now that the evolution of AoII has been clarified, we further discuss the system's evolution.

### 4.2.2 System Dynamic

In this subsection, we tackle the system dynamics, which will play a key role in later sections. We notice that the system's status at the beginning of time slot $k$ can be fully captured by the triplet $s_k \triangleq (\Delta_k, t_k, i_k)$ where $t_k \in \mathbb{N}^0$ indicates the time the current transmission has been in progress. We define $t_k = 0$ if there is no transmission in progress. $i_k \in \{-1, 0, 1\}$ indicates the state of the channel. We define $i_k = -1$ when the channel is idle, $i_k = 0$ if the channel is busy and the transmitting update is the same as the receiver's current estimate, and $i_k = 1$ when the transmitting update is different from the receiver's current estimate.

**Remark 10.** *According to the definitions of $t_k$ and $i_k$, $i_k = -1$ if and only if $t_k = 0$. In this case, the channel is idle.*

Then, characterizing the system dynamics is equivalent to characterizing the value of $s_{k+1}$ using $s_k$ and the transmitter's action. We use $a_k \in \{0, 1\}$ to denote the transmitter's decision, where $a_k = 0$ when the transmitter decides not to initiate a transmission and $a_k = 1$ otherwise. Hence, the system dynamics can be fully characterized by $P_{s_k, s_{k+1}}(a_k)$, which is the probability that action $a_k$ at $s_k$ leads to $s_{k+1}$. We will revisit $P_{s_k, s_{k+1}}(a_k)$ with an in-depth analysis later.

### 4.2.3 Problem Formulation

We define a policy $\phi$ as the one that specifies the transmitter's decision in each time slot. This chapter aims to find the policy that minimizes the expected AoII of the system. Mathematically, the problem can be formulated as the following optimization problem.

$$\underset{\phi \in \Phi}{\arg\min} \quad \lim_{K\to\infty} \frac{1}{K}\mathbb{E}_{\phi}\left(\sum_{k=0}^{K-1} \Delta_k\right), \tag{4.2}$$

where $\mathbb{E}_{\phi}$ is the conditional expectation, given that policy $\phi$ is adopted, and $\Phi$ is the set of all admissible policies.

**Definition 6** (Optimal policy)**.** *A policy is said to be optimal if it yields the minimal expected AoII.*

In the next section, we characterize the problem reported in (4.2) using a Markov decision process (MDP).

## 4.3 MDP Characterization

The minimization problem reported in (4.2) can be characterized by an infinite horizon with average cost MDP $\mathcal{M}$, which consists of the following components.

- The state space $\mathcal{S}$. The state $s = (\Delta, t, i)$ is the triplet defined in Section 4.2.2 without the time stamp. For the remainder of this chapter, we will use $s$ and $(\Delta, t, i)$ to represent the state interchangeably. They will synchronize any superscript or subscript.

- The action space $\mathcal{A}$. When $i = -1$, the feasible action is $a \in \{0, 1\}$ where $a = 0$ if the

transmitter decides not to initiate a new transmission and $a = 1$ otherwise. When $i \neq -1$, the feasible action is $a = 0$.

- The state transition probability $\mathcal{P}$. The probability that the operation of action $a$ at state $s$ leads to state $s'$ is denoted by $P_{s,s'}(a)$, whose value will be discussed in the next subsection.

- The immediate cost $\mathcal{C}$. The immediate cost for being at state $s$ is $C(s) = \Delta$.

Let $V(s)$ be the value function of state $s \in \mathcal{S}$. It is well known that the value function satisfies the Bellman equation [78].

$$V(s) + \theta = \min_{a \in \mathcal{A}} \left\{ C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V(s') \right\} \quad s \in \mathcal{S}, \tag{4.3}$$

where $\theta$ is the expected AoII achieved by the optimal policy. We will write $V(s)$ as $V(\Delta, t, i)$ in some parts of this chapter to better distinguish between states. We notice that the state transition probability is essential for solving the Bellman equation. Hence, we delve into $P_{s,s'}(a)$ in the following subsection.

### 4.3.1 State Transition Probability

We recall that $P_{s,s'}(a)$ is the probability that action $a$ at state $s$ will lead to state $s'$. Then, we define $Pr(T > k + 1 \mid t)$ as the probability that the current transmission will take more than $t + 1$ time slots, given that the current transmission has been in progress for $t$ time slots. Hence,

$$Pr(T > t + 1 \mid t) = \frac{1 - Pr(T \leq t + 1)}{Pr(T > t)} = \frac{1 - P_{t+1}}{1 - P_t},$$

where $P_t \triangleq \sum_{k=1}^{t} p_k$. Leveraging this, $P_{s,s'}(a)$ can be obtained easily. For the sake of space, the complete state transition probabilities are detailed in Appendix C.1.

We notice that we do not impose any restrictions on the update transmission time, which would make the theoretical analysis very difficult and lead to long channel occupancy by a single update. Therefore, to ease the theoretical analysis and be closer to the practice, we consider the following two independent assumptions. [1]

- **Assumption 1**: We assume that the update will always be delivered and the transmission lasts at most $t_{max}$ time slots. More precisely, we assume $1 \leq T \leq t_{max}$ and

$$\sum_{t=1}^{t_{max}} p_t = 1, \quad p_t \geq 0,\ 1 \leq t \leq t_{max}.$$

  In practice, we can make the probability of the transmission time exceeding $t_{max}$ negligible by choosing a sufficiently large $t_{max}$.

- **Assumption 2**: We assume the transmission can last for a maximum of $t_{max}$ time slots. At the end of the $t_{max}$th time slot, the update will be discarded if not delivered, and the channel will be available for a new transmission immediately. We define $p_{t^+} \triangleq \sum_{t=t_{max}+1}^{\infty} p_t$ as the probability that the update will be discarded. In practice, similar techniques, such as time-to-live (TTL) [79], are used to prevent an update from occupying the channel for too long.

**Remark 11.** *$t_{max}$ is a predetermined system parameter and is not a parameter to be optimized. When $t_{max} = 1$, the system reduces to the one considered in [36], according to which the optimal*

[1] The results presented in this chapter apply to both assumptions unless stated otherwise.

*policy is to transmit a new update whenever possible. Therefore, in the rest of this chapter, we focus on the case of* $t_{max} > 1$.

Under both assumptions, the transmission will last at most $t_{max}$ time slots, and the channel will be immediately available for a new transmission when the current transmission finishes. Hence, the state space $\mathcal{S}$ is reduced as $t$ is now bounded by $0 \leq t \leq t_{max} - 1$. Moreover, the state transition probabilities in Appendix C.1 will be adjusted as follows.

- Under **Assumption 1**, updates are bound to be delivered after $t_{max}$ time slots. Hence, $Pr(T > t + 1 \mid t) = 0$ for $t \geq t_{max} - 1$.
- Under **Assumption 2**, updates will be discarded at the end of the $t_{max}$th time slot if not delivered. Hence, $s' = (\Delta', t_{max}, i')$ will be replaced by $s' = (\Delta', 0, -1)$.

Having clarified the state transition probabilities, we evaluate a canonical policy in terms of the achieved expected AoII in the next section.

## 4.4 Policy Performance Analysis

As is proved in [36, 37, 48], the AoII-optimal policy often has a threshold structure. Hence, we consider the threshold policy.

**Definition 7** (Threshold policy)**.** *Under threshold policy* $\tau$*, the transmitter initiates a transmission only when the current AoII is no less than threshold* $\tau \in \mathbb{N}^0$ *and the channel is idle.*

**Remark 12.** *We define* $\tau \triangleq \infty$ *as the policy under which the transmitter never initiates any transmissions.*

We notice that the system dynamics under threshold policy can be characterized by a discrete-time Markov chain (DTMC). Without loss of generality, we assume the DTMC starts at state $(0, 0, -1)$. Then, the state space of the Markov chain $\mathcal{S}^{MC}$ consists of all the states accessible from state $(0, 0, -1)$. Since state $(0, 0, -1)$ is positive recurrent and communicates with each state $s \in \mathcal{S}^{MC}$, the stationary distribution exists. Let $\pi_s$ be the steady-state probability of state $s$. Then, $\pi_s$ satisfies the following balance equation.

$$\pi_s = \sum_{s' \in \mathcal{S}^{MC}} P_{s',s}(a)\pi_{s'} \quad s \in \mathcal{S}^{MC},$$

where $P_{s',s}(a)$ is the single-step state transition probability as define in Section 4.3, and the action $a$ depends on the threshold policy. Then, the first step in calculating the expected AoII achieved by the threshold policy is to calculate the stationary distribution of the induced DTMC. However, the problem arises as the state space $\mathcal{S}^{MC}$ is infinite and intertwined. To simplify the state transitions, we recall that the transmitter can only stay idle (i.e., $a = 0$) when the channel is busy. Let $\mathcal{S}^{MC}_{-1} = \{s = (\Delta, t, i) : i \neq -1\}$ be the set of the state where the channel is busy. Then, for $s' \in \mathcal{S}^{MC}_{-1}$, $P_{s',s}(a) = P_{s',s}(0)$ and is independent of the threshold policy. Hence, for any threshold policy and each $s \in \mathcal{S} \setminus \mathcal{S}^{MC}_{-1}$, we can repeatedly replace $\pi_{s'}$, where $s' \in \mathcal{S}^{MC}_{-1}$, with the corresponding balance equation until we get the following equation.

$$\pi_s = \sum_{s' \in \mathcal{S} \setminus \mathcal{S}^{MC}_{-1}} P_{\Delta',\Delta}(a)\pi_{s'} \quad s \in \mathcal{S} \setminus \mathcal{S}^{MC}_{-1}, \tag{4.4}$$

where $P_{\Delta',\Delta}(a)$ is the multi-step state transition probability from state $s' = (\Delta', 0, -1)$ to state

$s = (\Delta, 0, -1)$ under action $a$. For simplicity, we write (4.4) as

$$\pi_\Delta = \sum_{\Delta' \geq 0} P_{\Delta',\Delta}(a)\pi_{\Delta'} \quad \Delta \geq 0. \tag{4.5}$$

As we will see in the following subsections, $\pi_\Delta$ is sufficient to calculate the expected AoII obtained by any threshold policy.

**Remark 13.** *The intuition behind the simplification of the balance equations is as follows. We recall that the system dynamics when the channel is busy are independent of the adopted policy. Hence, we can calculate these dynamics in advance so that the balance equations contain only the states in which the transmitter needs to make decisions.*

In the next subsection, we derive the expression of $P_{\Delta,\Delta'}(a)$.

### 4.4.1 Multi-step State Transition Probability

We start with the case of $a = 0$. In this case, no update will be transmitted, and $P_{\Delta,\Delta'}(0)$ is independent of the transmission delay. Then, according to Appendix C.1,

$$P_{0,\Delta'}(0) = \begin{cases} 1-p & \Delta' = 0, \\ p & \Delta' = 1, \end{cases}$$

and for $\Delta > 0$,

$$P_{\Delta,\Delta'}(0) = \begin{cases} p & \Delta' = 0, \\ 1-p & \Delta' = \Delta + 1. \end{cases}$$

In the sequel, we focus on the case of $a = 1$. We define $P^t_{\Delta,\Delta'}(a)$ as the probability that action $a$ at state $s = (\Delta, 0, -1)$ will lead to state $s' = (\Delta', 0, -1)$, given that the transmission takes $t$ time slots. Then, under **Assumption 1**,

$$P_{\Delta,\Delta'}(1) = \sum_{t=1}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1).$$

Hence, it is sufficient to obtain the expressions of $P^t_{\Delta,\Delta'}(1)$. To this end, we define $p^{(t)}$ as the probability that the dynamic source will remain in the same state after $t$ time slots. Since the Markov chain is symmetric, $p^{(t)}$ is independent of the state and can be calculated by

$$p^{(t)} = \left( \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}^t \right)_{11},$$

where the subscript indicates the row number and the column number of the target probability. For the consistency of notation, we define $p^{(0)} \triangleq 1$. Then, we have the following lemma.

**Lemma 4.** *Under **Assumption 1**,*

$$P_{\Delta,\Delta'}(1) = \sum_{t=1}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1), \tag{4.6}$$

*where*

$$P^t_{0,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ p^{(t-k)} p (1-p)^{k-1} & 1 \le \Delta' = k \le t, \\ 0 & otherwise, \end{cases}$$

*and for* $\Delta > 0$*,*

$$P^t_{\Delta,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ (1-p^{(t-1)})(1-p) & \Delta' = 1, \\ (1-p^{(t-k)})p^2(1-p)^{k-2} & 2 \le \Delta' = k \le t-1, \\ p(1-p)^{t-1} & \Delta' = \Delta + t, \\ 0 & otherwise. \end{cases}$$

*Under* ***Assumption 1****, equation* (4.6) *can be written equivalently as*

$$P_{\Delta,\Delta'}(1) = \begin{cases} \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) & 0 \le \Delta' \le t_{max}-1, \Delta \ge \Delta', \\ \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) + p_{t'} P^{t'}_{\Delta,\Delta'}(1) & 0 \le \Delta' \le t_{max}-1, \Delta < \Delta', \\ p_{t'} P^{t'}_{\Delta,\Delta'}(1) & \Delta' \ge t_{max}, \end{cases}$$

*where* $t' \triangleq \Delta' - \Delta$ *and* $P^{t'}_{\Delta,\Delta'}(1) \triangleq 0$ *when* $t' \le 0$ *or when* $t' > t_{max}$*. Meanwhile,* $P_{\Delta,\Delta'}(1)$ *possesses the following properties.*

1. $P_{\Delta,\Delta'}(1)$ *is independent of* $\Delta$ *when* $0 \le \Delta' \le t_{max}-1$ *and* $\Delta \ge \Delta'$*.*
2. $P_{\Delta,\Delta'}(1) = P_{\Delta+\delta,\Delta'+\delta}(1)$ *when* $\Delta' \ge t_{max}$ *and* $\Delta \ge 0$ *for any* $\delta \ge 1$*.*
3. $P_{\Delta,\Delta'}(1) = 0$ *when* $\Delta' > \Delta + t_{max}$ *or when* $t_{max} - 1 < \Delta' < \Delta + 1$*.*

*Proof.* The expression of $P^t_{\Delta,\Delta'}(1)$ is obtained by analyzing the system dynamics. The complete proof can be found in Appendix C.2. □

The state transition probabilities under **Assumption 2** can be obtained similarly. To this

end, we define $P^{t^+}_{\Delta,\Delta'}(a)$ as the probability that action $a$ at state $s = (\Delta, 0, -1)$ will result in state $s' = (\Delta', 0, -1)$, given that the transmission is terminated. Then, we have the following lemma.

**Lemma 5.** *Under **Assumption 2**,*

$$P_{\Delta,\Delta'}(1) = \sum_{t=1}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1), \tag{4.7}$$

*where*

$$P^t_{0,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ p^{(t-k)} p (1-p)^{k-1} & 1 \le \Delta' = k \le t, \\ 0 & otherwise, \end{cases}$$

$$P^{t^+}_{0,\Delta'}(1) = P^{t_{max}}_{0,\Delta'}(1),$$

*and for $\Delta > 0$,*

$$P^t_{\Delta,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ (1-p^{(t-1)})(1-p) & \Delta' = 1, \\ (1-p^{(t-k)}) p^2 (1-p)^{k-2} & 2 \le \Delta' = k \le t-1, \\ p(1-p)^{t-1} & \Delta' = \Delta + t, \\ 0 & otherwise, \end{cases}$$

$$P^{t^+}_{\Delta,\Delta'}(1) = \begin{cases} 1 - p^{(t_{max})} & \Delta' = 0, \\ (1 - p^{(t_{max}-k)})p(1-p)^{k-1} & 1 \leq \Delta' = k \leq t_{max} - 1, \\ (1-p)^{t_{max}} & \Delta' = \Delta + t_{max}, \\ 0 & otherwise. \end{cases}$$

*Under **Assumption 2**, equation (4.7) can be written equivalently as*

$$P_{\Delta,\Delta'}(1) = \begin{cases} \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1) & 0 \leq \Delta' \leq t_{max} - 1, \Delta \geq \Delta', \\ \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) + p_{t'} P^{t'}_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1) & 0 \leq \Delta' \leq t_{max} - 1, \Delta < \Delta', \\ p_{t'} P^{t'}_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1) & \Delta' \geq t_{max}. \end{cases}$$

*Meanwhile, $P_{\Delta,\Delta'}(1)$ possesses the following properties.*

1. *$P_{\Delta,\Delta'}(1)$ is independent of $\Delta$ when $0 \leq \Delta' \leq t_{max} - 1$ and $\Delta \geq \max\{1, \Delta'\}$.*

2. *$P_{\Delta,\Delta'}(1) = P_{\Delta+\delta,\Delta'+\delta}(1)$ when $\Delta' \geq t_{max}$ and $\Delta > 0$ for any $\delta \geq 1$.*

3. *$P_{\Delta,\Delta'}(1) = 0$ when $\Delta' > \Delta + t_{max}$ or when $t_{max} - 1 < \Delta' < \Delta + 1$.*

*Proof.* The proof follows similar steps as presented in the proof of Lemma 4. The complete proof can be found in Appendix C.3. □

As the expressions and properties of $P_{\Delta,\Delta'}(a)$ under both assumptions are clarified, we solve for $\pi_\Delta$ in the next subsection.

### 4.4.2 Stationary Distribution

Let $ET$ be the expected transmission time of an update. Since the channel remains idle if no transmission is initiated and the expected transmission time of an update is $ET$, $\pi_\Delta$ satisfies the following equation.

$$\sum_{\Delta=0}^{\tau-1} \pi_\Delta + ET \sum_{\Delta=\tau}^{\infty} \pi_\Delta = 1, \tag{4.8}$$

where $ET = \sum_{t=1}^{t_{max}} t p_t$ under **Assumption 1** and $ET = \sum_{t=1}^{t_{max}} t p_t + t_{max} p_{t^+}$ under **Assumption 2**. We notice that there is still infinitely many $\pi_\Delta$ to calculate. To overcome the infinity, we recall that, under threshold policy, the suggested action is $a = 1$ for all the state $(\Delta, 0, -1)$ with $\Delta \geq \tau$. Hence, we define $\Pi \triangleq \sum_{\Delta=\omega}^{\infty} \pi_\Delta$ where $\omega \triangleq t_{max} + \tau + 1$. As we will see in the following subsections, $\Pi$ and $\pi_\Delta$ for $0 \leq \Delta < \omega - 1$ are sufficient for calculating the expected AoII achieved by the threshold policy. With $\Pi$ in mind, we have the following theorem.

**Theorem 6.** *For $0 < \tau < \infty$, $\Pi$ and $\pi_\Delta$ for $0 \leq \Delta < \omega - 1$ are the solution to the following system of linear equations.*

$$\pi_0 = (1-p)\pi_0 + p \sum_{i=1}^{\tau-1} \pi_i + P_{1,0}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right).$$

$$\pi_1 = p\pi_0 + P_{1,1}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right).$$

$$\Pi = \sum_{i=\tau+1}^{\omega-1} \left( \sum_{k=\tau+1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{max}} \left( P_{\omega,\omega+i}(1) \right) \Pi.$$

$$\sum_{i=0}^{\tau-1} \pi_i + ET \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right) = 1.$$

*For each* $2 \leq \Delta \leq t_{max} - 1$,

$$\pi_\Delta = \begin{cases} (1-p)\pi_{\Delta-1} + P_{\tau,\Delta}(1)\left(\sum_{i=\tau}^{\omega-1}\pi_i + \Pi\right) & \Delta - 1 < \tau, \\ \sum_{i=\tau}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\left(\sum_{i=\Delta}^{\omega-1}\pi_i + \Pi\right) & \Delta - 1 \geq \tau. \end{cases}$$

*For each* $t_{max} \leq \Delta \leq \omega - 1$,

$$\pi_\Delta = \begin{cases} (1-p)\pi_{\Delta-1} & \Delta - 1 < \tau, \\ \sum_{i=\tau}^{\Delta-1} P_{i,\Delta}(1)\pi_i & \Delta - 1 \geq \tau. \end{cases}$$

*Proof.* We delve into the definition of $\Pi$. By leveraging the structural property of the threshold policy and the properties of $P_{\Delta,\Delta'}(a)$, we obtain the above system of linear equations. The complete proof can be found in Appendix C.4. □

**Remark 14.** *The size of the system of linear equations detailed in Theorem 6 is* $\omega + 1$.

**Corollary 6.** *When* $\tau = 0$,

$$\pi_0 = \frac{P_{1,0}(1)}{ET[1 - P_{0,0}(1) + P_{1,0}(1)]}.$$

*For each* $1 \leq \Delta \leq t_{max}$,

$$\pi_\Delta = \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\left(\frac{1}{ET} - \sum_{i=0}^{\Delta-1}\pi_i\right).$$

$$\Pi = \frac{\sum_{i=1}^{t_{max}} \left( \sum_{k=1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i}{1 - \sum_{i=1}^{t_{max}} P_{t_{max}+1,t_{max}+1+i}(1)}.$$

*When* $\tau = 1$,

$$\pi_0 = \frac{P_{1,0}(1)}{pET + P_{1,0}(1)}, \quad \pi_1 = \frac{pP_{1,0}(1) + pP_{1,1}(1)}{pET + P_{1,0}(1)}.$$

*For each* $2 \leq \Delta \leq t_{max} + 1$,

$$\pi_\Delta = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \left( \frac{1 - \pi_0}{ET} - \sum_{i=1}^{\Delta-1} \pi_i \right).$$

$$\Pi = \frac{\sum_{i=2}^{t_{max}+1} \left( \sum_{k=2}^{i} P_{i,t_{max}+k}(1) \right) \pi_i}{1 - \sum_{i=1}^{t_{max}} P_{t_{max}+2,t_{max}+2+i}(1)}.$$

*Proof.* The calculations follow similar steps as detailed in the proof of Theorem 6. The complete proof can be found in Appendix C.5. □

We will calculate the expected AoII in the next subsection based on the above results.

### 4.4.3 Expected AoII

Let $\bar{\Delta}_\tau$ be the expected AoII achieved by threshold policy $\tau$. Then,

$$\bar{\Delta}_\tau = \sum_{\Delta=0}^{\tau-1} C(\Delta, 0)\pi_\Delta + \sum_{\Delta=\tau}^{\infty} C(\Delta, 1)\pi_\Delta, \tag{4.9}$$

where $C(\Delta, a)$ is the expected sum of AoII during the transmission of the update caused by the operation of $a$ at state $(\Delta, 0, -1)$. Note that $C(\Delta, a)$ includes the AoII for being at state $(\Delta, 0, -1)$.

**Remark 15.** *In order to have a more intuitive understanding of the definition of $C(\Delta, a)$, we use $\eta$ to denote a possible path of the state during the transmission of the update and let $H$ be the set of all possible paths. Moreover, we denote by $C_\eta$ and $P_\eta$ the sum of AoII and the probability associated with path $\eta$, respectively. Then,*

$$C(\Delta, a) = \sum_{\eta \in H} P_\eta C_\eta.$$

*For example, we consider the case of $p_2 = 1$, where the transmission takes 2 time slots to be delivered. Also, action $a = 1$ is taken at state $(2, 0, -1)$. Then, a sample path $\eta$ of the state during the transmission can be the following.*

$$(2, 0, -1) \to (3, 1, 1) \to (4, 0, -1).$$

*By our definition, $C_\eta = 2 + 3 = 5$ and $P_\eta = Pr[(3, 1, 1) \mid (2, 0, -1), a = 1] \cdot Pr[(4, 0, -1) \mid (3, 1, 1), a = 1]$ for the above sample path.*

In the following, we calculate $C(\Delta, a)$. Similar to Section 4.4.1, we define $C^t(\Delta, a)$ as the expected sum of AoII during the transmission of the update caused by action $a$ at state $(\Delta, 0, -1)$,

given that the transmission takes $t$ time slots. Then, under **Assumption 1**,

$$C(\Delta, a) = \begin{cases} \Delta & a = 0, \\ \sum_{t=1}^{t_{max}} p_t C^t(\Delta, 1) & a = 1, \end{cases} \tag{4.10}$$

and, under **Assumption 2**,

$$C(\Delta, a) = \begin{cases} \Delta & a = 0, \\ \sum_{t=1}^{t_{max}} p_t C^t(\Delta, 1) + p_{t^+} C^{t_{max}}(\Delta, 1) & a = 1. \end{cases} \tag{4.11}$$

Hence, obtaining the expressions of $C^t(\Delta, 1)$ is sufficient. To this end, we define $C^k(\Delta)$ as the expected AoII $k$ time slots after the transmission starts at state $(\Delta, 0, -1)$, given that the transmission is still in progress. Then, we have the following lemma.

**Lemma 6.** *$C^t(\Delta, 1)$ is given by*

$$C^t(\Delta, 1) = \sum_{k=0}^{t-1} C^k(\Delta),$$

*where $C^k(\Delta)$ is given by*

$$C^k(\Delta) = \begin{cases} \sum_{h=1}^{k} h p^{(k-h)} p (1-p)^{h-1} & \Delta = 0, \\ \sum_{h=1}^{k-1} h (1 - p^{(k-h)}) p (1-p)^{h-1} + (\Delta + k)(1-p)^k & \Delta > 0. \end{cases}$$

*Proof.* The expression of $C^k(\Delta)$ is obtained by analyzing the system dynamics. The complete proof can be found in Appendix C.6. □

Next, we calculate the expected AoII achieved by the threshold policy. We start with the case of $\tau = \infty$.

**Theorem 7.** *The expected AoII achieved by the threshold policy with $\tau = \infty$ is*

$$\bar{\Delta}_{\infty} = \frac{1}{2p}.$$

*Proof.* In this case, the transmitter never initiates any transmissions. Hence, the state transitions are straightforward. The complete proof can be found in Appendix C.7. □

In the following, we focus on the case where $\tau$ is finite. We recall that the expected AoII is given by (4.9). The problem arises because of the infinite sum. To overcome this, we adopt a similar approach as proposed in Section 4.4.2. More precisely, we leverage the structural property of the threshold policy and define $\Sigma \triangleq \sum_{\Delta=\omega}^{\infty} C(\Delta, 1)\pi_{\Delta}$. Then, equation (4.9) can be written as

$$\bar{\Delta}_{\tau} = \sum_{i=0}^{\tau-1} C(i,0)\pi_i + \sum_{i=\tau}^{\omega-1} C(i,1)\pi_i + \Sigma.$$

As we have obtained the expressions of $\pi_{\Delta}$ and $C(\Delta, a)$ in previous subsections, it is sufficient to obtain the expression of $\Sigma$.

**Theorem 8.** *Under **Assumption 1** and for $0 \leq \tau < \infty$,*

$$\Sigma = \frac{\sum_{t=1}^{t_{max}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C(i,1)\pi_i \right) + \Delta_t' \Pi_t \right]}{1 - \sum_{t=1}^{t_{max}} \left( p_t P^t_{1,1+t}(1) \right)},$$

*where*

$$\Pi_t = p_t P^t_{1,1+t}(1)\left(\sum_{i=\omega-t}^{\omega-1}\pi_i + \Pi\right),$$

$$\Delta'_t = \sum_{i=1}^{t_{max}} p_i \left(\frac{t - t(1-p)^i}{p}\right).$$

*Under* ***Assumption 2*** *and for* $0 \leq \tau < \infty$*,*

$$\Sigma = \frac{\sum_{t=1}^{t_{max}}\left[\left(\sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)C(i,1)\pi_i\right) + \Delta'_t\Pi_t\right]}{1 - \sum_{t=1}^{t_{max}}\Upsilon(\omega+t,t)},$$

*where*

$$\Upsilon(\Delta,t) = p_t P^t_{\Delta-t,\Delta}(1) + p_{t^+} P^{t^+}_{\Delta-t,\Delta}(1),$$

$$\Pi_t = \sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)\pi_i + \Upsilon(\omega+t,t)\Pi,$$

$$\Delta'_t = \sum_{i=1}^{t_{max}} p_i\left(\frac{t-t(1-p)^i}{p}\right) + p_{t^+}\left(\frac{t-t(1-p)^{t_{max}}}{p}\right).$$

*Proof.* We delve into the definition of $\Sigma$ and repeatedly use the properties of $C(\Delta,a)$ and $P_{\Delta,\Delta'}(a)$. The complete proof can be found in Appendix C.8. □

## 4.5 Optimal Policy

In this section, we find the optimal policy for $\mathcal{M}$ theoretically. First of all, we prove that the optimal policy exists.

### 4.5.1 Existence of the Optimal Policy

We introduce the infinite horizon $\gamma$-discounted cost of $\mathcal{M}$, where $0 < \gamma < 1$ is a discount factor. The expected $\gamma$-discounted cost under policy $\phi$ is

$$V_{\phi,\gamma}(s) = \mathbb{E}_{\phi}\left[\sum_{t=0}^{\infty} \gamma^t C(s_t) \mid s_0\right], \tag{4.12}$$

where $s_t$ is the state of $\mathcal{M}$ at time slot $t$. We define $V_{\gamma}(s) \triangleq \inf_{\phi} V_{\phi,\gamma}(s)$ as the best that can be achieved. Equivalently, $V_{\gamma}(s)$ is the value function associated with the $\gamma$-discounted version of $\mathcal{M}$. Hence, $V_{\gamma}(s)$ satisfies the corresponding Bellman equation.

$$V_{\gamma}(s) = \min_{a \in \mathcal{A}} \left\{ C(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V_{\gamma}(s') \right\}.$$

Value iteration algorithm is a canonical algorithm to calculate $V_{\gamma}(s)$. Let $V_{\gamma,\nu}(s)$ be the estimated value function at iteration $\nu$. Then, the estimated value function is updated in the following way.

$$V_{\gamma,\nu+1}(s) = \min_{a \in \mathcal{A}} \left\{ C(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V_{\gamma,\nu}(s') \right\}. \tag{4.13}$$

**Lemma 7.** *When updated following* (4.13)*,* $\lim_{\nu \to \infty} V_{\gamma,\nu}(s) = V_{\gamma}(s)$.

*Proof.* According to [80, Propositions 1 and 3], it is sufficient to show that $V_{\gamma}(s)$ is finite. To this end, we consider the policy $\phi$ being the threshold policy with $\tau = \infty$. According to (4.12), we

have

$$V_{\phi,\gamma}(s) = \mathbb{E}_{\phi}\left[\sum_{t=0}^{\infty}\gamma^t C(s_t) \mid s_0\right] \leq \sum_{t=0}^{\infty}\gamma^t(\Delta_0 + t) = \frac{\Delta_0}{1-\gamma} + \frac{\gamma}{(1-\gamma)^2} < \infty.$$

Then, by definition, we have $V_\gamma(s) \leq V_{\gamma,\phi}(s) < \infty$. Hence, the value iteration reported in (4.13) will converge to the value function. □

Leveraging the convergence of the value iteration algorithm, we can prove the following structural property of $V_\gamma(s)$.

**Lemma 8.** *$V_\gamma(s)$ is non-decreasing in $\Delta$ when $\Delta > 0$.*

*Proof.* We recall that $V_\gamma(s)$ can be calculated using the value iteration algorithm. Hence, the monotonicity of $V_\gamma(s)$ can be proved via mathematical induction. The complete proof can be found in Appendix C.9. □

Now, we proceed with showing the existence of the optimal policy. To this end, we first define the stationary policy.

**Definition 8** (Stationary policy)**.** *A stationary policy specifies a single action in each time slot.*

**Theorem 9.** *There exists a stationary policy that is optimal for $\mathcal{M}$. Moreover, the minimum expected AoII is independent of the initial state.*

*Proof.* We show that $\mathcal{M}$ verifies the two conditions given in [80]. Then, the results in the theorem is guaranteed by [80, Theorem]. The complete proof can be found in Appendix C.10. □

We denote by $\phi^*$ the optimal policy for $\mathcal{M}$. Then, the next problem is how to find $\phi^*$. To solve MDP, the value iteration algorithm and the policy iteration algorithm are two of the

most popular. In the value iteration algorithm, the value function $V(s)$ is computed iteratively until convergence. However, since the state space $\mathcal{S}$ is infinite, it is not feasible to compute the value function for all states. To make the calculation feasible, in Section 4.5.2, an approximation method is used to obtain an approximated optimal policy $\hat{\phi}^*$, and we rigorously prove that $\hat{\phi}^*$ converges to $\phi^*$. However, the choice of the approximation parameters can significantly affect the complexity of the algorithm and may even lead to a non-optimal policy. To avoid this problem, in Section 4.5.3, we introduce the policy iteration algorithm and find $\phi^*$ theoretically using the policy improvement theorem. We start with the value iteration algorithm in the following subsection.

### 4.5.2 Value Iteration Algorithm

In this subsection, we present the relative value iteration (RVI) algorithm that approximates $\phi^*$. Direct application of RVI becomes impractical as the state space $\mathcal{S}$ is infinite. Hence, we use approximating sequence method (ASM) [52]. To this end, we construct another MDP $\mathcal{M}^{(m)} = (\mathcal{S}^{(m)}, \mathcal{A}, \mathcal{P}^{(m)}, \mathcal{C})$ by truncating the value of $\Delta$. More precisely, we impose

$$\mathcal{S}^{(m)} : \begin{cases} \Delta \in \{0, 1, ..., m\}, \\ i \in \{-1, 0, 1\}, \\ t \in \{0, 1, ..., t_{max} - 1\}, \end{cases}$$

where $m$ is the predetermined maximal value of $\Delta$. The transition probabilities from $s \in \mathcal{S}^{(m)}$ to $z \in \mathcal{S} \setminus \mathcal{S}^{(m)}$ are redistributed to the states $s' \in \mathcal{S}^{(m)}$ in the following way.

$$P_{s,s'}^{(m)}(a) = \begin{cases} P_{s,s'}(a) & \Delta' < m, \\ P_{s,s'}(a) + \sum_{G(z,s')} P_{s,z}(a) & \Delta' = m, \end{cases}$$

where $G(z, s') = \{z = (\Delta, t, i) : \Delta > m, t = t', i = i'\}$. The action space $\mathcal{A}$ and the immediate cost $\mathcal{C}$ are the same as defined in $\mathcal{M}$.

**Theorem 10.** *The sequence of optimal policies for $\mathcal{M}^{(m)}$ will converge to the optimal policy for $\mathcal{M}$ as $m \to \infty$.*

*Proof.* The proof follows the same steps as those in the proof of [48, Theorem 1]. The complete proof can be found in Appendix C.11. □

Then, we can apply RVI to $\mathcal{M}^{(m)}$ and treat the resulting policy as an approximation of $\phi^*$. The pseudocode of RVI is given in Algorithm 5. However, the choice of the approximation parameter $m$ is crucial. A large $m$ can add unnecessary computational complexity, while a small $m$ can lead to a non-optimal policy. Therefore, in the following subsections, we use the policy iteration algorithm and the policy improvement theorem to find $\phi^*$ theoretically. We start with introducing the policy iteration algorithm.

### 4.5.3 Policy Iteration Algorithm

The policy iteration algorithm is an iterative algorithm that iterates between the following two steps until convergence. [2]

[2] The convergence happens when two consecutive iterations produce equivalent policies.

**Algorithm 5** Relative Value Iteration

**procedure** RVI($\mathcal{M}^{(m)}$,$\epsilon$)
  $V_0(s) \leftarrow 0$ for $s \in \mathcal{S}^{(m)}$; $\nu \leftarrow 0$
  Choose $s^{ref} \in \mathcal{S}^{(m)}$ arbitrarily
  **repeat**
    **for** $s \in \mathcal{S}^{(m)}$ **do**
      **for** $a \in \mathcal{A}$ **do**
        $H_{s,a} \leftarrow C(s) + \sum_{s'} P^{(m)}_{s,s'}(a) V_\nu(s')$
      $Q_{\nu+1}(s) \leftarrow \min_a\{H_{s,a}\}$
    **for** $s \in \mathcal{S}^{(m)}$ **do**
      $V_{\nu+1}(s) \leftarrow Q_{\nu+1}(s) - Q_{\nu+1}(s^{ref})$
    $\nu \leftarrow \nu + 1$
  **until** $\max_s\{|V_\nu(s) - V_{\nu-1}(s)|\} \leq \epsilon$
  **return** $\hat{\phi}^* \leftarrow \text{argmin}_a\{H_{s,a}\}$

1. The first step is policy evaluation. In this step, we calculate the value function $V^\phi(s)$ and the expected AoII $\theta^\phi$ resulting from the adoption of some policy $\phi$. More precisely, the value function and the expected AoII are obtained by solving the following system of linear equations.

$$V^\phi(s) + \theta^\phi = C(s) + \sum_{s' \in \mathcal{S}} P^\phi_{s,s'} V^\phi(s') \quad s \in \mathcal{S}, \tag{4.14}$$

   where $P^\phi_{s,s'}$ is the state transition probability from $s$ to $s'$ when policy $\phi$ is adopted. Note that (4.14) forms an underdetermined system. Hence, we can select a reference state $s$ arbitrarily and set the corresponding value function to $0$. In this way, we can obtain a unique solution.

2. The second step is policy improvement. In this step, we obtain a new policy $\phi'$ using the $V^\phi(s)$ obtained in the first step. More precisely, the action suggested by $\phi'$ at state $s$ is

**Algorithm 6** Policy Iteration

**procedure** PI($\mathcal{M}$)
  Choose $\phi'(s) \in \mathcal{A}$ arbitrarily for all $s \in \mathcal{S}$
  **repeat**
    $\phi(s) \leftarrow \phi'(s)$ for all $s \in \mathcal{S}$
    $(V^{\phi}(s), \theta^{\phi}) \leftarrow$ POLICYEVALUATION$(\mathcal{M}, \phi(s))$
    $\phi'(s) \leftarrow$ POLICYIMPROVEMENT$(\mathcal{M}, V^{\phi}(s))$
  **until** $\phi'(s) = \phi(s)$ for all $s \in \mathcal{S}$
  **return** $(\phi^*, \theta) \leftarrow (\phi(s), \theta^{\phi})$

determined by

$$\phi'(s) = \underset{a \in \mathcal{A}}{\operatorname{argmin}} \left\{ C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V^{\phi}(s') \right\}.$$

The pseudocode of the policy iteration algorithm is given in Algorithm 6. With policy iteration algorithm in mind, we can proceed with presenting the policy improvement theorem.

**Theorem 11** (Policy improvement theorem)**.** *Suppose that we have obtained the value function resulting from the operation of a policy $A$ and that the subsequent policy improvement step has produced a policy $B$, the following results hold.*

- *If $B$ is different from $A$, $\theta^A \geq \theta^B$.*
- *If $A$ and $B$ are equivalent, [3] both policies are optimal.*

*Proof.* The proof is based on [81, pp. 42-43]. The complete proof can be found in Appendix C.12. □

With the most important theorem proved, we proceed with finding $\phi^*$ theoretically. First, we simplify the Bellman equation shown in (4.3) to make the theoretical proof more concise and straightforward.

[3] Policies $A$ and $B$ are equivalent when they yield the same expected AoII.

### 4.5.4 Simplifying the Bellman Equation

We note that state transitions are complex and intertwined. Consequently, the direct analysis of the Bellman equation (4.3) is complicated. In the following, we will simplify the Bellman equation. To this end, we leverage the fact that the action space depends on the state space. More specifically, when the channel is busy (i.e., $i \neq -1$), the feasible action is $a = 0$. Hence, the transmitter's actions at these states are fixed, which leads to the fact that for these states, the minimum operators in (4.3) are avoided. Let $\mathcal{S}_{-1} \triangleq \{s : i = -1\}$ be the set of states at which the channel is idle. Then,

$$V(s) + \theta = \min_{a \in \mathcal{A}} \left\{ C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V(s') \right\} = C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(0) V(s') \quad s \in \mathcal{S} \setminus \mathcal{S}_{-1}. \tag{4.15}$$

Then, by repeatedly replacing the $V(s)$, where $s \in \mathcal{S} \setminus \mathcal{S}_{-1}$, with the expression given by (4.15), we can obtain the Bellman equation consists only $V(s)$ where $s \in \mathcal{S}_{-1}$. We know that $s = (\Delta, 0, -1)$ for $s \in \mathcal{S}_{-1}$. Hence, we abbreviate $V(\Delta, 0, -1)$ as $V(\Delta)$. Then, for each $\Delta \geq 0$, we have the following modified Bellman equation.

$$V(\Delta) + \theta = \min_{a \in \{0,1\}} \left\{ C(\Delta, a) - \theta(a) + \sum_{\Delta' \geq 0} P_{\Delta,\Delta'}(a) V(\Delta') \right\}, \tag{4.16}$$

where

$$\theta(a) = \begin{cases} 0 & a = 0, \\ (ET - 1)\theta & a = 1. \end{cases}$$

Note that $ET$, $P_{\Delta',\Delta}(a)$, and $C(\Delta, a)$ are those defined and discussion in Section 4.4. Hence, it is sufficient to use (4.16) instead of (4.3) to determine the optimal action at state $(\Delta, 0, -1)$. Although equation (4.16) may seem complicated at first glance, its advantages will be fully demonstrated in the following subsection.

### 4.5.5 Optimality Proof

In this subsection, we find $\phi^*$ theoretically. We first introduce the condition that is essential to the analysis later on.

**Condition 1.** *The condition is the following.*

$$\bar{\Delta}_1 \leq \min\left\{\bar{\Delta}_0, \frac{1+(1-p)\sigma}{2}\right\},$$

*where, for **Assumption 1**,*

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left(\frac{1-(1-p)^t}{p}\right)}{1-\sum_{t=1}^{t_{max}} pp_t(1-p)^{t-1}},$$

*and for **Assumption 2**,*

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left(\frac{1-(1-p)^t}{p}\right) + p_{t^+}\left(\frac{1-(1-p)^{t_{max}}}{p}\right)}{1-\left(\sum_{t=1}^{t_{max}} pp_t(1-p)^{t-1} + p_{t^+}(1-p)^{t_{max}}\right)}.$$

$\bar{\Delta}_0$ *and* $\bar{\Delta}_1$ *are the expected AoII resulting from the adoption of the threshold policy with* $\tau = 0$ *and* $\tau = 1$*, respectively.*

**Theorem 12.** *When Condition 1 is satisfied, the optimal policy for $\mathcal{M}$ is the threshold policy with $\tau = 1$.*

*Proof.* The value iteration algorithm detailed in Section 4.5.2 provides a good guess on the optimal policy. Then, we theoretically prove the optimality using the policy improvement theorem. The general procedure for the optimality proof can be summarized as follows.

1. *Policy Evaluation*: We calculate the value function resulting from the adoption of the threshold policy with $\tau = 1$.

2. *Policy Improvement*: We obtain a new policy using the value function obtained in the previous step and verify that the new policy is the threshold policy with $\tau = 1$.

Then, the policy improvement theorem tells us that the threshold policy with $\tau = 1$ is optimal. The complete proof can be found in Appendix C.13. □

**Remark 16.** *When the system fails to satisfy Condition 1, we can use the relative value iteration algorithm introduced in Section 4.5.2 to obtain a good estimate of $\phi^*$.*

## 4.6 Numerical Results

In this section, we numerically verify Condition 1 and analyze the performance of the optimal policy.

### 4.6.1 Verification of Condition 1

As the closed-form expressions of $\bar{\Delta}_0$ and $\bar{\Delta}_1$ are given in Section 4.4, the inequality in Condition 1 is easy to verify. We verify Condition 1 numerically for the following systems.

- The system adopts **Assumption 1/Assumption 2** and the transmission delay follows the Geometric distribution with success probability $p_s$. More precisely, $p_t = (1-p_s)^{t-1}p_s$.

- The system adopts **Assumption 1** and the transmission delay follows the Zipf distribution with constant $a$. More precisely, $p_t = \frac{t^{-a}}{\sum_{i=1}^{t_{max}} i^{-a}},\ 1 \leq t \leq t_{max}$.

- The system adopts **Assumption 1** and $p_t = \frac{1}{2}(\mathbb{1}\{t=1\} + \mathbb{1}\{t=t_{max}\})$.

For each of the above systems, the parameters take the following values.

- $0.05 \leq p \leq 0.45$ with step size being equal to $0.05$.

- $2 \leq t_{max} \leq 15$ with step size being equal to $1$.

- $0 \leq p_s \leq 0.95$ with step size being equal to $0.05$.

- $0 \leq a \leq 5$ with step size being equal to $0.25$.

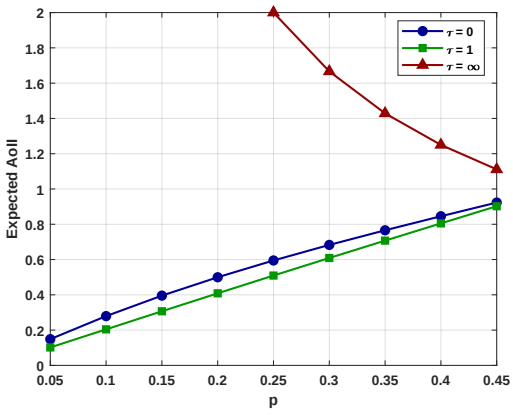
The numerical results show that all the above systems satisfy Condition 1. Then, according to Theorem 12, we can conclude that the corresponding optimal policy is the threshold policy with $\tau = 1$.

**Remark 17.** *The Zipf distribution reduces to the uniform distribution when $a = 0$, and the Geometric transmission delay reduces to a deterministic transmission delay when $p_s = 0$. We ignore the case of $p = 0$ because the dynamic source does not change state in this case. Similarly, we are not interested in the case of $p = 0.5$ because the state of the dynamic source is independent of the previous state in this case. Also, we exclude the case of $p_s = 1$ because, in this case, the transmission time is deterministic and equal to $1$ time slot. The corresponding optimal policies under various system settings are well studied in [36, 37, 39, 48, 64].*

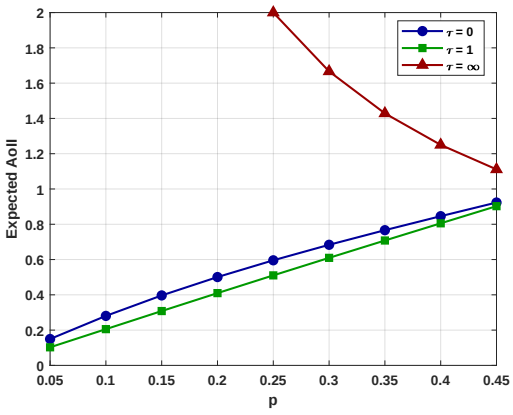
### 4.6.2 Performance of the Optimal Policy

In this subsection, we analyze the performance of the optimal policy. To this end, we consider the system where the transmission delay follows a Geometric distribution with success probability $p_s$. Moreover, we compare the performance of the optimal policy with that of the threshold policies with $\tau = 0$ and $\tau = \infty$. All the results are calculated using Section 4.4.

The effect of $p$: In this case, we fix $t_{max} = 5$ and $p_s = 0.7$. Then, we vary $p$ and plot the corresponding results in Figure 4.2. In the figure, to better show the performance of the optimal policy, we only show parts of the results for the threshold policy with $\tau = \infty$. We notice that, as $p$ increases, the expected AoIIs achieved by the threshold policies with $\tau = 0$ and $\tau = 1$ increase. This is because when $p$ is large, the dynamic source will be inclined to switch between states. Therefore, the state of the dynamic source is more unpredictable, leading to an increase in the achieved expected AoIIs. Meanwhile, the expected AoII achieved by the threshold policy with $\tau = \infty$ decreases as $p$ increases. To explain this, we first recall that, under the threshold policy with $\tau = \infty$, the receiver's estimate does not change. Also, when $p$ is large, the dynamic source will change states frequently. Therefore, the probability of a situation where the receiver's estimate is always incorrect is small, which makes the resulting AoII small. Also, we notice that **Assumption 1** and **Assumption 2** lead to almost the same performance. To explain this, we first note that the only difference between **Assumption 1** and **Assumption 2** is whether the update is delivered or discarded when the transmission lasts to the $t_{max}$th time slot after the start of the transmission. However, under our choices of $p_s$ and $t_{max}$, the transmission time of an update rarely reaches $t_{max}$ time slots. Even if it reaches $t_{max}$ time slots, delivery or discarding does




(a) Performance under **Assumption 1**.

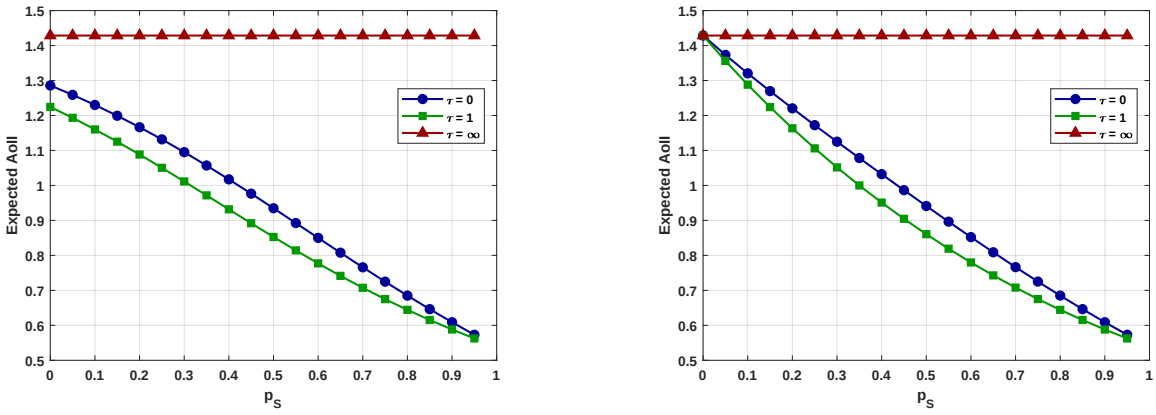



(b) Performance under **Assumption 2**.

Figure 4.2: Illustrations of the expected AoII as a function of $p$ and $\tau$. We set the upper limit of the transmission time $t_{max} = 5$ and the success probability in the Geometric distribution $p_s = 0.7$.

not significantly impact the performance, as the receiver's estimate can be correct or incorrect regardless of whether the update is delivered. Therefore, **Assumption 1** and **Assumption 2** yield almost the same performance.

The effect of $p_s$: In this case, we fix $t_{max} = 5$ and $p = 0.35$. Then, we vary $p_s$ and plot the corresponding results in Figure 4.3. The figure shows that the expected AoIIs achieved by the threshold policies with $\tau = 0$ and $\tau = 1$ decrease as $p_s$ increases. The reason behind this is as follows. As $p_s$ increases, the expected transmission time of an update decreases, meaning that updates are more likely to be delivered within the first few time slots. As a result, the receiver receives fresher information, and thus the expected AoII decreases. Moreover, the performance gap between the threshold policies with $\tau = 1$ and $\tau = 0$ is small when $p_s$ is large. To explain this, we notice that the threshold policy with $\tau = 0$ is not optimal because the updates transmitted when AoII is zero do not provide any new information to the receiver. Meanwhile, the transmission will occupy the channel for a few time slots. Therefore, such an action deprives the transmitter of the ability to send new updates for the next few time slots without providing the




(a) Performance under **Assumption 1**.

(b) Performance under **Assumption 2**.

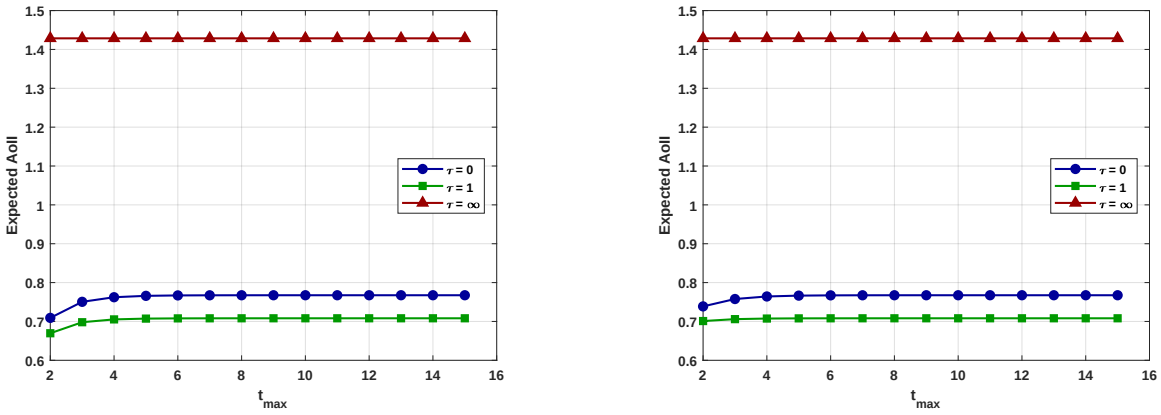
Figure 4.3: Illustrations of the expected AoII as a function of $p_s$ and $\tau$. We set the upper limit of the transmission time $t_{max} = 5$ and the source dynamic $p = 0.35$.

receiver with any new information. Hence, when $p_s$ is large, the expected transmission time of an update is small. Consequently, the transmission when AoII is zero becomes less costly. Hence, the gap narrows.

The effect of $t_{max}$: In this case, we fix $p_s = 0.7$ and $p = 0.35$. Then, we vary $t_{max}$ and plot the corresponding results in Figure 4.4. From the figure, we can see that the effect of $t_{max}$ on the performances is only noticeable when $t_{max}$ is small. This is because, under our choice of $p_s$, most updates will be delivered within the first few time slots. Therefore, increasing $t_{max}$ will not significantly affect the performance.

## 4.7 Conclusion

In this chapter, we investigate the problem of minimizing the Age of Incorrect Information over a channel with random delay. We study a slotted-time system where a transmitter observes a dynamic source and sends updates to a remote receiver over a channel with random delay. To




(a) Performance under **Assumption 1**.

(b) Performance under **Assumption 2**.

Figure 4.4: Illustrations of the expected AoII as a function of $t_{max}$ and $\tau$. We set the success probability in the Geometric distribution $p_s = 0.7$ and the source dynamic $p = 0.35$.

facilitate the analysis, we consider two cases. The first case assumes that the transmission time has an upper bound and that the update will always be delivered. The second case assumes that the system automatically discards updates if the transmission lasts too long. We aim to find when the transmitter should initiate transmission to minimize the AoII. To this end, we first characterize the optimization problem using the MDP and calculate the expected AoII achieved by the threshold policy precisely using the Markov chain. Next, we prove that the optimal policy exists and provide a computable relative value iteration algorithm to estimate the optimal policy. Then, with the help of the policy improvement theorem, we prove theoretically that, under Condition 1, the optimal policy is the threshold policy with $\tau = 1$. Finally, we numerically verify Condition 1 under various system parameters and analyze the performance of the optimal policy.

# Chapter 5: Freshness with Preemption

## 5.1 Overview

We recall that a fundamental assumption made in Chapter 4 is that the transmitter does not have the ability to preempt. In other words, the transmitter can only wait for the current transmission to complete before initiating a new transmission. Therefore, in this chapter, we build on the result in Chapter 4 and investigate the problem of finding the optimal policy when the transmitter is able to preempt. More precisely, we consider the case where the transmitter can preempt the transmission of updates and immediately transmit a new update. In this way, we also need to consider whether the transmitter needs to terminate the transmission to transmit new updates when the channel is busy.

The main contributions of this chapter can be summarized as follows. 1) We study the problem of optimizing the AoII with a generic time penalty function in a slotted-time system with a preemptible transmitter under a generic transmission delay. 2) We formulate the problem using the Markov decision process and prove the existence of the optimal policy. 3) We obtain the analytical expression of the expected AoII achieved by the strong preemptive policy. 4) We propose the value iteration algorithm that approximates the optimal policy. 5) We analyze two canonical transmission delays and theoretically find the corresponding optimal policies. Specifically, we first study the scenario where the transmission delay follows the Geometric

distribution, a typical example of an unbounded transmission time. Then, we investigate the case where the transmission delay follows the Zipf distribution, a typical example of a bounded transmission time. Finally, we extend the results to the generic transmission delay when the transmission time is upper bounded by 2.

This chapter summarizes the work in [82], and the rest is organized as follows. Section 5.2 describes the system model, specifies the choice of the penalty functions in AoII, and formulates the optimization problem. Then, we cast the optimization problem into a Markov decision process and specify the state transition probabilities in Section 5.3. In Section 5.4, we analyze the strong preemptive policy and obtain the analytical expression of the expected AoII it achieves. Then, in Section 5.5, we prove the existence of the optimal policy, propose a modified value iteration algorithm to approximate the optimal policy, and theoretically obtain the optimal policy when the transmission delay follows two canonical distributions. The chapter concludes with numerical results detailed in Section 5.6.

## 5.2 System Overview

### 5.2.1 System Model

We consider a system in which the transmitter observes a Markovian source by receiving updates from it and controls the transmission of the updates over a communication channel with random delays so that the receiver at the other end of the channel has the best real-time knowledge of the Markovian source. We assume that time is slotted and normalized to a unit time slot. Meanwhile, the end of one time slot is the beginning of the next time slot. An illustration of the system model is given in Figure 5.1. At the beginning of time slot $k$, the transmitter receives

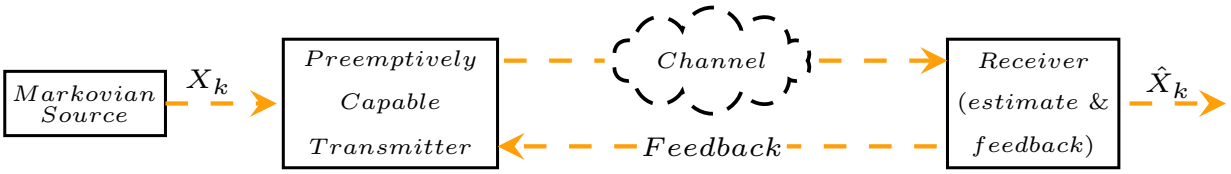



Figure 5.1: An illustration of the system model.

an update $X_k$ from the Markovian source and discards the old one. To not overcomplicate the system model, we assume that the Markovian source has two states and is symmetric with state transition probability $p$. Then, the transmitter decides whether to transmit $X_k$ based on the system's current status. When the channel is idle, the transmitter can choose whether to transmit the current update. We consider the case where the transmitter has the ability to terminate an ongoing transmission and immediately initiate a new one. We assume that terminating the current transmission and starting a new one can be done simultaneously, and the preempted update will be discarded. Thus, when the channel is busy, the transmitter can stay idle or terminate the ongoing transmission and immediately transmit the current update. We denote the action of the transmitter by $a_k$. $a_k = 0$ when the transmitter chooses to stay idle. Otherwise, $a_k = 1$. Note that the specific action represented by $a_k = 1$ depends on the status of the communication channel. If not preempted, the transmission will arrive at the receiver after a random amount of time slots. The receiver maintains an estimate of the state of the Markovian source based on the received updates. Whenever the receiver receives a new update, it sends a feedback signal to the transmitter to inform the transmitter that the update has been received.

In the following, we define the delay model. The communication channel is reliable but suffers from random delays, which means that if an update is not preempted, it will be delivered to the receiver losslessly after a random amount of time slots. The transmission time is a random variable denoted by $T$. For simplicity, we assume that $T$ is independent and identically distributed

for each update. The transmission time can be fully characterized by the probability mass function (PMF) denoted by $p_t \triangleq Pr(T = t)$, where $t \in \mathbb{N}^*$. We do not impose any restrictions on the transmission time, which means that the transmission time can be infinite.

Next, we describe the receiver's estimation strategy and the feedback mechanism. Let $\hat{X}_k$ denote the receiver's estimate at time slot $k$. Then, according to [75], the best estimator when $p \leq \frac{1}{2}$ is the last received update. Let $d_k = 1$ when an update is delivered to the receiver at time slot $k$ and $d_k = 0$ otherwise. Then,

$$\hat{X}_k = \begin{cases} \hat{X}_{k-1} & d_{k-1} = 0, \\ X_{k-T} & d_{k-1} = 1, \end{cases}$$

where $X_{k-T}$ is the update delivered at the end of time slot $k-1$. In the case of $p > \frac{1}{2}$, the best estimator depends on the realization of the transmission time. In this chapter, we consider only the case of $0 < p < \frac{1}{2}$. We exclude the case of $p = 0$ because, in this case, the Markovian source never changes state. We also exclude the case of $p = \frac{1}{2}$ because, in this case, the state of the Markovian source is independent of the previous state. The results can be extended to the case of $p > \frac{1}{2}$ by using the corresponding best estimator. Whenever the receiver receives a new update, it sends an $ACK$ packet to the transmitter so that the transmitter is aware of the change in the receiver's estimate and that the channel has become idle. In real-world applications, the size of $ACK$ packets is usually negligible compared to that of status updates. Therefore, we assume that $ACK$ packets are accurately and instantaneously received by the transmitter. This assumption is widely used in the relevant literature [11, 36]. The $ACK$ packet alone is sufficient for the transmitter to keep track of the receiver's estimate since we assume that the update will

necessarily be delivered unless it is preempted.

The system uses AoII to measure its performance. Since the dynamic source has two states, we let $X_t \in \{0,1\}$ and $\hat{X}_t \in \{0,1\}$. To avoid unnecessary complications, we choose $g(X_k, \hat{X}_k) = |X_k - \hat{X}_k| \in \{0,1\}$. Consequently, AoII can be written as

$$\Delta_{AoII}(X_k, \hat{X}_k, k) = f(k - U_k) \triangleq f(\Delta_k), \tag{5.1}$$

where $U_k$ is defined in Section 1.2.2. To facilitate the analysis, we make the following assumptions on the time penalty function $f(\Delta)$.

- $f(\Delta_1) \geq f(\Delta_2) \geq 0$ if $\Delta_1 \geq \Delta_2$.
- $f(\Delta) \to \infty$ when $\Delta \to \infty$.
- $\sum_{\Delta=0}^{\infty} \gamma^{\Delta} f(\Delta) < \infty$ for $0 < \gamma < 1$.

Some choices of the time penalty function $f(\Delta)$ are the following.

- $f(\Delta) = \alpha\Delta + \beta$ where $\alpha \geq 0$ and $\beta \geq 0$ are finite constants.
- $f(\Delta) = \kappa\Delta^2$ where $\kappa \geq 0$ is finite constant.
- $f(\Delta) = \log_a(\Delta + 1)$ where $a > 1$ is finite constant.

We notice that the evolution of $\Delta_k$ can fully characterize the evolution of $f(\Delta_k)$. Leveraging the definition of $U_k$, the evolution of $\Delta_k$ can be characterized by the following two cases.

- When the receiver's estimate is correct at time slot $k$, $U_k = k$ by definition. Hence, $\Delta_k = 0$.
- When the receiver's estimate is erroneous at time slot $k$, $U_k = U_{k-1}$ by definition. Hence, $\Delta_k = k - U_k = \Delta_{k-1} + 1$.

The evolution of $\Delta_k$ can be summarized as follows.

$$\Delta_k = \mathbb{1}\{X_k \neq \hat{X}_k\}(\Delta_{k-1} + 1).$$

We can conclude that $\Delta_k = 0$ if and only if $X_k = \hat{X}_k$. For later use in analysis, we reformulate the evolution of $\Delta_k$ by incorporating the dynamics of the Markovian source. To this end, we first define $\Gamma \triangleq \mathbb{1}\{(\Delta_{k-1} = 0 \wedge \hat{X}_k = \hat{X}_{k-1}) \vee (\Delta_{k-1} > 0 \wedge \hat{X}_k \neq \hat{X}_{k-1})\}$. [1] Then, we know that $\hat{X}_k = X_{k-1}$ if $\Gamma = 1$ and $\hat{X}_k \neq X_{k-1}$ if $\Gamma = 0$. Hence, we have

$$\Delta_k = \begin{cases} \mathbb{1}\{X_k \neq X_{k-1}\}(\Delta_{k-1} + 1) & \Gamma = 1, \\ \mathbb{1}\{X_k = X_{k-1}\}(\Delta_{k-1} + 1) & \Gamma = 0. \end{cases} \tag{5.2}$$

Note that the system's state at any time slot can correspond to either $\Gamma = 1$ or $\Gamma = 0$.

### 5.2.2 Problem Formulation

In this chapter, we investigate the problem of minimizing the AoII by controlling the transmitter's decision in each time slot. We define a policy as one that specifies the transmitter's decision in each time slot based on the current system status. Then, this chapter aims to find a policy that minimizes the AoII of the system. Mathematically, the problem can be formulated as the following minimization problem.

$$\underset{\psi \in \Psi}{\arg\min} \qquad \lim_{K \to \infty} \frac{1}{K} \mathbb{E}_{\psi} \left( \sum_{k=0}^{K-1} f(\Delta_k) \right), \tag{5.3}$$

[1] In the definition, $\wedge$ is the logical AND operator and $\vee$ is the logical OR operator.

where $\mathbb{E}_\psi$ is the conditional expectation, given that policy $\psi$ is adopted and $\Psi$ is the set of all admissible policies.

**Definition 9** (Optimal policy)**.** *A policy is optimal if it minimizes the AoII of the system. The action specified by the optimal policy is called the optimal action. We use $\psi^*$ and $a^*$ to denote the optimal policy and action, respectively.*

In the next section, we characterize the optimization problem (5.3) using the Markov decision process (MDP).

## 5.3 MDP Characterization

We use an infinite horizon with average cost MDP $\mathcal{M}$ to characterize the minimization problem (5.3). Specifically, $\mathcal{M}$ consists of the following components.

- The state space $\mathcal{S}$. The state $s = (\Delta, t, i)$ where $\Delta \in \mathbb{N}^0$ is the $\Delta_k$ defined in (5.1) without the time stamp. $t \in \mathbb{N}^0$ denotes the time that the transmission has been in progress. When the channel is idle, we define $t = 0$. $i \in \{-1, 0, 1\}$ indicates the channel status. When the channel is idle, $i = -1$. When the channel is busy transmitting an update, $i \in \{0, 1\}$, where $i = 0$ if the update being transmitted is the same as the receiver's current estimate. Otherwise, $i = 1$. To better distinguish between different states, we will use $s$ and $(\Delta, t, i)$ interchangeably to represent the state throughout the rest of the chapter. Therefore, $s$ and $(\Delta, t, i)$ will synchronize all changes, such as adding superscripts or subscripts.

- The action space $\mathcal{A}$. The feasible action is $a \in \{0, 1\}$. When $i \neq -1$, $a = 1$ if the transmitter decides to terminate the current transmission and immediately start a new one.

Otherwise, $a = 0$. When $i = -1$, $a = 1$ if the transmitter decides to transmit the new update, and $a = 0$ otherwise.

- The state transition probability $\mathcal{P}$. The probability that action $a$ at state $s$ leads to state $s'$ is denoted by $P_{s,s'}(a)$. The value of $P_{s,s'}(a)$ will be discussed in the next subsection.

- The immediate cost $\mathcal{C}$. The immediate cost for being at state $s$ is $C(s) = f(\Delta)$.

Let $V(s)$ be the value function of state $s$. Then, the optimal action at state $s$, denoted by $a^*(s)$, can be determined by the following equation.

$$a^*(s) = \underset{a \in \mathcal{A}}{\operatorname{argmin}} \left\{ \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V(s') \right\} \quad s \in \mathcal{S}.$$

Hence, computing the value function for each state $s \in \mathcal{S}$ is sufficient to obtain the optimal policy. It is well known that $V(s)$ satisfies the Bellman equation.

$$V(s) + \theta = \min_{a \in \mathcal{A}} \left\{ C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V(s') \right\} \quad s \in \mathcal{S},$$

where $\theta$ is the expected AoII achieved by the optimal policy. Hence, the state transition probability $P_{s,s'}(a)$ plays a vital role. In the following, we delve into the expression of $P_{s,s'}(a)$.

We first define and compute an auxiliary quantity $Pr(T = t \mid t-1)$, which is the probability that an update will be delivered in the next time slot, given that the transmission has been in progress for $t - 1$ time slots. It is easy to get

$$Pr(T = t \mid t - 1) = \frac{p_t}{1 - \sum_{i=1}^{t-1} p_i}.$$

For simplicity, we abbreviate $Pr(T = t \mid t-1)$ as $q_t$ for the remainder of this chapter. Leveraging $q_t$, we can proceed with deriving the state transition probability $P_{s,s'}(a)$. For the sake of space, the detailed discussion is provided in Appendix D.1.

## 5.4 Strong Preemptive Policy

In this section, we analyze and evaluate the performance of the strong preemptive policy by calculating the expected AoII it achieves.

**Definition 10** (Strong preemptive policy)**.** *The strong preemptive policy always starts a new transmission when the channel is idle and always preempts the transmitting update.*

**Remark 18.** *We consider the strong preemptive policy because it is intuitively desirable when the transmission delay follows a memoryless distribution. For example, the Geometric distribution. One of the essential properties of the Geometric distribution is that $q_t$ is independent of $t$, meaning that no matter how long an update has been in transmission, it has the same probability of being delivered in the next time slot. Therefore, it is desirable for the transmitter to preempt so that the update in the channel is always the freshest.*

The system dynamics under the strong preemptive policy can be fully characterized by a discrete-time Markov chain (DTMC). Without loss of generality, we assume the system starts at state $(0, 0, -1)$. Then, the state space of the induced DTMC $\mathcal{S}_{sp}^{MC}$ consists of all the states that are accessible from state $(0, 0, -1)$. For a better presentation, we introduce the *virtual state*. By definition, the DTMC will never visit the virtual state. Nevertheless, the existence of these virtual states will make the equations clearer. In the following, we elaborate on the state space $\mathcal{S}_{sp}^{MC}$ and identify the virtual states. We first recall that the strong preemptive policy always preempts the

transmitting updates. Hence, each update can only live for one slot in the channel before being preempted or delivered. Consequently, the DTMC will never reach state $s$ with $t > 1$. Hence, the system can only be in state $s$ with $t \in \{0, 1\}$. With this in mind, $\mathcal{S}_{sp}^{MC}$ consists of the following states.

- $s = (\Delta, 0, -1)$ where $\Delta \geq 0$. The DTMC will be in this state every time the channel is idle.
- $s = (\Delta, 1, 0)$ where $\Delta \geq 0$. The DTMC will be in this state when the channel is busy transmitting an update that is the same as the receiver's estimate. Then, we identify the virtual state. We note that $i = 0$ happens only when the transmitter initiates the transmission when AoII is zero. We recall that the transmitting update is either delivered or preempted one time slot after the transmission starts. Combined with the fact that, within one time slot, AoII can either increase by $1$ or decrease to zero, we know that $s = (\Delta, 1, 0)$ where $\Delta \geq 2$ is a virtual state.
- $s = (\Delta, 1, 1)$ where $\Delta \geq 0$. The DTMC will be in this state when the channel is busy transmitting an update that differs from the receiver's estimate. Then, we identify the virtual state. We notice that $i = 1$ occurs when the transmitter initiates the transmission when the AoII is not zero. Combined with the fact that, within one time slot, AoII either increases by $1$ or decreases to zero, we know that $s = (1, 1, 1)$ is a virtual state.

We notice that all the states in $\mathcal{S}_{sp}^{MC}$, except for the virtual states, communicate with state $(0, 0, -1)$. Combined with the fact that state $s = (0, 0, -1)$ is recurrent, we can conclude that the stationary distribution of the induced DTMC exists. We denote by $\pi_{-1}(\Delta)$ the steady state probability of state $s = (\Delta, 0, -1)$. Likewise, we denote the steady state probability of state $s = (\Delta, 1, 0)$

and state $s = (\Delta, 1, 1)$ as $\pi_0(\Delta)$ and $\pi_1(\Delta)$, respectively. For virtual states, we define the corresponding steady state probability as 0. We also define $\pi(\Delta) \triangleq \pi_{-1}(\Delta) + \pi_0(\Delta) + \pi_1(\Delta)$. Then, the expected AoII achieved by the strong preemptive policy is given by

$$\bar{\Delta}_{sp} = \sum_{\Delta=0}^{\infty} f(\Delta)\pi(\Delta). \tag{5.4}$$

Hence, it is sufficient to calculate $\pi(\Delta)$ for $\Delta \geq 0$.

**Lemma 9.** *The following gives the expressions of* $\pi(\Delta)$ *for* $\Delta \geq 0$.

$$\pi(0) = \frac{p + q_1 - 2q_1p}{1 - (1 - q_1)(1 - 2p)}.$$

*For each* $\Delta \geq 1$,

$$\pi(\Delta) = \frac{(1 - q_1 - p + 2q_1p)^{\Delta-1}(p^2 + q_1p - 2q_1p^2)}{1 - (1 - q_1)(1 - 2p)}.$$

*Proof.* The balance equations the steady state probabilities satisfy can be obtained easily by exploiting the state transition probabilities discussed and detailed in Appendix D.1. Then, the stationary distribution can be obtained by solving the resulting system of linear equations. The complete proof can be found in Appendix D.2. □

**Remark 19.** *We can approximate the infinite sum in* (5.4) *using a finite sum with a sufficiently large upper bound on* $\Delta$.

We notice that

$$\begin{aligned}\bar{\Delta}_{sp} =& f(0)\pi(0) + c\sum_{\Delta=1}^{\infty}(1-q_1-p+2q_1p)^{\Delta}f(\Delta)\\ <& f(0)\pi(0) + c\sum_{\Delta=0}^{\infty}(1-q_1-p+2q_1p)^{\Delta}f(\Delta)\\ <& \infty,\end{aligned}$$

where $c \triangleq \frac{p^2+q_1p-2q_1p^2}{(1-q_1-p+2q_1p)[1-(1-q_1)(1-2p)]} \in [0,1]$. Then, the finiteness of $\bar{\Delta}_{sp}$ is guaranteed by the assumption on $f(\Delta)$ introduced in Section 5.2.1. In the following, we provide the closed-form expression of the expected AoII achieved by the strong preemptive policy when $f(\Delta) = \alpha\Delta + \beta$, where $\alpha \geq 0$ and $\beta \geq 0$ are two finite constants.

**Corollary 7.** *When* $f(\Delta) = \alpha\Delta + \beta$,

$$\bar{\Delta}_{sp} = \frac{\alpha p}{(p+q_1-2q_1p)(q_1+2p-2q_1p)} + \beta.$$

*Proof.* To deal with the infinite sum, we introduce an auxiliary quantity $\Sigma \triangleq \sum_{\Delta=2}^{\infty}\Delta\pi(\Delta)$. Then, $\bar{\Delta}_{sp} = \alpha(\pi(1)+\Sigma)+\beta$. To obtain the closed-form expression of $\Sigma$, we introduce another auxiliary quantity $\Pi \triangleq \sum_{\Delta=2}^{\infty}\pi(\Delta)$, whose closed-form expression can be obtained using Lemma 9. The complete proof can be found in Appendix D.3. □

## 5.5 Optimal Policy

In this section, we first prove the existence of the optimal policy. Then, we provide a feasible relative value iterative algorithm that approximates the optimal policy. Next, using the

policy improvement theorem, we analyze the optimization problem (5.3) and theoretically find the optimal policy when the transmission delay follows the Geometric distribution and the Zipf distribution, respectively.

### 5.5.1 Existence of the Optimal Policy

For the $\mathcal{M}$ in Section 5.3, we define the expected $\gamma$-discounted cost under policy $\psi$ as

$$V_{\psi,\gamma}(s) = \mathbb{E}_{\psi}\left[\sum_{k=0}^{\infty}\gamma^{t}C(s_k) \mid s\right],$$

where $0 < \gamma < 1$ is a discount factor and $s_k$ is the state of $\mathcal{M}$ at time $k$. Let $V_{\gamma}(s)$ be the value function associated with $\mathcal{M}$ under $\gamma$-discounted cost. Then, we know that $V_{\gamma}(s) = inf_{\psi}V_{\psi,\gamma}(s)$. Moreover, $V_{\gamma}(s)$ satisfies the Bellman equation.

$$V_{\gamma}(s) = \min_{a\in\mathcal{A}}\left\{C(s) + \gamma\sum_{s'\in\mathcal{S}}P_{s,s'}(a)V_{\gamma}(s')\right\} \quad s \in \mathcal{S}.$$

The value iteration algorithm is one of the most commonly used algorithms to calculate the value function. Let $V_{\gamma,\nu}(s)$ be the estimated value function at iteration $\nu$. Then, the estimated value function is updated in the following way.

$$V_{\gamma,\nu+1}(s) = \min_{a\in\mathcal{A}}\left\{C(s) + \gamma\sum_{s'\in\mathcal{S}}P_{s,s'}(a)V_{\gamma,\nu}(s')\right\} \quad s \in \mathcal{S}. \tag{5.5}$$

Without loss of generality, we initialize $V_{\gamma,0}(s) = 0$ for $s \in \mathcal{S}$. Then, we can prove the following lemma.

**Lemma 10.** *When updated following* (5.5)*,* $\lim_{\nu\to\infty} V_{\gamma,\nu}(s) = V_\gamma(s)$ *for* $s \in \mathcal{S}$*.*

*Proof.* According to [77, Propositions 1 and 3], it is sufficient to show that $V_\gamma(s)$ is finite. To this end, we have

$$V_{\psi,\gamma}(s) = \mathbb{E}_\psi\left[\sum_{k=0}^{\infty}\gamma^k C(s_k) \mid s\right] \leq \sum_{k=0}^{\infty}\gamma^k f(\Delta+k) = \frac{1}{\gamma^\Delta}\sum_{k=\Delta}^{\infty}\gamma^k f(k) \leq \frac{1}{\gamma^\Delta}\sum_{k=0}^{\infty}\gamma^k f(k) < \infty.$$

The finiteness is guaranteed by the assumption on $f(\Delta)$ introduced in Section 5.2.1. Then, by definition, we have $V_\gamma(s) \leq V_{\gamma,\psi}(s) < \infty$. Hence, we can conclude that the value iteration reported in (5.5) will converge to the value function. □

Leveraging the iterative nature of the value iteration algorithm, we can prove the following structural property of $V_\gamma(s)$.

**Lemma 11.** $V_\gamma(s)$ *is non-decreasing in* $\Delta > 0$*.*

*Proof.* Given the convergence proved in Lemma 10, the monotonicity of $V_\gamma(s)$ can be proved via mathematical induction. The complete proof can be found in Appendix D.4. □

Now, we proceed with showing the existence of the optimal policy. To this end, we first define the stationary policy.

**Definition 11** (Stationary policy)**.** *A stationary policy specifies a single action in each time slot.*

**Theorem 13.** *There exists a stationary policy* $\psi$ *that is optimal for* $\mathcal{M}$*. Moreover, the minimum expected AoII is independent of the initial state.*

*Proof.* The proof follows the same steps as in [77, Theorem 4]. We define $h_\gamma(s) \triangleq V_\gamma(s) - V_\gamma(s^{ref})$ as the relative value function, where $s^{ref}$ is the reference state. Note that the reference

state is arbitrary but fixed. Then, we show that $\mathcal{M}$ satisfies the two conditions given in [77]. To avoid excessive repetition of the proof, we omit the specific reasoning and give only the proofs of two important intermediate results.

1. $h_\gamma(s)$ is non-decreasing in $\Delta$ when $\Delta > 0$. The result follows directly from Lemma 11 because the reference state is fixed.

2. There exists a policy $\psi$ that induces an irreducible ergodic Markov chain, and the expected cost is finite. $\psi$ can be the strong preemptive policy. Then, the result is true as discussed in Section 5.4.

Using these two results, we can verify that $\mathcal{M}$ satisfies the two conditions given in [77]. Then, the existence of the optimal policy is guaranteed by [77, Theorem]. Moreover, the minimum expected cost is independent of the initial state. □

### 5.5.2 Value Iteration Algorithm

In this section, we present the relative value iteration (RVI) algorithm that approximates the optimal policy for $\mathcal{M}$. Direct application of RVI becomes impractical as the state space $\mathcal{S}$ of $\mathcal{M}$ is infinite. Hence, we construct another $\mathcal{M}^{(m)} = (\mathcal{S}^{(m)}, \mathcal{A}, \mathcal{P}^{(m)}, \mathcal{C})$ by truncating the value of $\Delta$ and $t$. More precisely, we impose

$$\mathcal{S}^{(m)}: \begin{cases} \Delta \in \{0, 1, ..., \Delta_{max}\}, \\ i \in \{-1, 0, 1\}, \\ t \in \{0, 1, ..., t_{max}\}, \end{cases}$$

where $\Delta_{max}$ and $t_{max}$ are the predetermined maximal value of $\Delta$ and $t$, respectively. Then, the size of the state space reduces from infinite to $((\Delta_{max}+1)\times 3\times(t_{max}+1))$. The transition probabilities from $s\in\mathcal{S}^{(m)}$ to $z\in\mathcal{S}\setminus\mathcal{S}^{(m)}$ are redistributed to the states $s'\in\mathcal{S}^{(m)}$ according to

$$P_{s,s'}^{(m)}(a)=\begin{cases}P_{s,s'}(a) & \Delta'<\Delta_{max}\ and\ t'<t_{max},\\ P_{s,s'}(a)+\sum_{G_1(z,s')}P_{s,z}(a) & \Delta'=\Delta_{max}\ and\ t'<t_{max},\\ P_{s,s'}(a)+\sum_{G_2(z,s')}P_{s,z}(a) & \Delta'<\Delta_{max}\ and\ t'=t_{max},\\ P_{s,s'}(a)+\sum_{G_3(z,s')}P_{s,z}(a) & \Delta'=\Delta_{max}\ and\ t'=t_{max},\end{cases} \tag{5.6}$$

where $G_1(s,s')=\{s:\Delta>\Delta_{max},t=t',i=i',\}$, $G_2(s,s')=\{s:\Delta=\Delta',t>t_{max},i=i'\}$, and $G_3(s,s')=\{s:\Delta>\Delta_{max},t>t_{max},i=i'\}$. The action space $\mathcal{A}$ and the immediate cost $\mathcal{C}$ are the same as defined in $\mathcal{M}$.

We can rigorously show that the sequence of optimal policies for $\mathcal{M}^{(m)}$ will converge to the optimal policy for $\mathcal{M}$ as $\Delta_{max}\to\infty$ and $t_{max}\to\infty$. More specifically, we can show that our system verifies the two assumptions given in [52]. Then, by [52, Theorem 2.2], we know the following results hold.

- There exists an average cost optimal stationary policy for $\mathcal{M}^{(m)}$.
- Any limit point of the sequence of optimal policies for $\mathcal{M}^{(m)}$ is optimal for $\mathcal{M}$.

Considering the similarity of the system model, the proof will be similar to the proof of [48, Theorem 1]. Therefore, we only prove one of the most important lemmas in the proof and omit the rest of the proof. Let $V_\gamma^{(m)}(s)$ be the value function associated with $\mathcal{M}^{(m)}$ under $\gamma$-discounted cost.

**Algorithm 7** Relative Value Iteration

**procedure** RVI($\mathcal{M}^{(m)}$,$\epsilon$)
  $\nu \leftarrow 0$; $V_\nu(s) \leftarrow 0$ for $s \in \mathcal{S}^{(m)}$
  Choose $s^{ref} \in \mathcal{S}^{(m)}$ arbitrarily
  **repeat**
    **for** $s \in \mathcal{S}^{(m)}$ **do**
      **for** $a \in \mathcal{A}$ **do**
        $Q_a(s) \leftarrow C(s) + \sum_{s'} P^{(m)}_{s,s'}(a) V_\nu(s')$
      $Q(s) \leftarrow \min_a\{Q_a(s)\}$
    **for** $s \in \mathcal{S}^{(m)}$ **do**
      $V_{\nu+1}(s) \leftarrow Q(s) - Q(s^{ref})$
    $\nu \leftarrow \nu + 1$
  **until** $\max_s\{|V_\nu(s) - V_{\nu-1}(s)|\} \leq \epsilon$
  **return** $\hat{\psi}^* \leftarrow \text{argmin}_a\{Q_a(s)\}$

**Lemma 12.** *$V^{(m)}_\gamma(s)$ is non-decreasing in $\Delta > 0$.*

*Proof.* The proof is very similar to the proof of Lemma 11 because the way of redistributing the state transition probabilities shown in (5.6) does not change the structural properties of the state transition probability presented in the proof of Lemma 11. Therefore, we omit the proof of this lemma. □

Then, we can apply RVI to the truncated MDP $\mathcal{M}^{(m)}$ and treat the resulting optimal policy as an approximation of the optimal policy for $\mathcal{M}$. The pseudocode of RVI is given in Algorithm 7. A similar approximation is also used in [38], according to which small $\Delta_{max}$ and $t_{max}$ can give an accurate estimate of the optimal policy for $\mathcal{M}$. However, the choices of $\Delta_{max}$ and $t_{max}$ are still problematic. If large values are chosen, the state space of $\mathcal{M}^{(m)}$ grows rapidly. Meanwhile, the RVI may result in a non-optimal policy if the chosen value is small. Hence, in the following subsections, we investigate two specific examples of the transmission delay, namely the transmission delay that follows the Geometric distribution and the Zipf distribution. For these two common delay models, we theoretically find their optimal policies. For this purpose, we first

**Algorithm 8** Policy Iteration

**procedure** PI($\mathcal{M}$)
  Choose $\psi'(s) \in \mathcal{A}$ for $s \in \mathcal{S}$ arbitrarily
  **repeat**
    $\psi \leftarrow \psi'$
    $V^{\psi}(s) \leftarrow$ POLICYEVALUATION$(\mathcal{M}, \psi)$
    $\psi' \leftarrow$ POLICYIMPROVEMENT$(\mathcal{M}, V^{\psi}(s))$
  **until** $\psi' = \psi$
  **return** $\psi^* \leftarrow \psi$

introduce the policy iteration algorithm and the policy improvement theorem.

### 5.5.3 Policy Iteration Algorithm

The pseudocode of policy iteration algorithm is given in Algorithm 8. We elaborate on the POLICYEVALUATION function and the POLICYIMPROVEMENT function.

- The POLICYEVALUATION function takes the MDP $\mathcal{M}$ and the policy $\psi$ as input and produce the value function $V^{\psi}(s)$ and the expected AoII $\theta^{\psi}$ resulting from the adoption of $\psi$. To be more specific, $V^{\psi}(s)$ and $\theta^{\psi}$ are the solution to the following system of linear equations.

$$V^{\psi}(s) + \theta^{\psi} = C(s) + \sum_{s' \in \mathcal{S}} P^{\psi}_{s,s'} V^{\psi}(s') \quad s \in \mathcal{S}, \tag{5.7}$$

  where $P^{\psi}_{s,s'}$ is the probability that the system will transit from state $s$ to state $s'$ under policy $\psi$. Note that (5.7) forms a underdetermined system. Hence, we can select a reference state $s^{ref}$ arbitrarily and set $V^{\psi}(s^{ref}) = 0$. In this way, we can obtain a unique solution.

- The POLICYIMPROVEMENT function takes the MDP $\mathcal{M}$ and the value function $V^{\psi}(s)$ as input and produce the optimal policy $\psi'$ under $V^{\psi}(s)$. Let $\psi'(s)$ be the action suggested by

the new policy $\psi'$ at state $s$. Then, $\psi'(s)$ is given by the following equation.

$$\psi'(s) = \underset{a \in \mathcal{A}}{\operatorname{argmin}} \left\{ C(s) + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V^{\psi}(s') \right\}.$$

The policy iteration algorithm iterates between the two functions until convergence, the criterion of which is defined at line 7 of Algorithm 8. Although the policy iteration algorithm appears to be more computationally demanding than the value iteration algorithm and is not as commonly used as the value iteration algorithm, it has the advantage that we can prove the policy improvement theorem, which will be the basis of our theoretical analysis in the next two subsections.

**Theorem 14** (Policy improvement theorem)**.** *Suppose that we have obtained the value function resulting from the operation of a policy $A$ and that the policy improvement function has produced a new policy $A'$. When policy $A$ and policy $A'$ are identical, we say the policy improvement function converges and policy $A$ is optimal.*

*Proof.* The proof is based on [81, pp.42-43]. We first assume that the policy improvement function converges to a non-optimal policy $A$. Then, we prove that there is a contradiction under this assumption. Thus, the assumption that $A$ is non-optimal must be false, and its opposite must be true. The complete proof can be found in Appendix D.5. □

With the policy iteration algorithm and Theorem 14 in mind, we can proceed with finding the optimal policy through theoretical analysis.

### 5.5.4 Geometric Delay

In this subsection, we consider the case where the transmission delay is Geometrically distributed with success probability $0 < p_s < 1$. More precisely,

$$p_t = p_s(1 - p_s)^{t-1} \quad t \geq 1.$$

**Remark 20.** *We omit the case of $p_s = 0$ as, in this case, the update will never be delivered. We also do not discuss the case of $p_s = 1$ because the transmission time is deterministic and normalized in this case. The corresponding optimal policy has been well studied in related literature [36, 37, 48].*

Under Geometric distribution, $q_t = p_s$ for $t \geq 1$. Then, leveraging the policy improvement theorem, we can prove the following theorem.

**Theorem 15.** *The strong preemptive policy is optimal if the transmission delay follows the Geometric distribution.*

*Proof.* According to Theorem 14, it is sufficient to prove that the policy iteration algorithm converges to the strong preemptive policy. Specifically, we first calculate the value function resulting from the strong preemption policy. Then we use the resulting value function to derive a new policy and verify that the old and new policies are the same. The complete proof can be found in Appendix D.6. □

**Remark 21.** *The optimality of the strong preemptive policy is intuitive since $q_t$ is independent of $t$, which means that the probability of an update being delivered is independent of how long it*

*has been in transmission. Thus, the strong preemption policy ensures that updates in the channel will always be the latest ones without sacrificing the likelihood of their delivery.*

### 5.5.5 Zipf Delay

In this subsection, we consider the case where the transmission delay follows the Zipf distribution with the constant $a$. More precisely,

$$p_t = \frac{t^{-a}}{\sum_{i=1}^{t_{max}} i^{-a}} \quad 1 \leq t \leq t_{max},$$

where $t_{max} > 1$ is a predetermined constant.

**Remark 22.** *When $t_{max} = 1$, the transmission time is deterministic and normalized. Hence, we omit the discussion on this case for the same reason detailed in Remark 20.*

A transmission delay that follows the Zipf distribution is also considered in the literature on information freshness [38, 83]. Under the Zipf distribution,

$$q_t = \frac{t^{-a}}{\sum_{i=t}^{t_{max}} i^{-a}} \quad 1 \leq t \leq t_{max}.$$

We define that $q_t \triangleq 0$ when $t > t_{max}$. Hence, the transmission time of an update is upper bounded by $t_{max}$. Consequently, the state $s = (\Delta, t, i)$ in the corresponding $\mathcal{M}$ satisfies $0 \leq t \leq t_{max} - 1$. To simplify the analysis, we consider the case of $f(\Delta) = \alpha\Delta + \beta$. Before the optimal policy, we first introduce the threshold preemptive policy and evaluate its performance.

**Definition 12** (Threshold preemptive policy)**.** *The threshold preemptive policy always starts a new transmission when the channel is idle and does not preempt updates only at state $s =$*

$(\Delta, t_{max}-1, 1)$ *where* $\Delta \geq 1$.

The following theorem gives the expected AoII achieved by the threshold preemptive policy when $f(\Delta) = \alpha\Delta + \beta$.

**Theorem 16.** *When $f(\Delta) = \alpha\Delta + \beta$, the expected AoII achieved by the threshold preemptive policy $\bar{\Delta}_{tp}$ is give by*

$$\bar{\Delta}_{tp} = \frac{\alpha p}{(p + q_1 - 2q_1p)(q_1 + 2p - 2q_1p)} + \beta.$$

*Proof.* Although the threshold preemptive policy and the strong preemptive policy are not exactly the same, they yield the same expected AoII. This is because the actions suggested by the two policies differ only in the virtual states, which does not affect the long-term average performance.

□

We first introduce the following condition for direct use in the subsequent theoretical analysis.

**Condition 2.** *The conditions are the following.*

- $q_1 \geq q_t$ *for* $1 \leq t \leq t_{max} - 2$.
- *When* $t_{max} \geq 3$, $\mathcal{Q}_1 \geq 0$, $\mathcal{Q}_2 \geq 0$, *and* $\mathcal{Q}_3 \geq 0$, *where* $\mathcal{Q}_1$, $\mathcal{Q}_2$, *and* $\mathcal{Q}_3$ *are given by*

$$\begin{aligned}
\mathcal{Q}_1 \triangleq & (q_{t_{max}-1} - q_{t_{max}-1}p - p) + (1 - q_{t_{max}-1})p(q_1 + 2p - 2q_1p)^2. \\
\mathcal{Q}_2 \triangleq & \frac{(1-2p)\{(q_1 - 1) + (1 - q_{t_{max}-1})[p + q_1(1-p)]\}}{q_1 + p - 2q_1p}. \\
\mathcal{Q}_3 \triangleq & \frac{(1-q_1)(2p-1) - p(1 - q_{t_{max}-1})}{(2p + q_1 - 2q_1p)(p + q_1 - 2q_1p)} + \frac{(1 - q_{t_{max}-1})(1-p)p}{q_1 + p - 2q_1p} + \\
& (1 - q_{t_{max}-1})(1-p) + \mathcal{Q}_2.
\end{aligned}$$

Then, we can prove the following theorem.

**Theorem 17.** *When $f(\Delta) = \alpha\Delta + \beta$ and under Condition 2, the threshold preemptive policy is optimal if the transmission delay follows the Zipf distribution.*

*Proof.* We follow the same methodology presented in the proof of Theorem 15. The complete proof can be found in Appendix D.7. □

**Remark 23.** *For the system that fails to satisfy Condition 2, we can use the RVI algorithm introduced in Section 5.5.2 to approximate the corresponding optimal policy.*

**Corollary 8.** *The following results can be derived from Theorem 17.*

1. *When $f(\Delta) = \alpha\Delta + \beta$ and under Condition 2, the threshold preemptive policy is optimal.*
2. *For a generic transmission delay with transmission time upper bounded by 2 time slots, the threshold preemptive policy is optimal.*

*Proof.* We note that in the proof of Theorem 17, we only use Condition 2 and the fact that $q_t \geq 0$. Therefore, the first result can be directly derived from the proof of Theorem 17. For the second result, since the transmission time is upper bounded by 2, the proof follows the same steps as detailed in the proof of Theorem 17 with $t_{max} = 2$. The difference is that only the first three structural properties in Lemma 18 hold. Nevertheless, we can still complete the proof because the case that needs to use the fourth structural property in Lemma 18 does not exist in the case of $t_{max} = 2$. For the same reason, we also do not need to verify Condition 2. Consequently, we omit the detailed proof. □

Table 5.1: Condition 2 Verification

| $a$ | 0 | 0.25 | 0, 5 | 0.75 | 1 | 1.25 | 1.5 |
|---|---|---|---|---|---|---|---|
| Result | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| $a$ | 1.75 | 2 | 2.25 | 2.5 | 2.75 | 3 | 3.25 |
| Result | ✗ | ✗ | ○ | ✓ | ✓ | ✓ | ✓ |
| $a$ | 3.5 | 3.75 | 4 | 4.25 | 4.5 | 4.75 | 5 |
| Result | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

## 5.6 Numerical Results

In this section, we present numerical results regarding the verification of Condition 2 as well as a performance analysis of the optimal policy and the performance improvement compared to the non-preemptive policy.

### 5.6.1 Verification of Condition 2

In this subsection, we numerically verify Condition 2 under various system parameters. More specifically, the system parameters are chosen as follows.

- $0.05 \leq p \leq 0.45$ with an increment of $0.05$.
- $0 \leq a \leq 5$ with an increment of $0.25$.
- $3 \leq t_{max} \leq 11$ with an increment of $1$.

We choose $f(\Delta) = \Delta$ for better illustration. The results are summarized in Table 5.1, where the cross means that Condition 2 is not satisfied, the check mark means that Condition 2 is satisfied, and the circle means the result depends on the specific parameters. When $a = 2.25$, the results are visualized in Figure 5.2. We emphasize here that the optimal policy depends not only on the

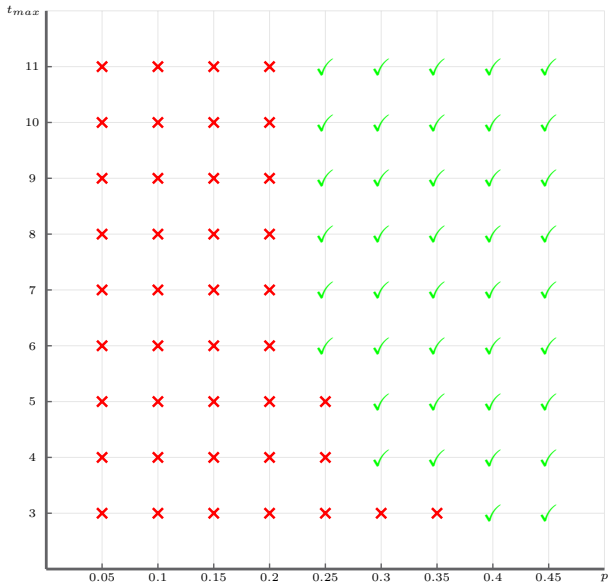



Figure 5.2: A visual representation of the results of numerical check of Condition 2 when the transmission delay follows the Zipf distribution with $a = 2.25$ under different $t_{max}$ and $p$. In the figure, the check mark indicates that Condition 2 is verified, and the cross indicates that Condition 2 is not verified.

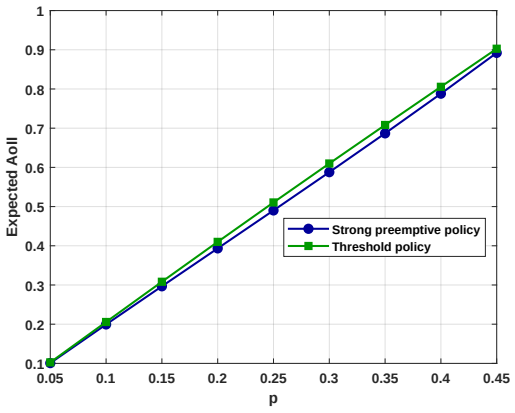
type of probability distribution of the delay but also on the probability distribution parameters.

### 5.6.2 Performance of the Optimal Policy

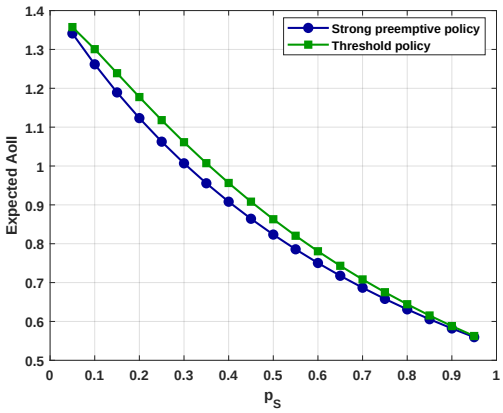
In this subsection, we compare the performance of the optimal policy with non-preemptive policy to highlight the performance improvements brought about by the preemption capability. To this end, we first define a specific type of non-preemptive policy.

**Definition 13** (Threshold policy)**.** *The threshold policy starts a new transmission when the channel is idle, and the AoII is not zero. When the channel is busy, the threshold policy never preempts the transmitting update.*

We choose $f(\Delta) = \Delta$ for better illustrations. Note that the threshold policy is a special case of the threshold policy defined in [77, Definition 2]. Then, we can compute the performances of the optimal policy and the threshold policy using Corollary 7 and [77, Theorem 3], respectively.




(a) $p_s = 0.7$.

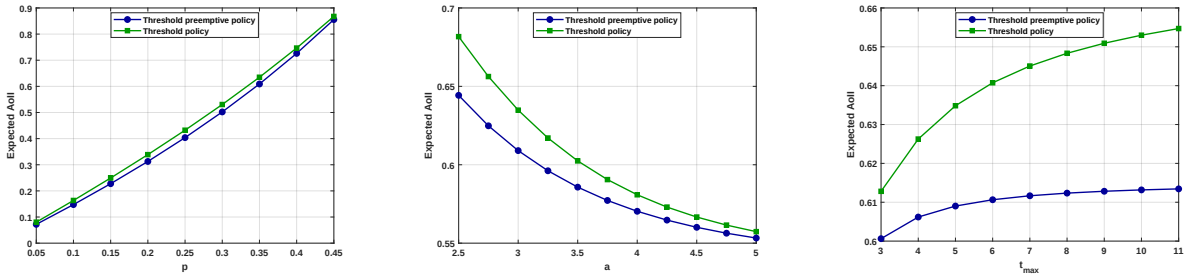



(b) $p = 0.35$.

Figure 5.3: The performance comparison when the transmission delay follows the Geometric distribution. In this case, there are two system parameters. One is the Markovian source dynamics $p$, and the other is the success probability $p_s$ in the Geometric distribution. Therefore, we fix one of the parameters and plot the corresponding results when the other parameter varies.

To accommodate assumption 1 in [77], we set the upper bound on the transmission time to $40$ when calculating the performance of the threshold policy. We also choose the system parameters that have been verified in [77] to satisfy [77, Condition 1]. Consequently, the threshold policy is optimal when the transmitter has no preemption capability. Then, we plot the corresponding performances for the two typical transmission delay models studied in this chapter.

When the transmission delay follows the Geometric distribution, the numerical results are given in Figure 5.3. The plots show that the performance gain from the transmitter's preemption capability is not significant. One possible reason is that the source process modeling and the time penalty function choice in this chapter are simple. Meanwhile, the expected AoII achieved by the optimal policy increases as $p$ increases. This is because when $p$ is large, the Markovian source jumps between states more frequently, making it more difficult for the receiver to maintain a correct estimate about the state of the Markovian source. On the contrary, the expected AoII resulting from the optimal policy decreases as $p_s$ increases. The reason for this is as follows.




(a) $a = 3$ and $t_{max} = 5$. (b) $p = 0.35$ and $t_{max} = 5$. (c) $p = 0.35$ and $a = 3$.

Figure 5.4: The performance comparison when the transmission delay follows the Zipf distribution. In this case, there are three system parameters: the Markovian source dynamics $p$, the constant $a$ in the Zipf distribution, and the upper bound on the transmission time $t_{max}$. We fix two of these parameters in the calculations, then vary the remaining one and plot the corresponding results.

When $p_s$ is large, the expected transmission time of an update is small. As a result, the receiver receives more updates per unit of time, which allows a more accurate estimation of the state of the Markovian source.

When the transmission delay follows the Zipf distribution, the numerical results are given in Figure 5.4. Again, the performance gain from the preemption capability is not significant. The expected AoII achieved by the optimal policy grows with $p$ for the same reason as in the case of the Geometric distribution. As $a$ decreases and $t_{max}$ increases, the expected transmission time of an update increases, which leads to a decrease in the number of updates received by the receiver per unit of time. Therefore, the expected AoII increases.

## 5.7 Conclusion

In this chapter, we optimize the performance of a transmitter-receiver pair in a system using the AoII with a generic time penalty function as the performance metric. In the system, the transmitter decides when to transmit status updates about a Markovian source to a distant receiver

over a channel with a random delay to achieve the minimum expected AoII. The transmitter we consider can preempt the transmitting update to transmit a new update when the channel is busy, and the receiver will predict the state of the Markovian source based on the received update. First, we cast the optimization problem into an infinite horizon with average cost MDP and provide the analytical expressions of the expected AoII achieved by the strong preemptive policy. Then, we prove the existence of the optimal policy and introduce the RVI algorithm to find the optimal policy. To implement the RVI algorithm, we truncate the MDP so that its state space becomes finite. However, the optimal policy resulting from the RVI algorithm is only an approximation. Therefore, we perform a theoretical analysis of the system when the delay distribution follows the Geometric and Zipf distributions, respectively. To this end, we introduce the policy iteration algorithm. Then, leveraging the policy improvement theorem, we theoretically find the corresponding optimal policies. For the system considered in this chapter, it is always optimal to transmit new updates when the channel is idle. When the channel is transmitting an old update, whether the transmitter preempts is closely related to whether the transmitting update can bring new information to the receiver and whether the transmitting update carries correct information about the Markovian source. Finally, we present the numerical results on the validation of Condition 2, the performance comparison between the optimal policy and the non-preemptive policy, and the effect of system parameters on the performance.

# Chapter 6: Conclusion and Outlook

## 6.1 Conclusion

In Chapter 2, we study the optimization of AoII when the resources are limited. In this case, the transmitter must choose wisely how to allocate the limited resources to best achieve the objective. We cast the problem into a constrained Markov decision process and prove that the optimal policy is a mixture of two deterministic threshold policies. For the deterministic threshold policy, we precisely calculate the expected AoII and the expected transmission ratio it achieves. This allows us to develop a low-complexity iterative algorithm to find the optimal policy and the mixing coefficient.

In Chapter 3, we study the performance and optimization of AoII in the scheduling problem. In this problem, the transmitter needs to choose a subset of the users to update each time, and should carefully choose which ones to update so that all users can achieve the best possible performance. We first investigate the structural properties of the optimal policy. Then, we develop the Whittle's index policy. However, the indexability requirement limits its application. Hence, we develop a new scheduling policy called the indexed priority policy, based on the optimal policy for the relaxed problem and the prime-dual heuristic. Although the indexed priority policy is more computationally expensive than Whittle's index policy, it has a broader applicability.

In Chapter 4 and Chapter 5, we investigate the characteristics of AoII when the channel

suffers from a random delay. First, in Chapter 4, we consider the case where the transmitter cannot do anything when the channel is busy transmitting a previous update. In this case, we precisely calculate the performance of the threshold policy and theoretically prove that, for a generic transmission delay and under an easily verifiable condition, the optimal policy for the transmitter is to start a new transmission whenever the AoII is not zero, and the channel is idle. In Chapter 5, we go one step further by considering the case where the transmitter can preempt the transmitting update to transmit a newer one. In this case, we precisely calculate the performance of the strong preemptive policy. Furthermore, we theoretically prove that the optimal policy is the strong preemptive policy when the transmission delay follows the Geometric distribution or the Zipf distribution satisfying certain characteristics.

## 6.2 Outlook

Although the previous chapters have demonstrated the characteristics, performance, and optimizations of AoII under various system settings, many problems and research directions remain to be explored.

- *Generalization*: The methodologies presented in the previous chapters have the potential to be applied to other problems. Therefore, we are interested in whether and how these methodologies can be generalized or applied to other related problems.

- *Dynamic source modeling*: In previous chapters and existing literature, the dynamic source is typically modeled by a Markov chain and possesses a symmetric structure. Meanwhile, the statistical parameters are clear. However, this is often not true in real-life applications. Hence, we seek to extend our efforts to the case where the modeling of the dynamic source

is more sophisticated and closer to the physical world.

- *Feedback mechanism*: One feature that distinguishes AoII from AoI is that it cares about the content of the information being transmitted. Therefore, we often assume that the feedback is timely and accurate. However, such an assumption is rare in practice. Therefore, we would like to understand how AoII behaves when the feedback is delayed or corrupted during the transmission.

- *Continuous time domain*: Current research on AoII is still limited to discrete-time systems. Therefore, the characteristics of AoII in continuous-time systems remain to be discovered. At the same time, the mathematical tools for solving problems in continuous-time systems are quite different from those in discrete-time systems. Therefore, the study of AoII in continuous-time systems is challenging but very important and rewarding.

- *Applications*: The ultimate goal of theoretical research is to apply the results to practical applications. Therefore, we want to combine the existing theoretical results with practical applications. For example, we are interested in analyzing the performance of AoII under canonical network protocols, developing AoII-based network protocols, and combining various learning techniques in artificial intelligence with AoII analysis and optimization.

# Appendix A: Freshness under Limited Resources

## A.1 Proof of Lemma 1

To better distinguish between different states of the system, we denote by $V_\nu(x_d, x_\Delta)$ the estimated value function of state $x$ at iteration $\nu$. To show the desired results, it is sufficient to prove that, at any iteration $\nu > 0$, the following holds

$$V_\nu(d, \Delta_1) > V_\nu(d, \Delta_2) \quad \forall\, \Delta_1 > \Delta_2 \geq 0, \tag{A.1}$$

$$V_\nu(d_1, \Delta) > V_\nu(d_2, \Delta) \quad \forall\, N - 1 \geq d_1 > d_2 \geq 0. \tag{A.2}$$

Leveraging the iterative nature of RVI, we use induction to prove the desired results. Without loss of generality, we choose $x = (0, 0)$ as the reference state. Since we initialize $V_0(d, \Delta) = \Delta$, equations (A.1) and (A.2) hold when $\nu = 0$. We suppose it holds up till iteration $\nu = t$ and examine whether it still holds at iteration $\nu = t + 1$.

We first notice that the transition probabilities which dictate the structure of Bellman update depend only on $d$. Combining with the monotonic property of $V_\nu(\cdot)$, we conclude that (A.1) holds at iteration $\nu + 1$.

We next show the relationship between $V_{\nu+1}(d_1, \Delta)$ and $V_{\nu+1}(d_2, \Delta)$. To this end, we define $V^0_{\nu+1}(\cdot)$ and $V^1_{\nu+1}(\cdot)$ as the estimated value function if action $a = 0$ and $a = 1$ is chosen,

respectively. Hence, we can combine and rewrite the Bellman update reported in (2.6) and (2.7) as follows.

$$V_{\nu+1}(d,\Delta) = \min\left\{V^0_{\nu+1}(d,\Delta), V^1_{\nu+1}(d,\Delta)\right\}, \tag{A.3}$$

where $V^a_{\nu+1}(d,\Delta)$ is calculated by

$$V^a_{\nu+1}(d,\Delta) = \Delta + \lambda a + (1-p_s a)\sum_{d'=0}^{N-1} P_{d,d'}V_\nu(d',\Delta') + ap_s\big((1-2p)V_\nu(0,0) + 2pV_\nu(1,1)\big) - Q_{\nu+1}(x^{ref}), \tag{A.4}$$

where $\Delta' = \mathbb{1}_{\{d'\neq 0\}} \times (\Delta + d')$ and $P_{d,d'}$ is specified in (2.1). With this in mind, we divide our discussion into the following cases.

- $d_1 = 1$ *and* $d_2 = 0$: We know that $\Delta = 0$ if and only if $d = 0$. Then, we only need to compare $V_{\nu+1}(1,\Delta)$ with $V_{\nu+1}(0,0)$. Applying (2.1) to (A.4), we arrive at the following results.

$$V^0_{\nu+1}(1,\Delta) - V^0_{\nu+1}(0,0) = \Delta + \kappa_1,$$

$$V^1_{\nu+1}(1,\Delta) - V^1_{\nu+1}(0,0) = \Delta + p_f\kappa_1,$$

  where $\Delta > 0$ and

$$\kappa_1 = (1-3p)[V_\nu(1,\Delta+1) - V_\nu(0,0)] + p[V_\nu(2,\Delta+2) - V_\nu(1,1)] + p[V_\nu(1,\Delta+1) - V_\nu(1,1)].$$

- $d_1 = 2$ *and* $d_2 = 1$: We need to compare $V_{\nu+1}(2,\Delta)$ with $V_{\nu+1}(1,\Delta)$. Following the same

trajectory, we have the following.

$$V^0_{\nu+1}(2,\Delta) - V^0_{\nu+1}(1,\Delta) = \kappa_2,$$

$$V^1_{\nu+1}(2,\Delta) - V^1_{\nu+1}(1,\Delta) = p_f\kappa_2,$$

where

$$\kappa_2 = p[V_\nu(1,\Delta+1) - V_\nu(0,0)] + p[V_\nu(3,\Delta+3) - V_\nu(2,\Delta+2)] + (1-2p)[V_\nu(2,\Delta+2) - V_\nu(1,\Delta+1)].$$

- $2 \leq d_2 < d_1 \leq N-2$: We need to compare $V_{\nu+1}(d_1,\Delta)$ with $V_{\nu+1}(d_2,\Delta)$. Following again the same trajectory, we have the following.

$$V^0_{\nu+1}(d_1,\Delta) - V^0_{\nu+1}(d_2,\Delta) = \kappa_3,$$

$$V^1_{\nu+1}(d_1,\Delta) - V^1_{\nu+1}(d_2,\Delta) = p_f\kappa_3,$$

where

$$\kappa_3 = (1-2p)[V_\nu(d_1,\Delta+d_1) - V_\nu(d_2,\Delta+d_2)] + p[V_\nu(d_1-1,\Delta+d_1-1) - V_\nu(d_2-1,\Delta+d_2-1)] + p[V_\nu(d_1+1,\Delta+d_1+1) - V_\nu(d_2+1,\Delta+d_2+1)].$$

- $d_1 = N-1$ *and* $d_2 = N-2$: We need to compare $V_{\nu+1}(N-1,\Delta)$ with $V_{\nu+1}(N-2,\Delta)$. Following again the same trajectory, we have the following.

$$V^0_{\nu+1}(N-1,\Delta) - V^0_{\nu+1}(N-2,\Delta) = \kappa_4,$$

$$V^1_{\nu+1}(N-1,\Delta) - V^1_{\nu+1}(N-2,\Delta) = p_f\kappa_4,$$

where

$$\kappa_4 = p[V_\nu(N-2,\Delta+N-2) - V_\nu(N-3,\Delta+N-3)] + (1-3p)[V_\nu(N-1,\Delta+N-1) - V_\nu(N-2,\Delta+N-2)].$$

Baring in mind the monotonicity of $V_\nu(d,\Delta)$ and $p \in [0,\frac{1}{3}]$, we can easily see that $\kappa_1$, $\kappa_2$, $\kappa_3$, and $\kappa_4$ are all positive. Since the estimated value function is updated following (A.3), we can easily verify that (A.2) holds at iteration $t+1$ which concludes our proof.

## A.2 Proof of Proposition 1

We continue with the same notations as in the proof of Lemma 1. We recall that RVI is an iterative algorithm and the estimated value function will converge to the value function. Hence, it is sufficient to show that the properties hold for the optimal policy at any iteration of RVI.

We define $\delta V_\nu(d,\Delta) = V^1_\nu(d,\Delta) - V^0_\nu(d,\Delta)$. Without loss of generality, we assume $t > 0$. Then, the optimal action at iteration $\nu$ is captured by the sign of $\delta V_\nu(d,\Delta)$. More precisely, the optimal action $a^*_t = 1$ if $\delta V_\nu(d,\Delta) \leq 0$ and $a^*_t = 0$ otherwise. Then, we can prove the following

lemma.

**Lemma 13.** *$\delta V_\nu(d, \Delta)$ is decreasing in $\Delta$ when $d \neq 0$ and $t > 0$.*

*Proof.* We distinguish between following cases.

- When $d = 1$, applying (2.1) to (A.4) yields

$$\delta V_\nu(1, \Delta) = \lambda + p_s\{p[V_{\nu-1}(1, 1) - V_{\nu-1}(1, \Delta + 1)] + (1 - 3p)[V_{\nu-1}(0, 0) - V_{\nu-1}(1, \Delta + 1)] + p[V_{\nu-1}(1, 1) - V_{\nu-1}(2, \Delta + 2)]\}. \quad \text{(A.5)}$$

  We notice that $(1 - 3p)$ is non-negative as $p \in [0, \frac{1}{3}]$.

- When $2 \leq d \leq N - 2$, following the same trajectory, we have

$$\delta V_\nu(d, \Delta) = \lambda + p_s\{(1 - 2p)[V_{\nu-1}(0, 0) - V_{\nu-1}(d, \Delta + d)] + p[V_{\nu-1}(1, 1) - V_{\nu-1}(d - 1, \Delta + d - 1)] + p[V_{\nu-1}(1, 1) - V_{\nu-1}(d + 1, \Delta + d + 1)]\}. \quad \text{(A.6)}$$

- When $d = N - 1$, following again the same trajectory, we have

$$\delta V_\nu(N - 1, \Delta) = \lambda + p_s\{(1 - 2p)[V_{\nu-1}(0, 0) - V_{\nu-1}(N - 1, \Delta + N - 1)] + 2p[V_{\nu-1}(1, 1) - V_{\nu-1}(N - 2, \Delta + N - 2)]\}. \quad \text{(A.7)}$$

We recall that $\lambda$ is a non-negative constant and $V_{\nu-1}(d, \Delta)$ is increasing in both $d$ and $\Delta$ by

Lemma 1. Then, we can see that (A.5), (A.6), and (A.7) are nothing but the sum of a constant and a negative term that is decreasing in $\Delta$. Combing together, we can conclude our proof. □

With the lemma given, we can see that, for fixed $d \neq 0$, $\delta V_\nu(d, \Delta)$ will decrease as $\Delta$ increases and, at some point, it will become negative. Therefore, for the states with fixed $d \neq 0$, the optimal action $a_t^*$ will switch from $a_t^* = 0$ to $a_t^* = 1$ as $\Delta$ increases. [1] We define the switching point for each $d \neq 0$ as the first $\Delta$ such that $\delta V_\nu(d, \Delta)$ is non-positive. Since the instant cost is unbounded, the value function must also be unbounded. Therefore, the switching points always exist. We notice that the expressions of $\delta V_\nu(d, \Delta)$ differ for different $d$. Consequently, the corresponding switching points will also be different. To investigate the relationships between the switching points, we provide the following lemma.

**Lemma 14.** *$\delta V_\nu(d, \Delta)$ is decreasing in $d$ when $d \neq 0$ and $\nu > 0$.*

*Proof.* It is equivalent to show that $\forall\ \Delta > 0$, $\delta V_\nu(d_1, \Delta) > \delta V_\nu(d_2, \Delta)$ if $1 \leq d_1 < d_2 \leq N - 1$. To this end, we distinguish between the following cases.

- When $d_1 = 1$ and $d_2 = 2$, leveraging (A.5) and (A.6), we have

$$\begin{aligned}\delta V_\nu(1, \Delta) - \delta V_\nu(2, \Delta) = p_s\{&(1 - 2p)[V_{\nu-1}(2, \Delta + 2) - V_{\nu-1}(1, \Delta + 1)] + \\ &p[V_{\nu-1}(3, \Delta + 3) - V_{\nu-1}(2, \Delta + 2)] + \\ &p[V_{\nu-1}(1, \Delta + 1) - V_{\nu-1}(0, 0)]\}. \quad \text{(A.8)}\end{aligned}$$

[1] It is worth noting that $\delta V_\nu(d, \Delta)$ can always be negative which means that the optimal action $a_t^*$ will always be $a_t^* = 1$.

- When $2 \leq d_1 < d_2 \leq N-2$, leveraging (A.6), we have

$$\begin{aligned}\delta V_\nu(d_1,\Delta) - \delta V_\nu(d_2,\Delta) = p_s\{&(1-2p)[V_{\nu-1}(d_2,\Delta+d_2) - V_{\nu-1}(d_1,\Delta+d_1)]+\\ &p[V_{\nu-1}(d_2-1,\Delta+d_2-1) - V_{\nu-1}(d_1-1,\Delta+d_1-1)]+\\ &p[V_{\nu-1}(d_2+1,\Delta+d_2+1) - V_{\nu-1}(d_1+1,\Delta+d_1+1)]\}. \quad \text{(A.9)}\end{aligned}$$

- Similarly, when $d_1 = N-2$ and $d_2 = N-1$, we have

$$\begin{aligned}&\delta V_\nu(N-2,\Delta) - \delta V_\nu(N-1,\Delta) =\\ &\quad p_s\{(1-3p)[V_{\nu-1}(N-1,\Delta+N-1) - V_{\nu-1}(N-2,\Delta+N-2)]+\\ &\quad\quad p[V_{\nu-1}(N-2,\Delta+N-2) - V_{\nu-1}(N-3,\Delta+N-3)]\}. \quad \text{(A.10)}\end{aligned}$$

According to Lemma 1, $V_{\nu-1}(d,\Delta)$ is increasing in both $d$ and $\Delta$. Combining with the fact that $p \in [0, \frac{1}{3}]$, we can easily verify that (A.8), (A.9), and (A.10) are all positive. Consequently, $\delta V_\nu(d_1,\Delta) > \delta V_\nu(d_2,\Delta)$ holds $\forall\ 1 \leq d_1 < d_2 \leq N-1$ which concludes our proof. □

Let $n_d^t$ denotes the switching point for the states with $d \neq 0$ at iteration $t > 0$. Then, $\delta V_\nu(d, n_d^t - 1) > 0$ and $\delta V_\nu(d, n_d^t) \leq 0$. Since $\delta V_\nu(d,\Delta)$ is decreasing in $d$, $\delta V_\nu(d', n_d^t) < \delta V_\nu(d, n_d^t) \leq 0$ if $d' > d$. This indicates that the ordering $n_{d'}^t \leq n_d^t$ must hold. Thus, we can conclude that the switching points $n_d^t$ when $d \neq 0$ are non-increasing in $d$.

Finally, we discuss the only missing case: $d = 0$. We know that $\Delta = 0$ if and only if $d = 0$. Thus, we only need to consider the state $(0,0)$. Then, we apply (2.1) to (A.4) which yields $\delta V_\nu(0,0) = \lambda$ for any $t > 0$. As $\lambda$ is a non-negative constant, we can conclude that the

optimal action at state $(0,0)$ is $a_t^* = 0$.

As the above results are valid for any $\nu > 0$ and RVI converges to the value function as $\nu \to +\infty$ (i.e. $\lim_{\nu \to +\infty} V_\nu(\cdot) = V(\cdot)$), we can conclude that the above results are valid for the optimal policy for (2.4).

## A.3 Proof of Theorem 1

We first introduce the infinite horizon $\gamma$-discounted cost of $\mathcal{M}$ where $0 < \gamma < 1$ is a discount factor. The expected $\gamma$-discounted cost under policy $\pi$ starting from state $x$ can be calculated as

$$V_{\pi,\gamma}(x) = \mathbb{E}_\pi \left[ \sum_{t=0}^{\infty} \gamma^t C(x_t, a_t) \mid x \right].$$

The quantity $V_\gamma(\cdot) \triangleq inf_\pi V_{\pi,\gamma}(\cdot)$ is the best that can be achieved. Equivalently, $V_\gamma(\cdot)$ is the value function associated with the infinite horizon $\gamma$-discounted MDP. Then $V_\gamma(\cdot)$ satisfies the following Bellman equation.

$$V_\gamma(x) = \min_a \left\{ C(x,a) + \gamma \sum_{x' \in \mathcal{X}} P_{xx'}(a) V_\gamma(x') \right\}.$$

We further define the quantity $v_{\gamma,n}(\cdot)$ as the minimum expected discounted cost for operating the system from time $t = 0$ to $t = n-1$. It is known that $lim_{n\to\infty} v_{\gamma,n}(x) = V_\gamma(x)$, for all $x \in \mathcal{X}$. We also define the expected cost under policy $\pi$ starting from state $x$ as

$$J_\pi(x) = \limsup_{n \to \infty} \frac{1}{n} \mathbb{E}_\pi \left[ \sum_{t=0}^{n-1} C(x_t, a_t) \mid x \right],$$

and $J(\cdot) \triangleq \inf_\pi J_\pi(\cdot)$ is the best that can be achieved. $V_{\pi,\gamma}^{(m)}(\cdot)$, $V_\gamma^{(m)}(\cdot)$, $v_{\gamma,n}^{(m)}(\cdot)$, $J_\pi^{(m)}(\cdot)$ and $J^{(m)}(\cdot)$ are defined analogously for the truncated MDP $\mathcal{X}^{(m)}$. We define $h_\gamma^{(m)}(x) \triangleq V_\gamma^{(m)}(x) - V_\gamma^{(m)}(0)$ as the relative value function and chose the reference state $0 = (0, 0)$. For the simplicity of notation, for any two state $x, y \in \mathcal{X}$, we say $x \leq y$ if and only if $x_d \leq y_d$ and $x_\Delta \leq y_\Delta$.

With the above definitions in mind, we claim that our system verifies the two assumptions given in [52]. That is

- *Assumption 1*: There exists a non-negative (finite) constant $L$, a non-negative (finite) function $M(\cdot)$ on $\mathcal{X}$, and constants $m_0$ and $\gamma_0 \in [0, 1)$, such that $-L \leq h_\gamma^{(m)}(x) \leq M(x)$, for $x \in \mathcal{X}^{(m)}$, $m \geq m_0$, and $\gamma \in (\gamma_0, 1)$: We recall that $V_\gamma^{(m)}(x)$ is the value function and satisfies the Bellman equation. Thus, we can show that $V_\gamma^{(m)}(x)$ is increasing in $x$ in a similar way as did in Lemma 1. The proof is omitted for the sake of space. Then, $h_\gamma^{(m)}(x) = V_\gamma^{(m)}(x) - V_\gamma^{(m)}(0) \geq 0$. Consequently, we can choose $L = 0$.

  Let $c_{x,0}(\psi)$ be the expected cost of a first passage from $x \in \mathcal{X}$ to the reference state 0 when policy $\psi$ is adopted and $c_{x,0}^{(m)}(\psi)$ is defined analogously for the truncated MDP $\mathcal{X}^{(m)}$. In the following, we consider the policy $\psi$ being always update policy where the transmitter makes transmission attempt at each time slot. Since the policy $\psi$ induces an irreducible ergodic Markov chain and the expected cost is finite, $h_\gamma^{(m)}(x) \leq c_{x,0}^{(m)}(\psi)$ from [80, Proposition 5] and $c_{x,0}(\psi)$ is finite from [80, Proposition 4]. We also know that $c_{x,0}(\psi)$ satisfies the following equation [52].

$$c_{x,0}(\psi) = C(x, a^\psi) + \sum_{x' \in \mathcal{X} - \{0\}} P_{xx'}(a^\psi) c_{x',0}(\psi). \tag{A.11}$$

  We can verify in a similar way to the proof of Lemma 1 that $c_{x,0}(\psi)$ is increasing in $x$. The

proof is omitted here for the sake of space. Then, we obtain

$$\begin{aligned}
&\sum_{y\in\mathcal{X}_{-1}^{(m)}} P_{xy}^{(m)}(a^{\psi})c_{y,0}(\psi) \\
&\qquad = \sum_{y\in\mathcal{X}_{-1}^{(m)}} P_{xy}(a^{\psi})c_{y,0}(\psi) + \sum_{y\in\mathcal{X}_{-1}^{(m)}} \left( \sum_{z\notin\mathcal{X}^{(m)}} P_{xz}(a^{\psi})q_z(y) \right) c_{y,0}(\psi) \\
&\qquad = \sum_{y\in\mathcal{X}_{-1}^{(m)}} P_{xy}(a^{\psi})c_{y,0}(\psi) + \sum_{z\notin\mathcal{X}^{(m)}} P_{xz}(a^{\psi}) \left( \sum_{y\in\mathcal{X}_{-1}^{(m)}} q_z(y)c_{y,0}(\psi) \right) \qquad \text{(A.12)} \\
&\qquad \leq \sum_{y\in\mathcal{X}_{-1}^{(m)}} P_{xy}(a^{\psi})c_{y,0}(\psi) + \sum_{z\notin\mathcal{X}^{(m)}} P_{xz}(a^{\psi})c_{z,0}(\psi) \\
&\qquad = \sum_{y\in\mathcal{X}-\{0\}} P_{xy}(a^{\psi})c_{y,0}(\psi),
\end{aligned}$$

where $\mathcal{X}_{-1}^{(m)} = \mathcal{X}^{(m)} - \{0\}$. Applying (A.12) to (A.11) yields

$$c_{x,0}(\psi) \geq C(x, a^{\psi}) + \sum_{y\in\mathcal{X}^{(m)}-\{0\}} P_{xy}^{(m)}(a^{\psi})c_{y,0}(\psi).$$

Bearing in mind that $c_{x,0}^{(m)}(\psi)$ satisfies the following.

$$c_{x,0}^{(m)}(\psi) = C(x, a^{\psi}) + \sum_{y\in\mathcal{X}^{(m)}-\{0\}} P_{xy}^{(m)}(a^{\psi})c_{y,0}^{(m)}(\psi),$$

we can conclude that $c_{x,0}^{(m)}(\psi) \leq c_{x,0}(\psi)$. Thus, we can choose $M(x) = c_{x,0}(\psi) < \infty$.

- *Assumption 2*: $\limsup_{m\to\infty} J^{(m)} \triangleq J^* < \infty$ and $J^* \leq J(x)$, for all $x \in \mathcal{X}$: We first show the hypothesis in [52, Proposition 5.1] is true. Since we redistribute the excess probabilities

in a way such that, for all $z \in \mathcal{X} - \mathcal{X}^{(m)}$,

$$\sum_{y \in \mathcal{X}^{(m)}} q_z(y) v_{\gamma,n}(y) = v_{\gamma,n}(x),$$

where $x_d = z_d$ and $x_\Delta = m$, we only need to verify that, for all $z \in \mathcal{X} - \mathcal{X}^{(m)}$

$$v_{\gamma,n}(x) \leq v_{\gamma,n}(z). \tag{A.13}$$

As $v_{\gamma,n}(x)$ adopts the following inductive form [52].

$$v_{\gamma,n+1}(x) = \min_a \left\{ C(x,a) + \gamma \sum_{x' \in \mathcal{X}} P_{xx'}(a) v_{\gamma,n}(x') \right\},$$

we can prove (A.13) is true in a similar way to Lemma 1 and the proof is omitted for the sake of space. $J(x)$ is trivially finite for $x \in \mathcal{X}$. Then, according to [52, Corollary 5.2], assumption 2 is valid.

Consequently, the following results are true.

- There exists an average cost optimal stationary policy in $\mathcal{M}^{(m)}$.
- Any limit point of the sequence of optimal policies in $\mathcal{M}^{(m)}$ is optimal in $\mathcal{M}$.

## A.4 Proof of Proposition 2

We first delve into the state space $\mathcal{X}$ of the MDP $\mathcal{M}$ and provide the condition it must satisfy. Without loss of generality, we suppose the system always starts from state $(0,0)$. We

claim that state $(d, \Delta)$ with $d \neq 0$ must satisfy the following condition.

$$\Delta \geq l_d = \sum_{i=1}^{d} i = \frac{d^2 + d}{2}. \tag{A.14}$$

To see the condition, we notice that the transition to state $(0, 0)$ is equivalent to restarting the system. Thus, it is sufficient to consider the sequence of transitions starting from the last time the system is at state $(0, 0)$. Therefore, the age $\Delta$ will always increase. We recall that the maximum jump of $d$ is 1 as specified in (2.1). Thus, we can conclude that there always exists a lower bound $l_d$ on the age $\Delta$ for any given $d \neq 0$. Combing with the system dynamic discussed in Section 2.2.2, the lower bound in (A.14) is easy to obtain. To make the structure of equations consistent, we define the states that violate the condition (A.14) as *virtual* states since the system will never reach these states.

As the state space is clarified, we proceed with deriving the main results. We first recall that the threshold policy $\boldsymbol{n}$ possesses the properties detailed in Proposition 1. More precisely, for the state with given $d \neq 0$, the action suggested by $\boldsymbol{n}$ is $a^* = 1$ if the age $\Delta$ is larger than or equal to the corresponding threshold $n_d$. Hence, we define $\tau = max\{\boldsymbol{n}\}$. An important property of $\tau$ is that, for the states with $\Delta \geq \tau$, the actions suggested by $\boldsymbol{n}$ are the same. We define $a_{d,\Delta}$ as the action suggested by $\boldsymbol{n}$ at state $(d, \Delta)$. For each $1 \leq d \leq N - 1$, we define

$$\Pi_d(\tau) = \sum_{\Delta=\tau}^{+\infty} a_{d,\Delta} \pi_d(\Delta) = \sum_{\Delta=\tau}^{+\infty} \pi_d(\Delta).$$

The last equality holds since $a_{d,\Delta} = 1$ for all the states with $\Delta \geq \tau$. Then, the expected

transmission rate can be calculated as

$$\bar{R}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \sum_{\Delta=n_d}^{+\infty} \pi_d(\Delta) = \sum_{d=1}^{N-1} \left( \sum_{\Delta=n_d}^{\tau-1} \pi_d(\Delta) + \Pi_d(\tau) \right).$$

We claim that $\Pi_d(\tau)$'s, along with the stationary distribution $\pi_d(\Delta)$'s, can be obtained by solving a finite system of linear equations induced from the balance equation (2.8). Leveraging the results in Section 2.2.2, we distinguish between the following cases.

- For state $(0,0)$, (2.8) can be written as

$$\begin{aligned}
\pi_0(0) &= (1-2p)\pi_0(0) + p\sum_{\Delta=1}^{+\infty}(1-p_s a_{1,\Delta})\pi_1(\Delta) + p_s(1-2p)\sum_{d=1}^{N-1}\sum_{\Delta=l_d}^{+\infty} a_{d,\Delta}\pi_d(\Delta) \\
&= (1-2p)\pi_0(0) + p\sum_{\Delta=1}^{\tau-1}(1-p_s a_{1,\Delta})\pi_1(\Delta) + p_f p\Pi_1(\tau) + \\
&\quad p_s(1-2p)\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right).
\end{aligned}$$

  This recovers (2.9).

- For state $(1,1)$, (2.8) can be written as

$$\begin{aligned}
\pi_1(1) &= 2p\pi_0(0) + 2p_s p\sum_{d=1}^{N-1}\sum_{\Delta=l_d}^{+\infty} a_{d,\Delta}\pi_d(\Delta) \\
&= 2p\pi_0(0) + 2p_s p\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right).
\end{aligned}$$

  This recovers (2.10).

- For the virtual states, we define the steady-state probabilities as zero since the system will never reach these states. This recovers (2.12).

- For other states, leveraging the definition of virtual states, we obtain an alternative form of (2.8) which is

$$\pi_d(\Delta) = \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d})\pi_{d'}(\Delta - d). \tag{A.15}$$

As $\Delta \in \mathbb{N}^*$, there are infinitely many equations to solve. Inspired by the definition of $\Pi_d(\tau)$, we can combine the states with $\Delta \geq \tau$ and eliminate the infinity. More precisely, for each $1 \leq d \leq N-1$, we do the following.

$$\begin{aligned}
\sum_{\Delta=\tau}^{+\infty} \pi_d(\Delta) = \Pi_d(\tau) &= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau}^{+\infty} (1 - p_s a_{d',\Delta-d})\pi_{d'}(\Delta - d) \right) \\
&= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau-d}^{\tau-1} (1 - p_s a_{d',\Delta})\pi_{d'}(\Delta) + p_f \sum_{\Delta=\tau}^{+\infty} \pi_{d'}(\Delta) \right) \\
&= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau-d}^{\tau-1} (1 - p_s a_{d',\Delta})\pi_{d'}(\Delta) + p_f \Pi_{d'}(\tau) \right).
\end{aligned} \tag{A.16}$$

Combining (A.15) and (A.16), we recover (2.13) and (2.14).

Equation (2.11) is obtained from the fact that the steady-state probabilities must add up to one.

By combining the states with $\Delta \geq \tau$, we actually cast the induced infinite-state Markov chain to a finite-state Markov chain. Therefore, the expected transmission rate can be calculated theorecially without any approximation.

## A.5 Proof of Corollary 1

We inherit the notations and definitions from the proof of Proposition 2. Before presenting the main results, we first introduce the key approximation used in the derivation of the main results. We note that when the thresholds in $\boldsymbol{n}$ are huge, the expected transmission rate will be

insignificant. More precisely, when the thresholds $n_d$'s are huge,

$$\bar{R}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \left( \sum_{\Delta=n_d}^{\tau-1} \pi_d(\Delta) + \Pi_d(\tau) \right) \approx 0.$$

We apply the above equation to (2.10) and obtain the following key approximation.

$$\pi_1(1) \approx 2p\pi_0(0). \tag{A.17}$$

Then, we claim that, for any state $(d, \Delta)$, the steady-state probability $\pi_d(\Delta)$ can be approximated as

$$\pi_d(\Delta) \approx c^{\boldsymbol{n}}_{d,\Delta}\pi_0(0),$$

where $c^{\boldsymbol{n}}_{d,\Delta}$ is a scalar depends on the policy and the state. To prove this, we first recall that the transitions in the induced Markov chain always go along the increasing direction of $\Delta$ unless it goes back to state $(0, 0)$ or $(1, 1)$. Combining with the approximation made in (A.17), we can see that any $\pi_d(\Delta)$ can be approximated as a multiple of $\pi_0(0)$.

With this in mind, we notice that, for any two threshold policies $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, the suggested actions by the two polices at states with $\Delta < \min\{[\boldsymbol{n}_1, \boldsymbol{n}_2]\}$ are the same (i.e. $a_{d,\Delta} = 0$). We denote by $G(\boldsymbol{n}_1, \boldsymbol{n}_2)$ the set of these states. Then, for state $(d, \Delta) \in G(\boldsymbol{n}_1, \boldsymbol{n}_2)$, regardless of whether $\boldsymbol{n}_1$ or $\boldsymbol{n}_2$ is adopted, the balance equation is the same. Consequently, the corresponding $c^{\boldsymbol{n}}_{d,\Delta}$ is independent of policy. Then, for $(d, \Delta) \in G(\boldsymbol{n}_1, \boldsymbol{n}_2)$,

$$\frac{\pi^1_d(\Delta)}{\pi^2_d(\Delta)} \approx \frac{c_{d,\Delta}\pi^1_0(0)}{c_{d,\Delta}\pi^2_0(0)} = \frac{\pi^1_0(0)}{\pi^2_0(0)}, \tag{A.18}$$

where $\pi_d^1(\Delta)$'s and $\pi_d^2(\Delta)$'s are the stationary distribution when $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ is adopted, respectively. Leveraging (A.18), we can obtain the main results in the corollary.

We first define $\eta = min\{\boldsymbol{n}\} - 1$ and $\Pi_d(\eta) = \sum_{\Delta=1}^{\eta} \pi_d(\Delta)$. Similar to what we did in the proof of Proposition 2, we rewrite the balance equation (2.8) as follows.

- For state $(0,0)$, (2.8) can be written as

$$\begin{aligned}\pi_0(0) &= (1-2p)\pi_0(0) + p\sum_{\Delta=1}^{+\infty}(1-p_s a_{1,\Delta})\pi_1(\Delta) + p_s(1-2p)\sum_{d=1}^{N-1}\sum_{\Delta=l_d}^{+\infty} a_{d,\Delta}\pi_d(\Delta) \\ &= (1-2p)\pi_0(0) + p\sum_{\Delta=\eta+1}^{\tau-1}(1-p_s a_{1,\Delta})\pi_1(\Delta) + p\Pi_1(\eta) + p_f p\Pi_1(\tau) + \\ &\quad p_s(1-2p)\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right).\end{aligned}$$

  This recovers (2.15).

- For state $(1,1)$, (2.8) can be written as

$$\begin{aligned}\pi_1(1) &= 2p\pi_0(0) + 2p_s p\sum_{d=1}^{N-1}\sum_{\Delta=l_d}^{+\infty} a_{d,\Delta}\pi_d(\Delta) \\ &= 2p\pi_0(0) + 2p_s p\sum_{d=1}^{N-1}\left(\sum_{\Delta=n_d}^{\tau-1}\pi_d(\Delta) + \Pi_d(\tau)\right).\end{aligned}$$

  This recovers (2.16).

- For other states, leveraging the definition of virtual states, we obtain an alternative form of (2.8) which is

$$\pi_d(\Delta) = \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d})\pi_{d'}(\Delta - d). \tag{A.19}$$

  Instead of applying (A.19) directly, we combine the states with $\Delta \leq \eta$ to reduce the number

of equations. More precisely, for each $2 \leq d \leq N-1$, we have

$$\begin{aligned}\sum_{\Delta=1}^{\eta+d} \pi_d(\Delta) &= \Pi_d(\eta) + \sum_{\Delta=\eta+1}^{\eta+d} \pi_d(\Delta) \\ &= \sum_{\Delta=1}^{\eta+d} \left( \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d}) \pi_{d'}(\Delta - d) \right) \\ &= \sum_{d'=1}^{N-1} P_{d',d} \Pi_{d'}(\eta).\end{aligned} \tag{A.20}$$

When $d = 1$, due to the particularity of state $(1, 1)$, we have

$$\sum_{\Delta=2}^{\eta+1} \pi_1(\Delta) = \Pi_1(\eta) - \pi_1(1) + \sum_{\Delta=\eta+1}^{\eta+1} \pi_1(\Delta) = \sum_{d'=1}^{N-1} P_{d',1} \Pi_{d'}(\eta). \tag{A.21}$$

We notice that (A.20) or (A.21) involves $\pi_d(\Delta)$'s where $1 \leq d \leq N-1$ and $\eta + 1 \leq \Delta \leq \eta + d$. Under usual circumstances, these steady-state probabilities can be calculated using (A.19). However, $\pi_d(\Delta)$'s where $\Delta \leq \eta$ are required when applying (A.19) and we have no access to them as we combined them together as $\Pi_d(\eta)$. To circumvent this, we use the approximation reported in (A.18). More precisely, for the states with $1 \leq d \leq N-1$ and $\eta + 1 \leq \Delta \leq \eta + d$, we have

$$\begin{aligned}\pi_d(\Delta) &= \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d}) \pi_{d'}(\Delta - d) \\ &\approx \rho \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d}) \sigma_{d'}(\Delta - d),\end{aligned} \tag{A.22}$$

where $\rho \triangleq \frac{\pi_0(0)}{\sigma_0(0)}$ and $\sigma_d(\Delta)$ is the stationary distribution of the Markov chain induced by another policy $\boldsymbol{n}'$ and can be calculated using Proposition 2. In order to utilize the approximation reported

in (A.18), the policy $\boldsymbol{n}'$ must satisfy

$$min\{[\boldsymbol{n}, \boldsymbol{n}']\} > \eta.$$

We recall in Proposition 2, the computational complexity of calculating the stationary distribution of a Markov chain induced by a threshold policy depends on the maximal threshold. To make the calculation of $\sigma_d(\Delta)$ as cheap as possible, we choose $\boldsymbol{n}' = [\eta', ..., \eta']$ where $\eta' = \min\{\boldsymbol{n}\}$. Combining (A.20), (A.21), and (A.22), we recover (2.18) ,(2.19), and (2.20).

For state with $1 \leq d \leq N-1$ and $\eta + d + 1 \leq \Delta \leq \tau - 1$, leveraging the above approximation, we can calculate the steady-state probabilities using (A.19) which recovers (2.21).

Finally, for state with $1 \leq d \leq N-1$ and $\Delta \geq \tau$, we combine them as did in Proposition 2. Then, we can recover (2.22).

Equation (2.17) is obtained form the fact that the sum of all steady-state probabilities must be one.

By combining the states with $\Delta \leq \eta$, we reduce the size of the finite-state Markov chain cast to. Although approximation is used, as the thresholds increase, the approximation in (A.17) will become more and more accurate.

## A.6 Proof of Corollary 2

We still inherit the notations and definitions from the proof of Proposition 2. We first recall that the AoII at state $(d, \Delta)$ is nothing but $\Delta$. Then, similar to what we did in the proof of

Proposition 2, the expected AoII under threshold policy $\boldsymbol{n}$ can be calculated as

$$\bar{\Delta}_{\boldsymbol{n}} = \sum_{d=1}^{N-1} \left( \sum_{\Delta=l_d}^{\tau-1} \omega_d(\Delta) + \Omega_d(\tau) \right),$$

where $\tau = max\{\boldsymbol{n}\}$, $l_d = \frac{d^2+d}{2}$, and

$$\omega_d(\Delta) \triangleq \Delta \pi_d(\Delta),$$
$$\Omega_d(\tau) \triangleq \sum_{\Delta=\tau}^{+\infty} \omega_d(\Delta).$$

Note that $\pi_d(\Delta)$'s are the stationary distribution of the infinite-state Markov chain induced from the same threshold policy $\boldsymbol{n}$. We claim that $\Omega_d(\tau)$'s, along with $\omega_d(\Delta)$'s, can be obtained by solving a finite system of linear equations. To this end, we distinguish between the following cases.

- For the virtual states, we have $\omega_d(\Delta) = 0$ because $\pi_d(\Delta) = 0$ for these state by definition. Meanwhile, $\omega_0(0) = 0$ because no cost is paid for being at state $(0, 0)$. This recovers (2.24).

- For the states with $1 \leq d \leq N-1$ and $l_d \leq \Delta \leq \tau - 1$, we have

$$\omega_d(\Delta) = \Delta \pi_d(\Delta). \tag{A.23}$$

  This recovers (2.25).

- For the states with $1 \leq d \leq N-1$ and $\Delta \geq \tau$, we can use (A.23). But in this case, we need to calculate an infinite number of values. To eliminate the infinity, we notice that $\omega_d(\Delta)$'s can also be calculated by multiplying both sides of (A.15) by $(\Delta - d)$. More precisely, we

have

$$(\Delta-d)\pi_d(\Delta) = \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d})(\Delta - d)\pi_{d'}(\Delta - d).$$

Applying the definition of $\omega_d(\Delta)$, we have

$$\omega_d(\Delta)-d\pi_d(\Delta) = \sum_{d'=1}^{N-1} P_{d',d}(1 - p_s a_{d',\Delta-d})\omega_{d'}(\Delta - d).$$

Like we did in the proof of Proposition 2, we combine the states with $\Delta \geq \tau$ to eliminate the infinity. More precisely, for each $1 \leq d \leq N-1$, we have

$$\begin{aligned}
\sum_{\Delta=\tau}^{+\infty} \Big(\omega_d(\Delta) - d\pi_d(\Delta)\Big) &= \Omega_d(\tau) - d\Pi_d(\tau) \\
&= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau}^{+\infty} (1 - p_s a_{d',\Delta-d})\omega_{d'}(\Delta - d) \right) \\
&= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau-d}^{\tau-1} (1 - p_s a_{d',\Delta})\omega_{d'}(\Delta) + p_f \sum_{\Delta=\tau}^{+\infty} \omega_{d'}(\Delta) \right) \\
&= \sum_{d'=1}^{N-1} P_{d',d} \left( \sum_{\Delta=\tau-d}^{\tau-1} (1 - p_s a_{d',\Delta})\omega_{d'}(\Delta) + p_f \Omega_{d'}(\tau) \right).
\end{aligned}$$

This recovers (2.26).

## A.7 Proof of Theorem 2

We first make the following definitions. When the MDP $\mathcal{M}$ is at state $x$ and action $a$ is chosen, cost $C_1(x,a) = x_\Delta$ and $C_2(x,a) = \lambda a$ are incurred. We define the expected $C_1$-cost and the expected $C_2$-cost under policy $\pi$ as $\bar{C}_1(\pi)$ and $\bar{C}_2(\pi)$, respectively. Let $G$ be a nonempty set

and $\mathcal{R}^*(i, G)$ be the class of policies $\pi$ such that

- $P_\pi(x_n \in G \; for \; some \; n \geq 1 \mid x_0 = i) = 1$ where $x_n$ is the state of $\mathcal{M}$ at time $n$.
- The expected time $m_{iG}(\pi)$ of a first passage from $i$ to $G$ under $\pi$ is finite.
- The expected $C_1$-cost $\bar{C}_1^{i,G}(\pi)$ and the expected $C_2$-cost $\bar{C}_2^{i,G}(\pi)$ of a first passage form $i$ to $G$ under $\pi$ are finite.

With the above definitions clarified, we proceed with presenting the assumptions given in [50] and verifying our system satisfies all the assumptions.

1. For all $w > 0$, the set $G(w) = \{x \mid$ *there exists an action a such that* $C_1(x, a) + C_2(x, a) \leq w\}$ is finite: For our system, we have $C_1(x, a) + C_2(x, a) = x_\Delta + \lambda a \geq x_\Delta$. Then, any state $x$ in $G(w)$ must satisfy $x_\Delta \leq w$. Bearing in mind that $x_\Delta \in \mathbb{N}_0$, we can conclude that, the set $G(w)$ is always finite.

2. There exists a stationary policy $e$ such that the induced Markov chain has the following properties: the state space $\mathcal{X}$ consists of a single (non-empty) positive recurrent class $R$ and a set $U$ of transient states such that $e \in \mathcal{R}^*(i, R)$, for $i \in U$. Moreover, both $\bar{C}_1(e)$ and $\bar{C}_2(e)$ on $R$ are finite: We consider the always update policy $\psi_{au}$ where the transmitter makes transmission attempt at every time slot. We take the set $R = \mathcal{X}$. Applying the system dynamic discussed in Section 2.2.2, we can see that, under $\psi_{au}$, all the states in $R$ communicate with state $(0, 0)$ and state $(0, 0)$ is positive recurrent. Consequently, we can conclude that the set $R$ forms a positive recurrent class. The set $U$ can simply be empty set. Finally, we notice that $\bar{C}_2(\psi_{au})$ is nothing but the expected transmission rate which is finite and $\bar{C}_1(\psi_{au})$ is the expected AoII which is also finite.

3. Given any two state $x \neq y$, there exists a policy $\pi$ such that $\pi \in \mathcal{R}^*(x, y)$: We first notice that any state $x \in \mathcal{X}$ communicates with state $(0, 0)$ with positive probability if the transmitter makes a transmission attempt at state $x$ and succeeds. We also notice that state $(0, 0)$ can reach any state $x \in \mathcal{X}$ as the minimum increase in both $d$ and $\Delta$ is one. Consequently, we can always find a policy that induces a Markov chain such that there exists a path with a positive probability between any two different states $x$ and $y$. The corresponding $\bar{C}_1^{x,y}(\pi)$, $\bar{C}_2^{x,y}(\pi)$ and $m_{x,y}(\pi)$ are trivially finite.

4. If a stationary policy $\pi$ has at least one positive recurrent state, then it has a single positive recurrent class $R$. Moreover, if $x \notin R$, then $\pi \in \mathcal{R}^*(x, R)$ where $x = (0, 0)$: We notice that, for any policy, the penalty can decrease only when the system reaches state $(0, 0)$ or $(1, 1)$. At the same time, $(0, 0)$ and $(1, 1)$ communicate with each other. Thus, any positive recurrent class must contain $(0, 0)$ and $(1, 1)$ which indicates that there can only be a single positive recurrent class.

5. There exists a policy $\pi$ such that $\bar{C}_1(\pi) < \infty$ and $\bar{C}_2(\pi) < \alpha$: We first note that $\bar{C}_2(\pi)$ is simply the expected transmission rate. Then, we can always find a policy with large enough thresholds such that $\bar{C}_2(\pi)$ is less than $\alpha$. We can easily verify that the corresponding $\bar{C}_1(\pi)$ is finite.

Some other results in [50] will be useful when constructing the optimal policy. More precisely, they are Proposition 3.2, Lemma 3.4, 3.7, 3.9 and 3.10. To this end, we define $\bar{R}_\lambda$ as the expected transmission rate associate with policy $\boldsymbol{n}_\lambda$ and $\lambda^* \triangleq \inf\{\lambda > 0 : \bar{R}_\lambda \leq \alpha\}$. We say a policy is $\lambda^*$-optimal if the policy is optimal for the MDP $\mathcal{M}$ with $\lambda = \lambda^*$.

We know that there exists $\lambda_+^* \downarrow \lambda^*$ and $\lambda_-^* \uparrow \lambda^*$ such that they both converge to $\lambda^*$. At

the same time, the corresponding optimal policies $\boldsymbol{n}_{\lambda_+^*}$ and $\boldsymbol{n}_{\lambda_-^*}$ will also converge and are both $\lambda^*$-optimal [50, Lemmas 3.4 and 3.7]. Since the Markov chains induced by policies $\boldsymbol{n}_{\lambda_+^*}$ and $\boldsymbol{n}_{\lambda_-^*}$ are both irreducible and state $(0,0)$ is positive recurrent in both Markov chains, we can choose which policy to adopt every time the system reaches state $(0,0)$ independently without changing its optimality [50, Proposition 3.2, Lemma 3.9]. Thus, we can mix the two policies in the following way: when the system reaches state $(0,0)$, the system will choose $\boldsymbol{n}_{\lambda_-^*}$ with probability $\mu$ and $\boldsymbol{n}_{\lambda_+^*}$ with probability $1-\mu$. Then the system will follow the chosen policy until the next choice. The probability $\mu$ is chosen such that the expected transmission rate of the mixed policy $\boldsymbol{n}_{\lambda^*}$ is equal to $\alpha$. More precisely,

$$\mu = \frac{\alpha - \bar{R}_{\lambda_+^*}}{\bar{R}_{\lambda_-^*} - \bar{R}_{\lambda_+^*}}.$$

Then, we can conclude that the mixed policy $\boldsymbol{n}_{\lambda^*}$ is optimal for the constrained problem (2.2) [50, Lemma 3.10].

# Appendix B: Scheduling for Freshness

## B.1 Proof of Lemma 2

We consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, that differ only in the value of $s_j$. Without the loss of generality, we assume $s_{1,j} < s_{2,j}$. Then, it is sufficient to show that, for any $1 \leq j \leq N$, $V(\boldsymbol{x}_1) \leq V(\boldsymbol{x}_2)$. Leveraging the iterative nature of VIA, we use mathematical induction to prove the monotonicity. First of all, the base case (i.e., $\nu = 0$) is true by initialization. We assume the lemma holds at iteration $\nu$. Then, we want to examine whether it holds at iteration $\nu + 1$. The update step reported in problem (3.3) can be rewritten as follows.

$$V_{\nu+1}(\boldsymbol{x}) = \min_{\boldsymbol{a} \in \mathcal{A}_N(1)} V^{\boldsymbol{a}}_{\nu+1}(\boldsymbol{x}), \tag{B.1}$$

where

$$V^{\boldsymbol{a}}_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}' - \{x'_j\}} \left\{ \left( \prod_{i \neq j} P_{x_i, x'_i}(a_i) \right) \sum_{\hat{r}'_j} P(\hat{r}'_j) U^j_\nu(\boldsymbol{x}, \boldsymbol{x}') \right\},$$

$$U^j_\nu(\boldsymbol{x}, \boldsymbol{x}') = \sum_{s'_j} P_{s_j, s'_j}(a_j, \hat{r}_j) V_\nu(\boldsymbol{x}').$$

To prove the desired results, we distinguish between the following cases.

- We first consider the case of $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 0$. When $a_j = 1$ and for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') = p_j V_\nu(\boldsymbol{x}'; s'_j = 1) + (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

$$U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') = \beta_j V_\nu(\boldsymbol{x}'; s'_j = s_{2,j} + 1) + (1 - \beta_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

where $V_\nu(\boldsymbol{x}'; s'_j = 0)$ is the estimated value function of the state $\boldsymbol{x}'$ with $s'_j = 0$ at iteration $\nu$ (at the risk of abusing the notation, we use $V(\boldsymbol{x}; s_j = s_1)$ and $V(\boldsymbol{x}; s_j = s_2)$ to represent the value functions of two states that differ only in the value of $s_j$). Then, we get

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') - U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') \leq (p_j - \beta_j)\big(V_\nu(\boldsymbol{x}'; s'_j = 1) - V_\nu(\boldsymbol{x}'; s'_j = 0)\big) \leq 0.$$

The inequalities hold since $\beta_j > p_j$ and Lemma 2 are true at iteration $\nu$ by assumption. Therefore, we have $U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') \leq U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}')$ when $a_j = 1$ for any $\boldsymbol{x}' - \{s'_j\}$.

For the case of $a_i = 1$ where $i \neq j$, we notice that $a_j = 0$. Then, for any $\boldsymbol{x}' - \{s'_j\}$, we obtain

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') = p_j V_\nu(\boldsymbol{x}'; s'_j = 1) + (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

$$U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') = (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = s_{2,j} + 1) + p_j V_\nu(\boldsymbol{x}'; s'_j = 0).$$

Therefore, when $a_i = 1$, we have

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') - U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') \leq (2p_j - 1)\big(V_\nu(\boldsymbol{x}'; s'_j = 1) - V_\nu(\boldsymbol{x}'; s'_j = 0)\big) \leq 0.$$

The inequalities hold since $2p_j - 1 < 0$ and Lemma 2 is true at iteration $\nu$ by assumption. Combining with the case of $a_j = 1$, $U_\nu^j(\boldsymbol{x}_1, \boldsymbol{x}') \leq U_\nu^j(\boldsymbol{x}_2, \boldsymbol{x}')$ holds for any $\boldsymbol{x}' - \{s_j'\}$ under any feasible action. Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ differ only in the value of $s_j$ and $C(\boldsymbol{x})$ is non-decreasing in $s_i$ for $1 \leq i \leq N$, we can see that $V_{\nu+1}^{\boldsymbol{a}}(\boldsymbol{x}_1) \leq V_{\nu+1}^{\boldsymbol{a}}(\boldsymbol{x}_2)$ for any feasible $\boldsymbol{a}$. Then, by (B.1), we can conclude that the lemma holds at iteration $\nu+1$ when $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 0$.

- When $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 1$, by replacing the $\beta_j$'s in the above case with $\alpha_j$'s, we can achieve the same result.

- When $0 < s_{1,j} < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j}$, we notice that

$$P_{s_{1,j},s_{1,j}+1}(a_j, \hat{r}_{1,j}) = P_{s_{2,j},s_{2,j}+1}(a_j, \hat{r}_{2,j}),$$

$$P_{s_{1,j},0}(a_j, \hat{r}_{1,j}) = P_{s_{2,j},0}(a_j, \hat{r}_{2,j}).$$

  Then, leveraging the monotonicity of $V_\nu(\boldsymbol{x})$ and $C(\boldsymbol{x})$, we can conclude with the same result.

Combining the three cases, we prove that the lemma also holds at iteration $\nu + 1$ of VIA. Therefore, the lemma holds at any iteration $\nu$ by mathematical induction. Since the results hold for any $1 \leq j \leq N$ and VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

## B.2 Proof of Lemma 3

We inherit the notations in the proof of Lemma 2. We still use mathematical induction to obtain the desired results. The base case $\nu = 0$ is true by initialization. We assume the lemma holds at iterative $\nu$ and examine whether it still holds at iteration $\nu + 1$. In the case of $M = 1$, we rewrite (3.3) as

$$V_{\nu+1}(\boldsymbol{x}) = \min_{1 \le j \le N} V^j_{\nu+1}(\boldsymbol{x}), \tag{B.2}$$

where

$$V^j_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'} \left\{ \left( \prod_{i \neq j} P^i_{x_i, x'_i}(0) \right) P^j_{x_j, x'_j}(1) V_\nu(\boldsymbol{x}') \right\}, \tag{B.3}$$

and $P^i_{x,x'}(a_i)$ is the probability that action $a_i$ will lead to state $x'$ when user $i$ is at state $x$. To get the desired results, we distinguish between the following cases

- We first show that $V^j_{\nu+1}(\boldsymbol{x}) = V^k_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$. According to (B.3), we have

$$V^j_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'} \left\{ \left( \prod_{i \neq j,k} P^i_{x_i, x'_i}(0) \right) P^k_{x_k, x'_k}(0) P^j_{x_j, x'_j}(1) V_\nu(\boldsymbol{x}') \right\}.$$

$$V^k_{\nu+1}(\mathscr{P}(\boldsymbol{x})) = C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\mathscr{P}(\boldsymbol{x})'} \left( \prod_{i \neq j,k} P^i_{\mathscr{P}(\boldsymbol{x})_i, \mathscr{P}(\boldsymbol{x})'_i}(0) \right) P^k_{\mathscr{P}(\boldsymbol{x})_k, \mathscr{P}(\boldsymbol{x})'_k}(1) P^j_{\mathscr{P}(\boldsymbol{x})_j, \mathscr{P}(\boldsymbol{x})'_j}(0) V_\nu(\mathscr{P}(\boldsymbol{x})').$$

It is obvious that for any $\mathscr{P}(\boldsymbol{x})'$, there always exists $\mathscr{P}(\boldsymbol{x}'') = \mathscr{P}(\boldsymbol{x})'$. Then, we obtain

$$\begin{aligned} V^k_{\nu+1}(\mathscr{P}(\boldsymbol{x})) &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \\ &\sum_{\mathscr{P}(\boldsymbol{x}'')} \left( \prod_{i \neq j,k} P^i_{x_i, x''_i}(0) \right) P^k_{x_j, \mathscr{P}(\boldsymbol{x}'')_k}(1) P^j_{x_k, \mathscr{P}(\boldsymbol{x}'')_j}(0) V_\nu(\mathscr{P}(\boldsymbol{x}'')) \\ &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\boldsymbol{x}''} \left( \prod_{i \neq j,k} P^i_{x_i, x''_i}(0) \right) P^k_{x_j, x''_j}(1) P^j_{x_k, x''_k}(0) V_\nu(\boldsymbol{x}'') \\ &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\boldsymbol{x}'} \left( \prod_{i \neq j,k} P^i_{x_i, x'_i}(0) \right) P^k_{x_j, x'_j}(1) P^j_{x_k, x'_k}(0) V_\nu(\boldsymbol{x}'). \end{aligned}$$

The second equality follows from the definition of $\mathscr{P}(\cdot)$, the property of summation, and the assumption at iteration $\nu$. The last equality follows from the variable renaming. Then, by the definition of statistically identical, we have $P^k_{x_j, x'_j}(1) = P^j_{x_j, x'_j}(1)$, $P^j_{x_k, x'_k}(0) = P^k_{x_k, x'_k}(0)$, and $C(\boldsymbol{x}) = C(\mathscr{P}(\boldsymbol{x}))$. Therefore, we can conclude that $V^j_{\nu+1}(\boldsymbol{x}) = V^k_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$.

- Along the same lines, we can easily show that $V^k_{\nu+1}(\boldsymbol{x}) = V^j_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$ and $V^i_{\nu+1}(\boldsymbol{x}) = V^i_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$ for $i \neq j, k$.

Combining the above cases with (B.2), we prove that $V_{\nu+1}(\boldsymbol{x}) = V_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$. Then, by induction, we have $V_\nu(\boldsymbol{x}) = V_\nu(\mathscr{P}(\boldsymbol{x}))$ at any iteration $\nu$. Since VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

## B.3 Proof of Theorem 3

For arbitrary $j$ and $k$

$$\delta^{j,k}(\boldsymbol{x}) = \sum_{\boldsymbol{x}' - \{x'_j, x'_k\}} \left\{ \left( \prod_{i \neq j,k} P_{x_i, x'_i}(0) \right) \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j) P(\hat{r}'_k) R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \right\}, \tag{B.4}$$

where

$$R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = \sum_{s'_j,s'_k}\left[\left(P_{s_k,s'_k}(0,\hat{r}_k)P_{s_j,s'_j}(1,\hat{r}_j) - P_{s_k,s'_k}(1,\hat{r}_k)P_{s_j,s'_j}(0,\hat{r}_j)\right)V(\boldsymbol{x}')\right]. \tag{B.5}$$

With this in mind, we will prove the properties one by one.

Property 1: $\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $\hat{r}_k = p^0_{e,k} = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$. When $\hat{r}_k = p^0_{e,k} = 0$, transmitting the update from user $k$ will necessarily fail. Therefore, $P_{s_k,s'_k}(0,0) = P_{s_k,s'_k}(1,0)$ for any $s_k$ and $s'_k$. Then, we have

$$R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = \sum_{s'_k} P_{s_k,s'_k}(0,0)\sum_{s'_j}\left[\left(P_{s_j,s'_j}(1,\hat{r}_j) - P_{s_j,s'_j}(0,\hat{r}_j)\right)V(\boldsymbol{x}')\right].$$

To identify the sign of $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$, we distinguish between the following cases

- When $s_j = 0$, we can easily show that $R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$ by noticing that the two possible actions with respect to user $j$ (i.e., $a_j = 1$ and $a_j = 0$) are equivalent when $s_j = 0$. Since $\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$'s with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x}) = 0$ in this case.

- When $s_j > 0$ and $\hat{r}_j = 1$, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, we have

$$\begin{aligned} R^{j,k}(\boldsymbol{x},\boldsymbol{x}') &= \sum_{s'_k} P_{s_k,s'_k}(0,0)(\alpha_j + p_j - 1)\big(V(\boldsymbol{x}';s'_j = s_j+1) - V(\boldsymbol{x}';s'_j = 0)\big) \\ &\leq 0. \end{aligned} \tag{B.6}$$

The inequality holds because of Lemma 2 and the fact that $\alpha_j + p_j < 1$. We recall that

$\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$'s with non-negative coefficients. Then, we can conclude that $\delta^{j,k}(\boldsymbol{x}) \leq 0$ in this case.

- When $s_j > 0$ and $\hat{r}_j = 0$, by replacing the $\alpha_j$ in (B.6) with $\beta_j$, we can get the same result. In this case, the equality holds when $\beta_j + p_j = 1$, or, equivalently, $p^0_{e,j} = 0$.

Combining the cases, we prove the first property.

Property 2: $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$ and is non-decreasing in $\hat{r}_k$ when $s_j, s_k > 0$. At the same time, $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j, k$. We first prove the monotonicity of $\delta^{j,k}(\boldsymbol{x})$ with respect to $\hat{r}_j$. To this end, we define $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as two states that differ only in the value of $\hat{r}_j$. Without a loss of generality, we assume $\hat{r}_{1,j} = 1$ and $\hat{r}_{2,j} = 0$. Then, we investigate the sign of $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$. We define $x_i \triangleq x_{1,i} = x_{2,i}$ for $i \neq j$. Then, according to (B.4), $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$ can be written as

$$\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}' - \{x'_j, x'_k\}} \left\{ \left( \prod_{i \neq j,k} P_{x_i, x'_i}(0) \right) \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j) P(\hat{r}'_k) \left( R^{j,k}(\boldsymbol{x}_1, \boldsymbol{x}') - R^{j,k}(\boldsymbol{x}_2, \boldsymbol{x}') \right) \right\}.$$

Since $x_{1,k} = x_{2,k}$, we have $P_{s_{1,k}, s'_k}(a, \hat{r}_{1,k}) = P_{s_{2,k}, s'_k}(a, \hat{r}_{2,k})$ for any $s'_k$. We recall that the transition probability is independent of $\hat{r}$ when $a = 0$. Combining with the fact that $s_{1,j} = s_{2,j}$, we also have $P_{s_{1,j}, s'_j}(0, \hat{r}_{1,j}) = P_{s_{2,j}, s'_j}(0, \hat{r}_{2,j})$ for any $s'_j$. Combining together, we obtain

$$P_{s_{1,k}, s'_k}(1, \hat{r}_{1,k}) P_{s_{1,j}, s'_j}(0, \hat{r}_{1,j}) = P_{s_{2,k}, s'_k}(1, \hat{r}_{2,k}) P_{s_{2,j}, s'_j}(0, \hat{r}_{2,j}),$$

$$P_{s_{1,k}, s'_k}(0, \hat{r}_{1,k}) = P_{s_{2,k}, s'_k}(0, \hat{r}_{2,k}).$$

Leveraging the above two problems, we have

$$R^{j,k}(\boldsymbol{x}_1, \boldsymbol{x}') - R^{j,k}(\boldsymbol{x}_2, \boldsymbol{x}') = \sum_{s'_j, s'_k} \left[ P_{s_k, s'_k}(0, \hat{r}_k) \Big( P_{s_{1,j}, s'_j}(1, \hat{r}_{1,j}) - P_{s_{2,j}, s'_j}(1, \hat{r}_{2,j}) \Big) V(\boldsymbol{x}') \right].$$

Consequently, we obtain

$$\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}' - \{x'_j\}} \left\{ \prod_{i \neq j} P_{x_i, x'_i}(0) \left[ \sum_{\hat{r}'_j} P(\hat{r}'_j) \sum_{s'_j} \Big( P_{s_{1,j}, s'_j}(1,1) - P_{s_{2,j}, s'_j}(1,0) \Big) V(\boldsymbol{x}') \right] \right\}.$$

In the following, we characterize the sign of

$$R_1 \triangleq \sum_{s'_j} \Big( P_{s_{1,j}, s'_j}(1,1) - P_{s_{2,j}, s'_j}(1,0) \Big) V(\boldsymbol{x}').$$

As $s_{1,j} = s_{2,j} > 0$, for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$R_1 = \big( (1 - \alpha_j) - (1 - \beta_j) \big) V(\boldsymbol{x}'; s'_j = 0) + (\alpha_j - \beta_j) V(\boldsymbol{x}'; s'_j = s_{1,j} + 1) \leq 0.$$

The inequality follows from Lemma 2 and the fact that $\beta_j > \alpha_j$. Since $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$ is a linear combination of $R_1$'s with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x}_1) \leq \delta^{j,k}(\boldsymbol{x}_2)$. Since $\hat{r}_{1,j} > \hat{r}_{2,j}$, we can see that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$.

In a very similar way, we can show that $\delta^{j,k}(\boldsymbol{x})$ is non-decreasing in $\hat{r}_k$. We recall that $\hat{r}_i$ will not affect the system dynamic if $a_i = 0$. Consequently, we can conclude that $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j, k$.

Combining together, we prove the second property.

Property 3: $\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_k = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$. Since the probabilities are non-negative, it is sufficient to show that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ satisfies Property 3 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. More precisely, it is sufficient to show that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$ when $s_k = 0$ and the equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$. We recall that $P_{s_k,s'_k}(1, \hat{r}_k) = P_{s_k,s'_k}(0, \hat{r}_k)$ for any $s'_k$ when $s_k = 0$. Hence, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, we have

$$R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{s'_k} \left[ P_{s_k,s'_k}(0, \hat{r}_k) \sum_{s'_j} \left( P_{s_j,s'_j}(1, \hat{r}_j) - P_{s_j,s'_j}(0, \hat{r}_j) \right) V(\boldsymbol{x}') \right].$$

Then, we investigate the following quantity for any $\boldsymbol{x}' - \{s'_j\}$

$$R_2 \triangleq \sum_{s'_j} \left( P_{s_j,s'_j}(1, \hat{r}_j) - P_{x_j,x'_j}(0, \hat{r}_j) \right) V(\boldsymbol{x}').$$

To this end, we distinguish between the following cases

- When $s_j = 0$, we have $P_{s_j,s'_j}(1, \hat{r}_j) = P_{s_j,s'_j}(0, \hat{r}_j)$ for any $s'_j$. Thus, we conclude that $R_2 = 0$ for any $\boldsymbol{x}' - \{s'_j\}$. Consequently, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') = 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$.

- When $s_j > 0$ and $\hat{r}_j = 1$, for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$R_2 = (\alpha_j - 1 + p_j) V(\boldsymbol{x}'; s'_j = s_j + 1) + (1 - \alpha_j - p_j) V(\boldsymbol{x}'; s'_j = 0) \leq 0 \tag{B.7}$$

The inequality follows from Lemma 2 and the fact that $\alpha_j + p_j < 1$. Thus, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$.

- When $s_j > 0$ and $\hat{r}_j = 0$, by replacing the $\alpha_j$ in (B.7) with $\beta_j$, we can get the same result. In this case, the equality holds when $\beta_j + p_j = 1$, or, equivalently, $p^0_{e,j} = 0$.

Combined together, we can conclude that Property 3 is true.

Property 4: $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $s_j$ if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ and is non-decreasing in $s_k$ if $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ when $s_j, s_k > 0$. We define $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ and $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ for $1 \leq i \leq N$. Such as we did in the proof of Property 3, it is sufficient to show that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ satisfies Property 4 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. We recall that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ depends on the values of $\hat{r}_j$ and $\hat{r}_k$. Therefore, we distinguish between the following cases

- In the case of $\hat{r}_j = \hat{r}_k = 1$ and $s_j, s_k > 0$, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, (B.5) can be written as

$$
\begin{aligned}
R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') &= \sum_{s'_j, s'_k} \left[ \Big( P_{s_k, s'_k}(0,1) P_{s_j, s'_j}(1,1) - P_{s_k, s'_k}(1,1) P_{s_j, s'_j}(0,1) \Big) V(\boldsymbol{x}') \right] \\
&= \big( p_k \alpha_j - (1-p_j)(1-\alpha_k) \big) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = 0) \\
&\quad + \big( (1-p_k)(1-\alpha_j) - p_j \alpha_k \big) V(\boldsymbol{x}'; s'_j = 0; s'_k = s_k + 1) \\
&\quad + \big( (1-p_k)\alpha_j - (1-p_j)\alpha_k \big) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = s_k + 1) \\
&\quad + \big( p_k(1-\alpha_j) - p_j(1-\alpha_k) \big) V(\boldsymbol{x}'; s'_j = 0; s'_k = 0).
\end{aligned}
$$

  As we can verify

$$
p_k \alpha_j - (1-p_j)(1-\alpha_k) < \frac{1}{2}(p_k + p_j - 1) < 0,
$$

$$
(1-p_k)(1-\alpha_j) - p_j \alpha_k > \frac{1}{2}(1 - p_k - p_j) > 0.
$$

We define $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ and $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ for $1 \leq i \leq N$. Then, we have

$$\Gamma_j^1 \lessgtr \Gamma_k^1 \Longrightarrow (1-p_k)\alpha_j - (1-p_j)\alpha_k \lessgtr 0.$$

Combining with Lemma 2, we can conclude that, for any $\boldsymbol{x}' - \{s_j', s_k'\}$, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ is non-increasing in $s_j$ if $\Gamma_j^1 \leq \Gamma_k^1$ and is non-decreasing in $s_k$ if $\Gamma_j^1 \geq \Gamma_k^1$.

- In the case of $\hat{r}_j = \hat{r}_k = 0$ and $s_j, s_k > 0$, by replacing the $\alpha$'s in the above case with $\beta$'s, we can conclude with the same result.

- In the case of $\hat{r}_j = 1$, $\hat{r}_k = 0$, and $s_j, s_k > 0$, for any $\boldsymbol{x}' - \{s_j', s_k'\}$, (B.5) can be written as

$$\begin{aligned}
R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') &= \sum_{s_j', s_k'} \left[ \left( P_{s_k, s_k'}(0,0) P_{s_j, s_j'}(1,1) - P_{s_k, s_k'}(1,0) P_{s_j, s_j'}(0,1) \right) V(\boldsymbol{x}') \right] \\
&= \big(p_k \alpha_j - (1-p_j)(1-\beta_k)\big) V(\boldsymbol{x}'; s_j' = s_j + 1; s_k' = 0) \\
&\quad + \big((1-p_k)(1-\alpha_j) - p_j \beta_k\big) V(\boldsymbol{x}'; s_j' = 0; s_k' = s_k + 1) \\
&\quad + \big((1-p_k)\alpha_j - (1-p_j)\beta_k\big) V(\boldsymbol{x}'; s_j' = s_j + 1; s_k' = s_k + 1) \\
&\quad + \big(p_k(1-\alpha_j) - p_j(1-\beta_k)\big) V(\boldsymbol{x}'; s_j' = 0; s_k' = 0).
\end{aligned}$$

As we can verify

$$p_k \alpha_j - (1-p_j)(1-\beta_k) < p_k \left( p_j - \frac{1}{2} \right) < 0,$$

$$(1-p_k)(1-\alpha_j) - p_j \beta_k > (1-p_k) \left( \frac{1}{2} - p_j \right) > 0.$$

At the same time

$$\Gamma_j^1 \lesseqgtr \Gamma_k^0 \Longrightarrow (1-p_k)\alpha_j - (1-p_j)\beta_k \lesseqgtr 0.$$

Combined with Lemma 2, we can conclude that, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ is non-increasing in $s_j$ if $\Gamma_j^1 \leq \Gamma_k^0$ and is non-decreasing in $s_k$ if $\Gamma_j^1 \geq \Gamma_k^0$.

- In the case of $\hat{r}_j = 0$, $\hat{r}_k = 1$, and $s_j, s_k > 0$, by swapping the $\alpha$'s and $\beta$'s in the above case, we can conclude with the same result.

Combined together, we conclude that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ satisfies Property 3 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. Consequently, $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $s_j$ if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ and is non-decreasing in $s_k$ if $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ when $s_j, s_k > 0$.

Property 5: $\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_j \geq s_k$, $\hat{r}_j \geq \hat{r}_k$, and users $j$ and $k$ are statistically identical. According to Property 3, it is sufficient to consider the case where $s_j, s_k > 0$. We notice that the sign of $\delta^{j,k}(\boldsymbol{x})$ can be captured by the sign of $Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \triangleq \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j)P(\hat{r}'_k)R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$. Thus, we divide our discussion into the following cases.

- We first consider the case of $s_j \geq s_k > 0$ and $\hat{r}_j = \hat{r}_k = 0$. Leveraging the definition of statistically identical, for any $\boldsymbol{x}' - \{x'_j, x'_k\}$, we have

$$Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j)P(\hat{r}'_k)\kappa_1 \Big( V(\boldsymbol{x}'; x'_j = (0, \hat{r}'_j); x'_k = (s_k + 1, \hat{r}'_k)) - V(\boldsymbol{x}'; x'_j = (s_j + 1, \hat{r}'_j); x'_k = (0, \hat{r}'_k)) \Big),$$

where $\kappa_1 = 1 - p_j - \beta_j \geq 0$. Then, by substituting the values of $P(\hat{r})$ and using Lemma 3, we obtain

$$
\begin{aligned}
Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') =& \gamma_j \gamma_k \kappa_1 V(\boldsymbol{x}'; x'_j = (s_k + 1, 1); x'_k = (0, 1)) - \\
& \gamma_j \gamma_k \kappa_1 V(\boldsymbol{x}'; x'_j = (s_j + 1, 1); x'_k = (0, 1)) + \\
& (1 - \gamma_j)(1 - \gamma_k) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_k + 1, 0); x'_k = (0, 0)) - \\
& (1 - \gamma_j)(1 - \gamma_k) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_j + 1, 0); x'_k = (0, 0)) + \\
& \gamma_k (1 - \gamma_j) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_k + 1, 1); x'_k = (0, 0)) - \\
& \gamma_k (1 - \gamma_j) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_j + 1, 0); x'_k = (0, 1)) + \\
& \gamma_j (1 - \gamma_k) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_k + 1, 0); x'_k = (0, 1)) - \\
& \gamma_j (1 - \gamma_k) \kappa_1 V(\boldsymbol{x}'; x'_j = (s_j + 1, 1); x'_k = (0, 0)).
\end{aligned}
$$

Since users $j$ and $k$ are statistically identical, we have $\gamma_j = \gamma_k$. Then, by Lemma 2, we have $Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq 0$ for any $\boldsymbol{x}' - \{x'_j, x'_k\}$. Since $\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$'s with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x}) \leq 0$.

- For the case of $s_j \geq s_k > 0$ and $\hat{r}_j = \hat{r}_k = 1$, by replacing $\beta_j$ in $\kappa_1$ with $\alpha_j$, we can conclude with the same result.

- Then, we consider the case of $s_j \geq s_k > 0$, $\hat{r}_j = 1$, and $\hat{r}_k = 0$. We first notice that, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$

$$
\begin{aligned}
R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') =& \big(p_k \alpha_j - (1 - p_j)(1 - \beta_k)\big) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = 0) + \\
& \big((1 - p_k)(1 - \alpha_j) - p_j \beta_k\big) V(\boldsymbol{x}'; s'_j = 0; s'_k = s_k + 1) + \\
& \big((1 - p_k)\alpha_j - (1 - p_j)\beta_k\big) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = s_k + 1) + \\
& \big(p_k(1 - \alpha_j) - p_j(1 - \beta_k)\big) V(\boldsymbol{x}'; s'_j = 0; s'_k = 0).
\end{aligned}
$$

As users $j$ and $k$ are statistically identical, we have $p_j = p_k$ and $\alpha_j < \beta_k$. Leveraging Lemma 2, we have

$$R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq (\alpha_j + p_j - 1)\Big(V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = 0) - V(\boldsymbol{x}'; s'_j = 0; s'_k = s_k + 1)\Big).$$

Then, for any $\boldsymbol{x}' - \{x'_j, x'_k\}$

$$Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j)P(\hat{r}'_k)\kappa_2\Big(V(\boldsymbol{x}'; x'_j = (0, \hat{r}'_j); x'_k = (s_k + 1, \hat{r}'_k)) - V(\boldsymbol{x}'; x'_j = (s_j + 1, \hat{r}'_j); x'_k = (0, \hat{r}'_k))\Big),$$

where $\kappa_2 = 1 - p_j - \alpha_j > 0$. Such as we did in the previous cases, we can leverage Lemmas 2 and 3 to conclude that $Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq 0$ for any $\boldsymbol{x}' - \{x'_j, x'_k\}$. Consequently, $\delta^{j,k}(\boldsymbol{x}) \leq 0$ in this case. The details are omitted for the sake of space.

Combined together, we conclude the proof of Property 5.

## B.4 Proof of Corollary 4

We follow the same steps as in the proof of Lemma 2. To prove the corollary, it is sufficient to show that $V(x_1) \leq V(x_2)$ when $s_1 < s_2$ and $\hat{r}_1 = \hat{r}_2$. We use mathematical induction to prove the monotonicity. First of all, the base case (i.e., $\nu = 0$) is true by initialization. We assume the lemma holds at iteration $\nu$. Then, we want to examine whether it holds at iteration $\nu + 1$. For the system with a single user, the update step reported in problem (3.3) can be simplified and

rewritten as follows

$$V_{\nu+1}(x) = \min_{a\in\{0,1\}} V^a_{\nu+1}(x), \tag{B.8}$$

where

$$V^a_{\nu+1}(x) = C(x,a) - \theta + \sum_{\hat{r}'} P(\hat{r}') \sum_{s'} P_{s,s'}(a,\hat{r}) V_\nu(x'),$$

and $\theta$ is the optimal value for $\mathcal{M}_1(\lambda,-1)$. To prove the desired results, we distinguish between the following cases

- We first consider the case of $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 0$. When $a = 1$, we have

$$V^1_{\nu+1}(x_1) = C(x_1,1) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big( pV_\nu(1,\hat{r}') + (1-p)V_\nu(0,\hat{r}') \Big),$$

$$V^1_{\nu+1}(x_2) = C(x_2,1) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big( \beta V_\nu(s_2+1,\hat{r}') + (1-\beta)V_\nu(0,\hat{r}') \Big).$$

  Subtracting the two expressions yields

$$\begin{aligned} &V^1_{\nu+1}(x_1) - V^1_{\nu+1}(x_2) \\ &\qquad \le C(x_1,1) - C(x_2,1) + \sum_{\hat{r}'} P(\hat{r}') \Big[ (p-\beta)\big(V_\nu(1,\hat{r}') - V_\nu(0,\hat{r}')\big) \Big] \le 0. \end{aligned}$$

  The inequalities hold since $\beta > p$, $C(x,a)$ is non-decreasing in $s$, and Corollary 4 is true at iteration $\nu$ by assumption.

  For the case of $a = 0$, we obtain

$$V^0_{\nu+1}(x_1) = C(x_1,0) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big( pV_\nu(1,\hat{r}') + (1-p)V_\nu(0,\hat{r}') \Big),$$

$$V^0_{\nu+1}(x_2) = C(x_2, 0) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big( (1-p) V_\nu(s_2 + 1, \hat{r}') + p V_\nu(0, \hat{r}') \Big).$$

Therefore, when $a = 0$, we have

$$\begin{aligned} & V^0_{\nu+1}(x_1) - V^0_{\nu+1}(x_2) \\ & \qquad \leq C(x_1, 0) - C(x_2, 0) + \sum_{\hat{r}'} P(\hat{r}') \Big[ (2p-1) \big( V_\nu(1, \hat{r}') - V_\nu(0, \hat{r}') \big) \Big] \leq 0. \end{aligned}$$

The inequalities hold since $2p - 1 < 0$, $C(x, a)$ is non-decreasing in $s$, and Corollary 4 is true at iteration $\nu$ by assumption. Combined together, we can see that $V^a_{\nu+1}(x_1) \leq V^a_{\nu+1}(x_2)$ for any feasible $a$. Then, by problem (B.8), we can conclude that the lemma holds at iteration $\nu + 1$ when $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 0$.

- When $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 1$, by replacing the $\beta$'s in the above case with $\alpha$'s, we can achieve the same result.

- When $0 < s_1 < s_2$ and $\hat{r}_1 = \hat{r}_2$, we notice that $P_{s_1, s_1+1}(a, \hat{r}_1) = P_{s_2, s_2+1}(a, \hat{r}_2)$ and $P_{s_1,0}(a, \hat{r}_1) = P_{s_2,0}(a, \hat{r}_2)$. Then, leveraging the monotonicity of $V_\nu(x)$ and $C(x, a)$, we can conclude with the same result.

Combining the three cases, we prove that the lemma holds at iteration $\nu + 1$ of VIA. Therefore, the lemma holds at any iteration $\nu$ by mathematical induction. Since VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

## B.5 Proof of Proposition 3

We define $\Delta V(x) \triangleq V^1(x) - V^0(x)$ where $V^a(x)$ is the value function resulting from taking action $a$ at state $x$. Then, $V^a(x)$ can be calculated as follows

$$V^a(x) = C(x,a) - \theta + \sum_{x' \in \mathcal{X}} P_{x,x'}(a) V(x'), \tag{B.9}$$

where $\theta$ is the optimal value for $\mathcal{M}_1(\lambda, -1)$. Hence, the optimal action at state $x$ can be fully characterized by the sign of $\Delta V(x)$. More precisely, the optimal action at state $x$ is $a = 1$ if $\Delta V(x) < 0$, and $a = 0$ is optimal otherwise. To determine the sign of $\Delta V(x)$ for each state, we distinguish between the following cases

- We first consider the state $x = (0, \hat{r})$. Applying the results in Section 3.2.2 to problem (B.9), we obtain

$$\begin{aligned} V^0(0,\hat{r}) = &- \theta + (1-\gamma)(1-p)V(0,0) + (1-\gamma)pV(1,0) + \\ &\gamma(1-p)V(0,1) + \gamma p V(1,1), \end{aligned}$$

$$V^1(0,\hat{r}) = \lambda + V^0(0,\hat{r}). \tag{B.10}$$

  Therefore, $\Delta V(0,\hat{r}) = \lambda \geq 0$. Thus, the optimal action at state $(0,\hat{r})$ is $a = 0$.

- Then, we consider the state $x = (s, 0)$ where $s > 0$. Applying the results in Section 3.2.2

to Equation (B.9), we obtain

$$V^0(s,0) = f(s) - \theta + (1-\gamma)pV(0,0) + (1-\gamma)(1-p)V(s+1,0) + \gamma pV(0,1) + \gamma(1-p)V(s+1,1),$$

$$V^1(s,0) = f(s) + \lambda - \theta + (1-\gamma)(1-\beta)V(0,0) + (1-\gamma)\beta V(s+1,0) + \gamma(1-\beta)V(0,1) + \gamma\beta V(s+1,1).$$

Then,

$$\Delta V(s,0) = \lambda + p_e^0(1-2p)\omega, \tag{B.11}$$

where $\omega = (1-\gamma)[V(0,0) - V(s+1,0)] + \gamma[V(0,1) - V(s+1,1)] \leq 0$.

- Finally, we consider the state $x = (s,1)$ where $s > 0$. Following the same trajectory, we have

$$\Delta V(s,1) = \lambda + (1-p_e^1)(1-2p)\omega.$$

According to Corollary 4 and the fact that $p < 0.5$, we can see that $\Delta V(s,0)$ and $\Delta V(s,1)$ are both a constant $\lambda$ plus a term that is non-increasing in $s$. As the time penalty function is unbounded, the value function must also be unbounded. Then, combining the three cases, we can conclude the following. For fixed $\hat{r}$, there always exists a threshold $n_{\hat{r}} > 0$ such that the optimal action at state $(s,\hat{r})$ where $s \geq n_{\hat{r}}$ is $a = 1$, otherwise $a = 0$ is optimal. Since $\hat{r} \in \{0,1\}$, the optimal policy can be fully captured by the pair $(n_0, n_1)$.

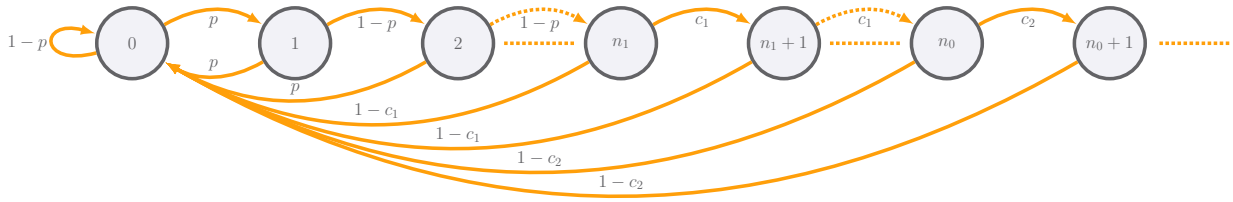



Figure B.1: Illustration of the DTMC induced by the threshold policy $\boldsymbol{n} = (n_0, n_1)$. In the figure, $c_1 = (1-\gamma)(1-p) + \gamma\alpha$ and $c_2 = (1-\gamma)\,\beta + \gamma\alpha$.

In the following, we determine the relationship between $n_0$ and $n_1$. We have

$$\Delta V(s,1) - \Delta V(s,0) = (1 - p_e^1 - p_e^0)(1-2p)\omega \leq 0.$$

At the same time, for the threshold $n_0$, we know $\Delta V(n_0, 0) < 0$. Then, we have $\Delta V(n_0, 1) \leq \Delta V(n_0, 0) < 0$. Combined with the fact that $\Delta V(s, \hat{r})$ is non-increasing in $s$, we can conclude that the ordering $n_0 \geq n_1$ is true.

## B.6 Proof of Proposition 4

We notice that the dynamic of AoII under threshold policy can be fully captured by a DTMC. Then, the expected AoII $\bar{\Delta}_{\boldsymbol{n}}$ and the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ under threshold policy $\boldsymbol{n} = (n_0, n_1)$ can be obtained from the stationary distribution of the induced DTMC. Let the states of the induced DTMC be the values of $s$. We recall that $\hat{r}$ is an independent Bernoulli random variable with parameter $\gamma$. Combined with the results in Section 3.2.2, we can easily obtain the state transition probabilities of the induced DTMC, which are shown in Figure B.1. The balance

equations of the induced DTMC are the following

$$(1-p)\pi_0 + p\sum_{k=1}^{n_1-1}\pi_k + (1-c_1)\sum_{k=n_1}^{n_0-1}\pi_k + (1-c_2)\sum_{k=n_0}^{+\infty}\pi_k = \pi_0.$$

$$p\pi_0 = \pi_1.$$

$$(1-p)\pi_{k-1} = \pi_k \;\; for\; 2 \le k \le n_1.$$

$$c_1\pi_{k-1} = \pi_k \;\; for\; n_1+1 \le k \le n_0.$$

$$c_2\pi_{k-1} = \pi_k \;\; for\; n_0+1 \le k.$$

$$\sum_{k=0}^{+\infty}\pi_k = 1.$$

Then, we can easily solve the above system of linear equations. After some algebraic manipulation, we obtain the following

$$\pi_0 = \frac{1}{2 + p(1-p)^{n_1-1}\left[\frac{1}{1-c_1} - \frac{1}{p} + c_1^{n_0-n_1}\left(\frac{1}{1-c_2} - \frac{1}{1-c_1}\right)\right]}.$$

$$\pi_k = p(1-p)^{k-1}\pi_0 \;\; for\; 1 \le k \le n_1.$$

$$\pi_k = p(1-p)^{n_1-1}c_1^{k-n_1}\pi_0 \;\; for\; n_1+1 \le k \le n_0.$$

$$\pi_k = p(1-p)^{n_1-1}c_1^{n_0-n_1}c_2^{k-n_0}\pi_0 \;\; for\; n_0+1 \le k.$$

Equipped with the above results, we proceed with calculating $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$. According to (3.4), the expected AoII is:

$$\bar{\Delta}_{\boldsymbol{n}} = \sum_{k=0}^{+\infty} f(k)\pi_k.$$

Substituting the expressions of $\pi_k$'s, we can get the expression of $\bar{\Delta}_{\boldsymbol{n}}$. Proposition 3 tells us the following.

- For state $(s, \hat{r})$ where $s < n_1$, it is optimal to stay idle (i.e., $a = 0$).
- For state $(s, \hat{r})$ where $n_1 \leq s < n_0$, it is optimal to make a transmission attempt only when $\hat{r} = 1$. We recall that $\hat{r}$ is an independent Bernoulli random variable with parameter $\gamma$. Therefore, the expected proportion of time that the system is at state $(s, 1)$ is $\gamma\pi_s$.
- For state $(s, \hat{r})$ where $s \geq n_0$, it is optimal to make transmission attempt regardless of $\hat{r}$.

Combined with (3.4), we have

$$\bar{\rho}_{\boldsymbol{n}} = \gamma \sum_{k=n_1}^{n_0-1} \pi_k + \sum_{k=n_0}^{+\infty} \pi_k.$$

Substituting the expressions of $\pi_k$'s, we can obtain the closed-form expression of $\bar{\rho}_{\boldsymbol{n}}$.

## B.7 Proof of Proposition 6

We first tackle the Whittle's indexes at state $(0, \hat{r})$ and $(s, 0)$ where $s > 0$. To this end, we distinguish between the following cases

- We first consider the state $x = (0, \hat{r})$. By definition, Whittle's index is the infimum $\lambda$ such that $V^0(x) = V^1(x)$. According to (B.10), we can conclude that $W_x = 0$ when $x = (0, \hat{r})$.

- Then, we consider the state $x = (s, 0)$ where $s > 0$. We recall that $p_e^0 = 0$. Then, we can conclude, from (B.11), that $W_x = 0$ for all $x = (s, 0)$ where $s > 0$.

Now, we tackle the Whittle's index at state $x = (s, 1)$ where $s > 0$. For convenience, we denote by $W_n$ the Whittle's index at state $x = (n, 1)$. According to the monotonicity of $\Delta V(x)$ shown in the proof of Proposition 3, we can conclude that threshold policy $\boldsymbol{n} = (+\infty, n + 1)$ is optimal when $V^0(n, 1) = V^1(n, 1)$. Then, we can prove the following

**Lemma 15.** *When* (3.7) *is satisfied and* $V^0(n, 1) = V^1(n, 1)$, $V(s, 1) = V(s, 0) \triangleq V(s)$ *for* $0 \leq s \leq n$.

*Proof.* Since the value function satisfies the Bellman equation, it is sufficient to show that $V(s, 1)$ and $V(s, 0)$ satisfy the same Bellman equation. We recall that the Bellman equation for $V(x)$ is given by

$$V(x) = \min_{a \in \{0,1\}} V^a(x),$$

where

$$V^a(x) = C(x, a) - \theta + \sum_{x'} P_{x,x'}(a) V(x'), \tag{B.12}$$

and $\theta$ is the optimal value of the decoupled problem. We recall, from Corollary 5, that the optimal action at state $(s, 0)$ is staying idle (i.e., $a = 0$) for any $s$. We also know that threshold policy $\boldsymbol{n} = (+\infty, n + 1)$ is optimal when $V^0(n, 1) = V^1(n, 1)$. Therefore, the optimal actions at states $(s, 0)$ and $(s, 1)$ where $s \leq n$ are the same (i.e., $a = 0$). Equivalently, we have

$$V(s, \hat{r}) = V^0(s, \hat{r}), \quad for\ s \leq n. \tag{B.13}$$

According to the system dynamic reported in Section 3.2.2, we know that the state transition

probabilities are independent of $\hat{r}$ when $a = 0$. Meanwhile, $\hat{r}$ does not affect the instant cost. Let $x_1 = (s, 1)$ and $x_2 = (s, 0)$. Then, for any $x'$, we have

$$P_{x_1,x'}(0) = P_{x_2,x'}(0).$$

$$C(x_1, 0) = C(x_2, 0).$$

Hence, according to (B.12), we can see that $V^0(s, 0) = V^0(s, 1)$ for any $s \leq n$. Combined with problem (B.13), we can conclude that $V(s, 0) = V(s, 1)$ for any $0 \leq s \leq n$. □

By definition, Whittle's index $W_n$ is the infimum $\lambda$ such that $V^0(n, 1) = V^1(n, 1)$. In this case, according to Lemma 15, $V(0, 1) = V(0, 0) = V(0)$. Then, $V^0(n, 1)$ and $V^1(n, 1)$ can be written as

$$V^0(n, 1) = f(n) - \theta + pV(0) + (1 - p)[(1 - \gamma)V(n + 1, 0) + \gamma V(n + 1, 1)]. \tag{B.14}$$

$$V^1(n, 1) = f(n) + W_n - \theta + (1 - \alpha)V(0) + \alpha[(1 - \gamma)V(n + 1, 0) + \gamma V(n + 1, 1)].$$

Without a loss of generality, we assume $V(0) = 0$. Then, equating the two expressions yields

$$W_n = (1 - p - \alpha)(\gamma V(n + 1, 1) + (1 - \gamma)V(n + 1, 0)). \tag{B.15}$$

Combining problems (B.14) and (B.15), we conclude that $W_n$ is

$$W_n = \frac{(1 - p - \alpha)(V^0(n, 1) + \theta - f(n))}{1 - p}.$$

Since the optimal action at state $(n, 1)$ is $a = 0$, we have $V^0(n, 1) = V(n, 1) = V(n)$. Finally, we obtain

$$W_n = \frac{(1 - p - \alpha)(V(n) + \theta - f(n))}{1 - p}. \tag{B.16}$$

Now, we tackle the expression of $V(n)$. When $V^0(n, 1) = V^1(n, 1)$, the optimal action at state $(s, \hat{r})$ where $0 \leq s < n$ is staying idle. Then, leveraging Lemma 15, value function $V(s)$ where $0 \leq s < n$ satisfies the following

$$V(s) = \begin{cases} -\theta + f(0) + pV(1) & s = 0, \\ -\theta + f(s) + (1 - p)V(s + 1) & 0 < s < n. \end{cases} \tag{B.17}$$

By backward induction, we end up with the following equation for $0 < s < n$.

$$V(s) = \frac{-\theta(1 - (1 - p)^{n-s})}{p} + \sum_{k=1}^{n-s} f(n - k)(1 - p)^{n-s-k} + (1 - p)^{n-s}V(n).$$

Letting $s = 1$ yields

$$V(1) = \frac{-\theta(1 - (1 - p)^{n-1})}{p} + \sum_{k=1}^{n-1} f(n - k)(1 - p)^{n-1-k} + (1 - p)^{n-1}V(n).$$

From problem (B.17), $V(1)$ also satisfies the following

$$V(1) = \frac{\theta - f(0)}{p}.$$

Equating the two expressions of $V(1)$, we obtain

$$V(n) = \frac{-f(0)}{p(1-p)^{n-1}} + \theta\left(\frac{2}{p(1-p)^{n-1}} - \frac{1}{p}\right) - \sum_{k=1}^{n-1} f(n-k)(1-p)^{-k}. \tag{B.18}$$

We recall that, when $V^0(n,1) = V^1(n,1)$, threshold policy $\boldsymbol{n} = (+\infty, n+1)$ is optimal and both actions at state $x = (n,1)$ are equally desirable. Thus, threshold policy $\boldsymbol{n} = (+\infty, n)$ is also optimal. Then, we know

$$\theta = \bar{\Delta}_n + W_n \bar{\rho}_n, \tag{B.19}$$

where $\bar{\Delta}_n$ and $\bar{\rho}_n$ are the expected AoII and the expected transmission rate under threshold policy $\boldsymbol{n} = (+\infty, n)$, respectively. Finally, combining problems (B.16), (B.18), and (B.19), we obtain

$$W_n = \frac{\dfrac{-f(0)}{p(1-p)^n} + \bar{\Delta}_n \dfrac{2-(1-p)^n}{p(1-p)^n} - (1-p)^{-n}\left(\displaystyle\sum_{k=1}^{n} f(k)(1-p)^{k-1}\right)}{\dfrac{1}{1-p-\alpha} - \bar{\rho}_n \dfrac{2-(1-p)^n}{p(1-p)^n}}.$$

After some algebraic manipulation, we have

$$W_n = \frac{(1-c_1)\displaystyle\sum_{k=n+1}^{+\infty} f(k) c_1^{k-n-1} - \bar{\Delta}_n}{\dfrac{(1-c_1)(1-p) - \gamma(1-p-\alpha)}{c_1(1-p-\alpha)} + \bar{\rho}_n},$$

where $c_1 = (1-\gamma)(1-p) + \gamma\alpha$.

In the following, we investigate some properties of Whittle's index. First of all, $W_n$ is non-negative since $1-p-\alpha$ and $V(n+1,\hat{r})$ in (B.15) are all non-negative. Meanwhile, combining (B.15) with the fact that $V(n,\hat{r})$ is non-decreasing in $n$, we can verify that $W_n$ is non-decreasing

in $n$. Combined with the Whittle's indexes in two other cases (i.e., $x = (0, \hat{r})$ and $x = (s, 0)$ where $s > 0$), we can easily obtain the properties of $W_x$ as detailed in Proposition 6.

## B.8 Proof of Proposition 7

We notice that $\mathcal{M}_1(\lambda, -1)$ coincides with the decoupled model studied in Section 3.4.2. When problem (3.7) is satisfied, the decoupled problem is indexable, and, by Corollary 5, we only need to show that $n$ is the optimal threshold for the states with $\hat{r} = 1$. We first tackle the case of $\lambda > 0$. To this end, we divide our discussion into the following cases

- For state $(s, 1)$ where $s < n$, $W_s \leq \lambda$ by definition. As the problem is indexable, we have $D(W_s) \subseteq D(\lambda)$. We recall that $W_s \triangleq \min\{\lambda' \geq 0 : V^0(s, 1) = V^1(s, 1)\}$. Equivalently, $W_s \triangleq \min\{\lambda' \geq 0 : (s, 1) \in D(\lambda')\}$. Then, we know $(s, 1) \in D(W_s)$. Combined together, we conclude that $(s, 1) \in D(\lambda)$. In other words, the optimal action at state $(s, 1)$ where $s < n$ is to stay idle (i.e., $a = 0$).

- For state $(s, 1)$ where $s \geq n$, we first recall that $W_s = \min\{\lambda' \geq 0 : (s, 1) \in D(\lambda')\}$. Consequently, for any $\lambda' < W_s$, we know $(s, 1) \notin D(\lambda')$. Meanwhile, we have $W_s \geq W_n > \lambda$ by the monotonicity of Whittle's index and the definition of $n$. Hence, we can conclude that $(s, 1) \notin D(\lambda)$. In other words, the optimal action at state $(s, 1)$ where $s \geq n$ is to make the transmission attempt.

Then, we conclude that $n$ is the optimal threshold for the states with $\hat{r} = 1$ when $\lambda > 0$. In the case of $\lambda = 0$, according to the proof of Proposition 3, we can easily verify that the optimal threshold is 1.

## B.9 Proof of Theorem 4

We first make the following definitions. When $\mathcal{M}_1(\lambda, -1)$ is at state $x$ and action $a$ is taken, cost $C_1(x,a) \triangleq f(s)$ and $C_2(x,a) \triangleq \lambda a$ are incurred. We denote the expected $C_1$-cost and the expected $C_2$-cost under policy $\phi$ as $\bar{C}_1(\phi)$ and $\bar{C}_2(\phi)$, respectively. Let $G$ be a non-empty set of states. For the given state $i$, we define $\mathcal{R}^*(i,G)$ as the class of policies $\phi$, for which the following hold

- The probability $P_\phi(x_n \in G \; for \; some \; n \geq 1 \mid x_0 = i) = 1$ where $x_n$ is the state of $\mathcal{M}_1(\lambda, -1)$ at time $n$.
- The expected time $m_{iG}(\phi)$ of a first passage from $i$ to $G$ under $\phi$ is finite.
- The expected $C_1$-cost $\bar{C}_1^{i,G}(\phi)$ and the expected $C_2$-cost $\bar{C}_2^{i,G}(\phi)$ of a first passage form $i$ to $G$ under $\phi$ are finite.

With the definitions in mind, we proceed with verifying the assumptions given in [50].

1. *For all $d > 0$, the set $A(d) = \{x \mid$ there exists an action $a$ such that $C_1(x,a) + C_2(x,a) \leq d\}$ is finite*: For any state $x$, the cost satisfies $C_1(x,a) + C_2(x,a) = f(s) + \lambda a \geq f(s)$. The equality holds when $a = 0$. Then, the states in $A(d)$ must satisfy $f(s) \leq d$. Combined with the fact that $f(s)$ is a non-decreasing and unbounded function when $s \in \mathbb{N}_0$, we can conclude that $A(d)$ is finite.

2. *There exists a stationary policy $e$ such that the induced Markov chain has the following properties: the state space $\mathcal{S}$ consists of a single (non-empty) positive recurrent class $R$ and a set $U$ of transient states such that $e \in \mathcal{R}^*(i,R)$ for $i \in U$. Moreover, both*

$\bar{C}_1(e)$ *and* $\bar{C}_2(e)$ *on* $R$ *are finite*: We consider the policy under which the base station makes a transmission attempt at every time slot. According to the system dynamic detailed in Section 3.2.2, we can see that all the states communicate with state $(0,0)$ and $(0,0)$ communicates with all other states. Thus, the state space $\mathcal{S}$ consists of a single (non-empty) positive recurrent class and the set of transient states can simply be an empty set. $\bar{C}_1(e)$ and $\bar{C}_2(e)$ are trivially finite as we can verify using Proposition 4.

3. *Given any two state* $x \neq y$*, there exists a policy* $\phi$ *such that* $\phi \in \mathcal{R}^*(x,y)$: We notice that, under any policy, the maximum increase of $s$ between two consecutive time slots is 1. Meanwhile, when $s$ decreases, it decreases to zero. Combined with the fact that $\hat{r}$ is an independent Bernoulli random variable, we can conclude that there always exists a path between any $x$ and $y$ with positive probability. $m_{xy}(\phi)$, $\bar{C}_1^{x,y}(\phi)$, and $\bar{C}_2^{x,y}(\phi)$ are trivially finite.

4. *If a stationary policy* $\phi$ *has at least one positive recurrent state, then it has a single positive recurrent class* $R$*. Moreover, if* $x = (0,0) \notin R$*, then* $\phi \in \mathcal{R}^*(x,R)$: Given that $\hat{r}$ is an independent Bernoulli random variable, we can easily conclude from the system dynamic that all the states communicate with state $(0,0)$ and $(0,0)$ communicates with all other states under any stationary policy. Therefore, any positive recurrent class must contain state $(0,0)$. Thus, there must have only one positive recurrent class which is $R = \mathcal{S}$.

5. *There exists a policy* $\phi$ *such that* $\bar{C}_1(\phi) < \infty$ *and* $\bar{C}_2(\phi) < K$ *where* $K \in (0,1]$: We notice that $\bar{C}_1(\phi)$ and $\bar{C}_2(\phi)$ are nothing but the expected AoII and the expected transmission rate achieved by $\phi$, respectively. Then, we can easily verify that such policy exists using Proposition 4.

As the assumptions are verified, we proceed with introducing the optimal randomized policy for given $\lambda$. We say a policy is $\lambda$-optimal if the policy is optimal for $\mathcal{M}_1(\lambda, -1)$. We consider two monotone sequences $\lambda_+^n \downarrow \lambda$ and $\lambda_-^n \uparrow \lambda$. Then, there exist subsequences of $\lambda_+^n$ and $\lambda_-^n$ such that the corresponding sequences of optimal policies converge. Then, according to [50, Lemma 3.7], the limit points, denoted by $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$, are both $\lambda$-optimal. By [50, Proposition 3.2], the Markov chains induced by $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ both contain a single non-empty positive recurrent class and state $(0,0)$ is positive recurrent in both induced Markov chains. Hence, the base station can choose which policy to follow each time the system reaches state $(0,0)$ while keeping the resulting randomized policy $\lambda$-optimal as suggested by [50, Lemma 3.9]. More precisely, we consider the following randomized policy: each time the system reaches state $(0,0)$, the base station will choose $\boldsymbol{n}_{\lambda_-}$ with probability $\mu$ and $\boldsymbol{n}_{\lambda_+}$ with probability $1-\mu$. The chosen policy will be followed until the next choice. We denote such policy as $\boldsymbol{n}_\lambda$ and conclude that $\boldsymbol{n}_\lambda$ is $\lambda$-optimal under any $\mu \in [0,1]$.

## B.10 Proof of Proposition 8

The value function $V(\boldsymbol{x})$ and $V^i(x_i)$ must satisfy their own Bellman equations. More precisely

$$V(\boldsymbol{x}) + \theta = \min_{\boldsymbol{a} \in \mathcal{A}_N(-1)} \left\{ C(\boldsymbol{x}, \boldsymbol{a}) + \sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x}, \boldsymbol{a}) V(\boldsymbol{x}') \right\},$$

$$V^i(x_i) + \theta_i = \min_{a_i \in \{0,1\}} \left\{ C(x_i, a_i) + \sum_{x_i'} Pr(x_i' \mid x_i, a_i) V^i(x_i') \right\}, \tag{B.20}$$

where $\theta$ and $\theta_i$ are the optimal values of $\mathcal{M}_N(\lambda, -1)$ and $\mathcal{M}_1^i(\lambda, -1)$, respectively. We recall from Section 3.2.2 that the users are independent when action $\boldsymbol{a}$ and current state $\boldsymbol{x}$ are given.

Thus

$$Pr(\boldsymbol{x}' \mid \boldsymbol{x}, \boldsymbol{a}) = \prod_{i=1}^{N} Pr(x_i' \mid \boldsymbol{x}, \boldsymbol{a}),$$

where $\boldsymbol{x}' = (x_1', ..., x_N')$. Then, we have

$$\sum_{\boldsymbol{x}' - \{x_i'\}} Pr(\boldsymbol{x}' - \{x_i'\} \mid \boldsymbol{x}, \boldsymbol{a}) = \sum_{\boldsymbol{x}' - \{x_i'\}} \prod_{j \neq i} Pr(x_j' \mid \boldsymbol{x}, \boldsymbol{a}) = 1.$$

We also recall from Section 3.2.2 that the state of user $i$ depends only on its previous state and the action with respect to user $i$. Thus

$$Pr(x_i' \mid \boldsymbol{x}, \boldsymbol{a}) = Pr(x_i' \mid x_i, a_i).$$

Combined together, we obtain

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{x_i'} Pr(x_i' \mid x_i, a_i) V^i(x_i') &= \sum_{i=1}^{N} \sum_{x_i'} \left[ \sum_{\boldsymbol{x}' - \{x_i'\}} \prod_{j \neq i} Pr(x_j' \mid \boldsymbol{x}, \boldsymbol{a}) \right] Pr(x_i' \mid x_i, a_i) V^i(x_i') \\
&= \sum_{i=1}^{N} \sum_{x_i'} \left( \sum_{\boldsymbol{x}' - \{x_i'\}} \prod_{i=1}^{N} Pr(x_i' \mid \boldsymbol{x}, \boldsymbol{a}) V^i(x_i') \right) \\
&= \sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x}, \boldsymbol{a}) \left( \sum_{i=1}^{N} V^i(x_i') \right).
\end{aligned} \tag{B.21}$$

Then, we sum problem (B.20) over all users which yields

$$\sum_{i=1}^{N} (V^i(x_i) + \theta_i) = \min_{\boldsymbol{a}} \left\{ \sum_{i=1}^{N} \left( C(x_i, a_i) + \sum_{x_i'} Pr(x_i' \mid x_i, a_i) V^i(x_i') \right) \right\}.$$

We recall that $C(\boldsymbol{x}, \boldsymbol{a}) = \sum_{i=1}^{N} C(x_i, a_i)$ by definition. Then, leveraging problem (B.21), we

obtain

$$\sum_{i=1}^{N} V^i(x_i) + \sum_{i=1}^{N} \theta_i = \min_{\boldsymbol{a} \in \mathcal{A}_N(-1)} \left\{ C(\boldsymbol{x}, \boldsymbol{a}) + \sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x}, \boldsymbol{a}) \left( \sum_{i=1}^{N} V^i(x_i') \right) \right\}.$$

Since the solution to the Bellman equation is unique [51], we must have $\sum_{i=1}^{N} V^i(x_i) = V(\boldsymbol{x})$ and $\sum_{i=1}^{N} \theta_i = \theta$. Then, we can conclude that it is optimal for $\mathcal{M}_N(\lambda, -1)$ if each user adopts its own optimal policy.

## B.11 Proof of Theorem 5

In this proof, we call a policy as $\lambda^*$-optimal if it is optimal for $\mathcal{M}_N(\lambda^*, -1)$. Meanwhile, in Section 3.4.2, we ensure that, for each user, there exists at least one threshold policy that yields a finite expected AoII. Therefore, we can conclude that, for RP, there exists at least one policy that causes the expected AoII and the expected transmission rate to be both finite. Then, according to [50, Lemma 3.10], a policy is optimal for RP if

1. It is $\lambda^*$-optimal;
2. The resulting expected transmission rate is equal to $M$.

We first construct a policy $\phi_{\lambda^*}$ that is $\lambda^*$-optimal. We recall from Proposition 8 that a policy is $\lambda^*$-optimal if it consists of the optimal policies for each $\mathcal{M}_1^i(\lambda^*, -1)$ where $1 \leq i \leq N$. According to Theorem 4, for any $i$, there exist $\boldsymbol{n}_{\lambda_-^*, i}$ and $\boldsymbol{n}_{\lambda_+^*, i}$ that are both optimal for $\mathcal{M}_1^i(\lambda^*, -1)$. Then, we can construct the policy $\phi_{\lambda^*}$ in the following way.

- For user $i$ with $\boldsymbol{n}_{\lambda_-^*, i} = \boldsymbol{n}_{\lambda_+^*, i} \triangleq \boldsymbol{n}_{\lambda^*, i}$, the threshold policy $\boldsymbol{n}_{\lambda^*, i}$ is used. Then, the deterministic policy $\boldsymbol{n}_{\lambda^*, i}$ is optimal for $\mathcal{M}_1^i(\lambda^*, -1)$ and

$$\bar{\rho}^i(\lambda^*) = \bar{\rho}^i(\lambda_-^*) = \bar{\rho}^i(\lambda_+^*).$$

In this case, the choice of $\mu_i$ makes no difference.

- For user $i$ with $\boldsymbol{n}_{\lambda_-^*,i} \neq \boldsymbol{n}_{\lambda_+^*,i}$, the randomized policy $\boldsymbol{n}_{\lambda^*,i}$ as detailed in Theorem 4 is used. Then, for any $\mu_i \in [0,1]$, the randomized policy $\boldsymbol{n}_{\lambda^*,i}$ is optimal for $\mathcal{M}_1^i(\lambda^*, -1)$ and

$$\bar{\rho}^i(\lambda^*) = \mu_i \bar{\rho}^i(\lambda_-^*) + (1-\mu_i)\bar{\rho}^i(\lambda_+^*).$$

Combing the two cases, we conclude that $\phi_{\lambda^*} = [\boldsymbol{n}_{\lambda^*,1}, ..., \boldsymbol{n}_{\lambda^*,N}]$ is $\lambda^*$-optimal under any $\mu_i \in [0,1]$. Hence, as long as the chosen $\mu_i$'s realize $\sum_{i=1}^N \bar{\rho}^i(\lambda^*) = M$, we can conclude that the randomized policy $\phi_{\lambda^*}$ is optimal for RP.

## B.12 Proof of Proposition 10

We notice that $\mathcal{M}_1^i(\lambda^*, -1)$ coincides with the decoupled model studied in Section 3.4.2. Therefore, we can use the results in Section 3.4.2 to prove the properties. Since the users share the same structure, we ignore the user index $i$ for simplicity. According to the definition of $I_x$, we have

$$\begin{aligned} I_x &= \sum_{x'} P_{x,x'}(0)V(x') - \sum_{x'} P_{x,x'}(1)V(x') - \lambda^* \\ &= -\Delta V(x). \end{aligned}$$

Leveraging the results in the proof of Proposition 3, we have the following

- For state $x = (0, \hat{r})$, $I_x = -\lambda^*$.
- For state $x = (s, 0)$ where $s > 0$, $I_x = -\lambda^* - p_e^0(1-2p)\omega$ where $\omega = (1-\gamma)[V(0,0) - V(s+1,0)] + \gamma[V(0,1) - V(s+1,1)] \leq 0$.
- For state $x = (s, 1)$ where $s > 0$, $I_x = -\lambda^* - (1-p_e^1)(1-2p)\omega$.

From the above three cases, we can easily conclude that $I_x \geq -\lambda^*$ and the equality holds when $\hat{r} = p_e^0 = 0$ or $s = 0$. As is proven in Corollary 4, $V(x)$ is non-decreasing in $s$. Hence, we can conclude that $I_x$ is also non-decreasing in $s$. To show that $I_x$ is monotone in $\hat{r}$, we consider two states $x_1 = (s, 1)$ and $x_2 = (s, 0)$. Then, we have

$$I_{x_2} - I_{x_1} = \Delta V(s,1) - \Delta V(s,0) = (1 - p_e^1 - p_e^0)(1-2p)\omega \leq 0.$$

Therefore, we can conclude that $I_x$ is non-decreasing in $\hat{r}$.

# Appendix C: Freshness against Generic Delay

## C.1 Details of State Transition Probability

We first discuss the individual transition of $\Delta$. We divide our discussion into the following cases.

- $\Delta = 0$ and the receiver's estimates are the same at state $s$ and $s'$. In this case, $\Delta' = 0$ when the dynamic source remains in the same state. Otherwise, $\Delta' = 1$.

$$\Delta' = \begin{cases} 0 & w.p.\ 1-p, \\ 1 & w.p.\ p. \end{cases}$$

- $\Delta = 0$ and the receiver's estimates are different at state $s$ and $s'$. In this case, $\Delta' = 0$ when the dynamic source flips the state. Otherwise, $\Delta' = 1$.

$$\Delta' = \begin{cases} 0 & w.p.\ p, \\ 1 & w.p.\ 1-p. \end{cases}$$

- $\Delta > 0$ and the receiver's estimates are the same at state $s$ and $s'$. In this case, $\Delta' = \Delta + 1$

when the dynamic source remains in the same state. Otherwise, $\Delta' = 0$.

$$\Delta' = \begin{cases} 0 & w.p.\ p, \\ \Delta' + 1 & w.p.\ 1-p. \end{cases}$$

- $\Delta > 0$ and the receiver's estimates are different at state $s$ and $s'$. In this case, $\Delta' = \Delta + 1$ when the dynamic source flips the state. Otherwise, $\Delta' = 0$.

$$\Delta' = \begin{cases} 0 & w.p.\ 1-p, \\ \Delta + 1 & w.p.\ p. \end{cases}$$

Hence, in the following, we only state whether the receiver's estimates are the same at state $s$ and $s'$ and omit the rest of the discussion on the transition of $\Delta$. To make the notation clearer, we write $P_{s,s'}(a)$ as $P[(\Delta', i', t') \mid (\Delta, t, i), a]$ and $Pr(T > t+1 \mid t)$ as $q_{t+1}$ in this proof. Then, we distinguish between the following cases.

- $s = (0, 0, -1)$. In this case, the channel is idle. Hence, the feasible action is $a \in \{0, 1\}$. When the transmitter decides not to initiate a new transmission (i.e., $a = 0$), $i' = 0$ and $t' = -1$. Moreover, the receiver's estimate remains the same. Hence,

$$Pr[(0, 0, -1) \mid (0, 0, -1), a = 0] = 1 - p.$$

$$Pr[(1, 0, -1) \mid (0, 0, -1), a = 0] = p.$$

When the transmitter decides to initiate a new transmission (i.e., $a = 1$), the update will be

delivered after a random amount of time $T$. When $T > 1$, which happens with probability $q_1$, the channel will be busy at the next time slot and $t' = 1$ as the transmission starts. Since $\Delta = 0$ when the transmission starts, we know $i' = 0$. Moreover, the receiver's estimate remains the same since no new update will be delivered. Hence,

$$Pr[(0, 1, 0) \mid (0, 0, -1), a = 1] = q_1(1 - p).$$

$$Pr[(1, 1, 0) \mid (0, 0, -1), a = 1] = q_1 p.$$

When $T = 1$, which happens with probability $1 - q_1$, the update will be delivered at the next time slot. Hence, the channel will be available for a new transmission at the next time slot, which means that $t' = 0$ and $i' = -1$. Since $\Delta = 0$ when the transmission starts, the newly arrived update brings no new information to the receiver. Hence, the receiver's estimate remains the same. Hence,

$$Pr[(0, 0, -1) \mid (0, 0, -1), a = 1] = (1 - q_1)(1 - p).$$

$$Pr[(1, 0, -1) \mid (0, 0, -1), a = 1] = (1 - q_1)p.$$

- $s = (0, t, 0)$. In this case, the channel is busy. Hence, the feasible action is $a = 0$. When the update will not arrive at the next time slot, which happens with probability $q_{t+1}$, $i' = i$ since both the transmitting update and the receiver's estimate remain the same. Apparently, $t' = t + 1$ as the transmission continues. Moreover, the receiver's estimate remains the

same. Hence,

$$Pr[(0, t+1, 0) \mid (0, t, 0)] = q_{t+1}(1-p).$$

$$Pr[(1, t+1, 0) \mid (0, t, 0)] = q_{t+1}p.$$

When the update arrives at the next time slot, which happens with probability $1 - q_{t+1}$, $t' = 0$ and $i' = -1$ by definition. Since $i = 0$, the newly arrived update brings no new information to the receiver. Hence, the receiver's estimate remains the same. Hence,

$$Pr[(0, 0, -1) \mid (0, t, 0)] = (1 - q_{t+1})(1-p).$$

$$Pr[(1, 0, -1) \mid (0, t, 0)] = (1 - q_{t+1})p.$$

- $s = (0, t, 1)$. The analysis is very similar to that for $s = (0, t, 0)$ except that when the update arrives, the receiver's estimate is flipped. Hence,

$$Pr[(0, t+1, 1) \mid (0, t, 1)] = q_{t+1}(1-p).$$

$$Pr[(1, t+1, 1) \mid (0, t, 1)] = q_{t+1}p.$$

$$Pr[(0, 0, -1) \mid (0, t, 1)] = (1 - q_{t+1})p.$$

$$Pr[(1, 0, -1) \mid (0, t, 1)] = (1 - q_{t+1})(1-p).$$

- $s = (\Delta, 0, -1)$ where $\Delta > 0$. In this case, the analysis is very similar to that for $s = (0, 0, -1)$, except that the receiver's estimate is incorrect at state $s$, and if the decision is

made to transmit, the transmitted update differs from the receiver's estimate. Therefore, the details are omitted here.

$$Pr[(\Delta+1,0,-1) \mid (\Delta,0,-1), a=0] = 1-p.$$

$$Pr[(0,0,-1) \mid (\Delta,0,-1), a=0] = p.$$

$$Pr[(\Delta+1,1,1) \mid (\Delta,0,-1), a=1] = q_1(1-p).$$

$$Pr[(0,1,1) \mid (\Delta,0,-1), a=1] = q_1 p.$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,0,-1), a=1] = (1-q_1)p.$$

$$Pr[(0,0,-1) \mid (\Delta,0,-1), a=1] = (1-q_1)(1-p).$$

- $s = (\Delta, t, 0)$ where $\Delta > 0$. The analysis is very similar to that for $s = (0, t, 0)$ except that the receiver's estimate is incorrect at state $s$. Hence,

$$Pr[(\Delta+1,t+1,0) \mid (\Delta,t,0)] = q_{t+1}(1-p).$$

$$Pr[(0,t+1,0) \mid (\Delta,t,0)] = q_{t+1}p.$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,t,0)] = (1-q_{t+1})(1-p).$$

$$Pr[(0,0,-1) \mid (\Delta,t,0)] = (1-q_{t+1})p.$$

- $s = (\Delta, t, 1)$ where $\Delta > 0$. The analysis is very similar to that for $s = (\Delta, t, 0)$ except that

the transmitted update differs from the receiver's estimate. Hence,

$$Pr[(\Delta+1, t+1, 1) \mid (\Delta, t, 1)] = q_{t+1}(1-p).$$

$$Pr[(0, t+1, 1) \mid (\Delta, t, 1)] = q_{t+1}p.$$

$$Pr[(\Delta+1, 0, -1) \mid (\Delta, t, 1)] = (1-q_{t+1})p.$$

$$Pr[(0, 0, -1) \mid (\Delta, t, 1)] = (1-q_{t+1})(1-p).$$

Combing the above cases, we fully characterized the state transitions and the corresponding probabilities.

## C.2 Proof of Lemma 4

We recall that $P^t_{\Delta,\Delta'}(1)$ is the probability that action $a$ at state $s = (\Delta, 0, -1)$ will lead to state $s' = (\Delta', 0, -1)$, given that the transmission takes $t$ time slots. With this in mind, we first distinguish between different values of $\Delta$.

- When $\Delta = 0$, the transmitted update is the same as the receiver's estimate. Hence, the receiver's estimate will not change due to receiving the transmitted update. Moreover, we recall that AoII will either increases by one or decreases to zero. Hence, $\Delta' \in \{0, 1, ..., t\}$. Then, we further distinguish our discussion into the following cases.
  - $\Delta' = 0$ happens when the receiver's estimate is correct as a result of receiving the update. Hence, the probability of this happening is $p^{(t)}$.
  - $\Delta' = k \in \{1, ..., t\}$ happens when the receiver's estimate is correct at $(t-k)$th time

slot after the transmission, which happens with probability $p^{(t-k)}$. Then, the estimate remains incorrect for the remainder of the transmission time. This happens when the source first changes state, then remains in the same state throughout the rest of the transmission. Hence, the probability of this happening is $p(1-p)^{k-1}$. Combining together, $\Delta' = k$ happens with probability $p^{(t-k)}p(1-p)^{k-1}$.

Combining together, we have

$$P^t_{0,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ p^{(t-k)}p(1-p)^{k-1} & 1 \leq \Delta' = k \leq t, \\ 0 & otherwise. \end{cases}$$

- When $\Delta > 0$, the transmitted update is different from the receiver's estimate. Hence, the receiver's estimate will flip as a result of receiving the transmitted update. Moreover, we know $\Delta' \in \{0, 1, ..., t-1, \Delta + t\}$. Hence, we further distinguish between the following cases.

  - $\Delta' = 0$ happens in the same case as discussed in the case of $\Delta = 0$. Hence, the estimate is correct with probability $p^{(t)}$.

  - $\Delta' = 1$ happens when the estimate is correct at $(t-1)$th time slot after the transmission, which happens with probability $1 - p^{(t-1)}$. Then, the estimate becomes incorrect as a result of receiving the update. Since the estimate flips upon the arrival of the transmitted update, it happens when the source remains in the same state. Hence, the probability of this happening is $1 - p$. Combing together, $\Delta' = 1$ happens with

probability $(1-p^{(t-1)})(1-p)$.

- $\Delta' = k \in \{2, ..., t-1\}$ happens when the estimate is correct at $(t-k)$th time slot after the transmission, which happens with probability $1-p^{(t-k)}$. Then, the estimate remains incorrect for the remainder of the transmission time. This happens when the dynamic source behaves the following way during the remaining transmission time. The dynamic source should first change state, then remain in the same state, and finally, change state again when the update arrives. This happens with probability $p^2(1-p)^{k-2}$. Hence, $\Delta' = k$ happens with probability $(1-p^{(t-k)})p^2(1-p)^{k-2}$.
- $\Delta' = \Delta + t$ happens when the estimate is incorrect throughout the transmission. Since the estimate will flip when the update is received, this happens when the source stays in the same state until the update arrives. Hence, $\Delta' = \Delta + t$ happens with probability $p(1-p)^{t-1}$.

Combining together, for $\Delta > 0$, we have

$$P^t_{\Delta,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ (1-p^{(t-1)})(1-p) & \Delta' = 1, \\ (1-p^{(t-k)})p^2(1-p)^{k-2} & 2 \leq \Delta' = k \leq t-1, \\ p(1-p)^{t-1} & \Delta' = \Delta + t, \\ 0 & otherwise. \end{cases}$$

By analyzing the above expressions, we can easily conclude that $P^t_{\Delta,\Delta'}(1)$ possesses the following properties.

- $P^t_{\Delta,0}(1)$ and $P^t_{\Delta,\Delta+t}(1)$ are both independent of $\Delta$.
- $P^t_{\Delta,\Delta'}(1)$ is independent of $\Delta$ when $\Delta > 0$ and $0 \leq \Delta' \leq t-1$.
- $P^t_{\Delta,\Delta'}(1) = 0$ when $\Delta' > \Delta + t$ or when $t - 1 < \Delta' < \Delta + t$.

Leveraging the above properties, we can prove the second part of the lemma. The equivalent expression can be obtained easily, so the details are omitted. In the following, we focus on proving the properties of $P_{\Delta,\Delta'}(a)$.

- **property 1**: When $\Delta' = 0$, $P_{\Delta,0}(1) = \sum_{t=1}^{t_{max}} p_t P^t_{\Delta,0}(1)$ for any $\Delta \geq 0$. Since $P^t_{\Delta,0}(1)$ is independent of $\Delta$, property 1 holds in this case. Then, we consider the case of $1 \leq \Delta' \leq t_{max} - 1$ and $\Delta \geq \Delta'$. In this case,

$$P_{\Delta,\Delta'}(1) = \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1),$$

  where $P^t_{\Delta,\Delta'}(1)$ is independent of $\Delta$. Hence, $P_{\Delta,\Delta'}(1)$ is independent of $\Delta$. Combining together, property 1 holds.

- **property 2**: We notice that, when $\Delta' \geq t_{max}$,

$$P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta+t'}(1).$$

  We recall that $P^{t'}_{\Delta,\Delta+t'}(1)$ is independent of $\Delta$. Then, we can conclude that $P_{\Delta,\Delta'}(1)$ depends only on $t'$. Thus, property 2 holds.

- **property 3**: The equivalent expression in corollary indicates that the property holds when

$\Delta' > \Delta + t_{max}$. In the case of $t_{max} - 1 < \Delta' < \Delta + 1$, we have

$$P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1),$$

where $t' \leq 0$. By definition, $P_{\Delta,\Delta'}(1) = 0$. Hence, property 3 holds.

## C.3 Proof of Lemma 5

The proof is similar to that of Lemma 4. We first derive the expressions of $P^{t}_{\Delta,\Delta'}(1)$ and $P^{t^+}_{\Delta,\Delta'}(1)$. To this end, we start with the case of $\Delta = 0$. In this case, the transmitted update is the same as the receiver's estimate. With this in mind, we distinguish between different values of $t$.

- When $1 \leq t < t_{max}$, the update is delivered after $t$ time slot. Hence, $\Delta' \in \{0, 1, ..., t\}$. Then, we further distinguish between different values of $\Delta'$.
  - $\Delta' = 0$ in the case where the receiver's estimate is correct when the update is delivered. Hence, $\Delta' = 0$ happens with probability $p^{(t)}$.
  - $\Delta' = k \in \{1, 2, ..., t\}$ when the receiver's estimate is correct at the $(t-k)$th time slots after the transmission occurs. Then, the source flips the state and remains in the same state for the remainder of the transmission. Hence, $\Delta' = k \in \{1, 2, ..., t\}$ happens with probability $p^{(t-k)}p(1-p)^{k-1}$.
- When $t = t_{max}$, the update either arrives or be discarded. In this case, $\Delta' \in \{0, 1, ..., t_{max}\}$. We recall that the update is the same as the receiver's estimate. Hence, the receiver's estimate will not change in both cases. Consequently, $P^{t_{max}}_{0,\Delta'}(1) = P^{t^+}_{0,\Delta'}(1)$, which can be obtained by setting the $t$ in the above case to $t_{max}$.

Combining together, for each $1 \leq t \leq t_{max}$,

$$P^{t}_{0,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ p^{(t-k)}p(1-p)^{k-1} & 1 \leq \Delta' = k \leq t, \\ 0 & otherwise. \end{cases}$$

$$P^{t^+}_{0,\Delta'}(1) = P^{t_{max}}_{0,\Delta'}(1).$$

Then, we consider the case of $\Delta > 0$. We notice that, in this case, the receiver's estimate will flip upon receiving the update. Then, we distinguish between different values of $t$.

- When $1 \leq t < t_{max}$, the update is delivered after $t$ time slots, and the receiver's estimate will flip. Hence, $\Delta' \in \{0, 1, ..., t-1, \Delta+t\}$. Then, we further distinguish between different values of $\Delta'$.
  - $\Delta' = 0$ in the case where the receiver's estimate is correct when the update is received. Hence, $\Delta' = 0$ happens with probability $p^{(t)}$.
  - $\Delta' = 1$ when the receiver's estimate is correct at $(t-1)$th time slot after transmission starts and becomes incorrect when the update arrives. Hence, $\Delta' = 1$ happens with probability $(1 - p^{(t-1)})(1-p)$.
  - $\Delta' = k \in \{2, 3, ..., t-1\}$ when the receiver's estimate is correct at $(t-k)$th time slot after the transmission starts. Then, the source changes state and remains in the same state. Finally, at the time slot when the update arrives, the source flips state again. Hence, $\Delta' = k \in \{2, 3, ..., t-1\}$ happens with probability $(1 - p^{(t-k)})p^2(1-p)^{k-2}$.
  - $\Delta' = \Delta + t$ when the estimate is incorrect throughout the transmission. We recall

that the receiver's estimate will flip when the update arrives. Hence, $\Delta' = \Delta + t$ when the source remains in the same state until the update arrives, which happens with probability $p(1-p)^{t-1}$.

- When $t = t_{max}$ and the transmitted update is delivered, the receiver's estimate flips. In this case, $\Delta' \in \{0, 1, ..., t_{max} - 1, \Delta + t_{max}\}$. Hence, $P^{t_{max}}_{\Delta,\Delta'}(1)$ can be obtained by setting the $t$ in the above case to $t_{max}$.

- When $t = t_{max}$ and the transmitted update is discarded, the receiver's estimate remains the same. In this case, $\Delta' \in \{0, 1, ..., t_{max} - 1, \Delta + t_{max}\}$. Then, we further divide our discussion into the following cases.

  - $\Delta' = 0$ when the receiver's estimate is correct at the $t_{max}$the time slot after the transmission starts, which happens when the state of the source at the time slot the update is discarded is different from that when the transmission started. Hence, $\Delta' = 0$ happens with probability $1 - p^{(t_{max})}$.

  - $\Delta' = k \in \{1, 2, ..., t_{max} - 1\}$ when the receiver's estimate is correct at $(t_{max} - k)$th time slot after the transmission starts. Then, the source changes state and remains in the same state for the remainder of the transmission. Hence, $\Delta' = k \in \{1, 2, ..., t_{max} - 1\}$ happens with probability $(1 - p^{(t_{max}-k)})p(1-p)^{k-1}$.

  - $\Delta' = \Delta + t_{max}$ when the source remains in the same state throughout the transmission. Combining with the source dynamic, we can conclude that $\Delta' = \Delta + t_{max}$ happens with probability $(1-p)^{t_{max}}$.

Combining together, for $\Delta > 0$ and each $1 \leq t \leq t_{max}$,

$$P^t_{\Delta,\Delta'}(1) = \begin{cases} p^{(t)} & \Delta' = 0, \\ (1-p^{(t-1)})(1-p) & \Delta' = 1, \\ (1-p^{(t-k)})p^2(1-p)^{k-2} & 2 \leq \Delta' = k \leq t-1, \\ p(1-p)^{t-1} & \Delta' = \Delta + t, \\ 0 & otherwise. \end{cases}$$

$$P^{t^+}_{\Delta,\Delta'}(1) = \begin{cases} 1-p^{(t_{max})} & \Delta' = 0, \\ (1-p^{(t_{max}-k)})p(1-p)^{k-1} & 1 \leq \Delta' = k \leq t_{max}-1, \\ (1-p)^{t_{max}} & \Delta' = \Delta + t_{max}, \\ 0 & otherwise. \end{cases}$$

By analyzing the above expressions, we can easily conclude that $P^t_{\Delta,\Delta'}(1)$ and $P^{t^+}_{\Delta,\Delta'}(1)$ possess the following properties.

- $P^t_{\Delta,\Delta+t}(1)$ and $P^{t^+}_{\Delta,\Delta+t_{max}}(1)$ are independent of $\Delta$ when $\Delta > 0$.
- $P^t_{\Delta,\Delta'}(1)$ is independent of $\Delta$ when $\Delta > 0$ and $0 \leq \Delta' \leq t-1$.
- $P^t_{\Delta,\Delta'}(1) = 0$ when $\Delta > 0$ and $t-1 < \Delta' < \Delta + t$.
- $P^{t^+}_{\Delta,\Delta'}(1)$ is independent of $\Delta$ when $\Delta > 0$ and $0 \leq \Delta' \leq t_{max}-1$.
- $P^{t^+}_{\Delta,\Delta'}(1) = 0$ when $\Delta > 0$ and $t_{max}-1 < \Delta' < \Delta + t_{max}$.

Leveraging the properties above, we proceed with proving the second part of the lemma. The

equivalent expression can be obtained easily by analyzing (4.7). Hence, the details are omitted. In the following, we focus on proving the presented properties.

- **property 1**: We notice that, when $0 \leq \Delta' \leq t_{max} - 1$ and $\Delta \geq \max\{1, \Delta'\}$,

$$P_{\Delta,\Delta'}(1) = \sum_{t=\Delta'}^{t_{max}} p_t P^t_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1).$$

  Then, we divide the discussion into the following two cases.

  - $\Delta \geq \max\{1, \Delta'\}$ indicates that $\Delta > 0$ and $\Delta' < \Delta + t_{max}$. Hence, $P^{t^+}_{\Delta,\Delta'}(1)$ is independent of $\Delta$.
  - $\Delta \geq \max\{1, \Delta'\}$ indicates that $\Delta > 0$ and $\Delta' < \Delta + t$. Hence, $P^t_{\Delta,\Delta'}(1)$ is independent of $\Delta$ for any feasible $t$.

  Combining together, we can conclude that property 1 holds.

- **property 2**: We notice that, when $\Delta' \geq t_{max}$,

$$P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1).$$

  Then, we divide the discussion into the following two cases.

  - Since $t' = \Delta' - \Delta$, $P^{t'}_{\Delta,\Delta'}(1) = P^{t'}_{\Delta,\Delta+t'}(1)$. Then, we know that $P^{t'}_{\Delta,\Delta'}(1)$ is independent of $\Delta > 0$ when $t' > 0$ and $P^{t'}_{\Delta,\Delta'}(1) = 0$ when $t' \leq 0$ by definition. Hence, $P^{t'}_{\Delta,\Delta'}(1)$ depends on $t'$.
  - When $\Delta' \geq t_{max}$ and $\Delta' \neq \Delta + t_{max}$, $P^{t^+}_{\Delta,\Delta'}(1) = 0$ for $\Delta > 0$. Also, $P^{t^+}_{\Delta,\Delta'}(1)$ is independent of $\Delta > 0$ when $\Delta' = \Delta + t_{max}$. Hence, $P^{t^+}_{\Delta,\Delta'}(1)$ depends only on $t'$.

Combining together, property 2 holds.

- **property 3**: When $\Delta' > \Delta + t_{max}$, the property holds apparently. When $t_{max} - 1 < \Delta' < \Delta + 1$,

$$P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1),$$

where $t' \leq 0$. Then, by definition, $P^{t'}_{\Delta,\Delta'}(1) = 0$. Moreover, we recall that $t_{max} > 1$, which indicates that $P^{t^+}_{\Delta,\Delta'}(1) = 0$. Hence, property 3 holds.

## C.4 Proof of Theorem 6

We recall that $\pi_\Delta$ satisfies (4.5) and (4.8). Then, plugging in the probabilities yields the following system of linear equations.

$$\begin{aligned} \pi_0 =& (1-p)\pi_0 + p\sum_{i=1}^{\tau-1}\pi_i + \sum_{i=\tau}^{\infty} P_{i,0}(1)\pi_i \\ =& (1-p)\pi_0 + p\sum_{i=1}^{\tau-1}\pi_i + P_{1,0}(1)\sum_{i=\tau}^{\infty}\pi_i. \end{aligned} \tag{C.1}$$

$$\pi_1 = p\pi_0 + \sum_{i=\tau}^{\infty} P_{i,1}(1)\pi_i = p\pi_0 + P_{1,1}(1)\sum_{i=\tau}^{\infty}\pi_i. \tag{C.2}$$

For each $2 \leq \Delta \leq t_{max} - 1$,

$$\pi_\Delta = \begin{cases} (1-p)\pi_{\Delta-1} + P_{\tau,\Delta}(1)\sum_{i=\tau}^{\infty}\pi_i & \Delta - 1 < \tau, \\ \sum_{i=\tau}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\sum_{i=\Delta}^{\infty}\pi_i & \Delta - 1 \geq \tau. \end{cases} \tag{C.3}$$

For each $t_{max} \leq \Delta \leq \omega - 1$,

$$\pi_\Delta = \begin{cases} (1-p)\pi_{\Delta-1} & \Delta - 1 < \tau, \\ \sum_{i=\tau}^{\Delta-1} P_{i,\Delta}(1)\pi_i & \Delta - 1 \geq \tau. \end{cases}$$

For each $\Delta \geq \omega$,

$$\pi_\Delta = \sum_{i=\Delta-t_{max}}^{\Delta-1} P_{i,\Delta}(1)\pi_i. \tag{C.4}$$

$$\sum_{i=0}^{\tau-1} \pi_i + ET \sum_{i=\tau}^{\infty} \pi_i = 1.$$

Note that we can pull the state transition probabilities in (C.1), (C.2), and (C.3) out of the summation due to property 1 in Lemma 4 and Lemma 5. Then, we sum (C.4) over $\Delta$ from $\omega$ to $\infty$.

$$\sum_{i=\omega}^{\infty} \pi_i = \sum_{i=\omega}^{\infty} \sum_{k=i-t_{max}}^{i-1} P_{k,i}(1)\pi_k. \tag{C.5}$$

We delve deep into the right hand side (RHS) of (C.5). To this end, we expand the first summation, which yields

$$RHS = \sum_{k=\tau+1}^{\omega-1} P_{k,\omega}(1)\pi_k + \sum_{k=\tau+2}^{\omega} P_{k,\omega+1}(1)\pi_k + \cdots + \sum_{k=\omega-1}^{\omega+t_{max}-2} P_{k,\omega+t_{max}-1}(1)\pi_k + \sum_{k=\omega}^{\omega+t_{max}-1} P_{k,\omega+t_{max}}(1)\pi_k + \cdots$$

Then, we rearrange the summation.

$$RHS = P_{\tau+1,\omega}(1)\pi_{\tau+1} + \sum_{k=1}^{2} P_{\tau+2,\omega+k-1}(1)\pi_{\tau+2} + \cdots + \sum_{k=1}^{t_{max}} P_{\omega-1,\omega+k-1}(1)\pi_{\omega-1} + \sum_{k=1}^{t_{max}} P_{\omega,\omega+k}(1)\pi_{\omega} + \sum_{k=1}^{t_{max}} P_{\omega+1,\omega+k+1}(1)\pi_{\omega+1} + \cdots$$

Leveraging property 2 in Lemma 4 and Lemma 5, we have

$$RHS = \sum_{i=\tau+1}^{\omega-1} \left( \sum_{k=\tau+1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{max}} \left( P_{\omega,\omega+i}(1) \right) \left( \sum_{k=\omega}^{\infty} \pi_k \right).$$

We define $\Pi \triangleq \sum_{i=\omega}^{\infty} \pi_i$. Then, equation (C.5) becomes the following.

$$\Pi = \sum_{i=\tau+1}^{\omega-1} \left( \sum_{k=\tau+1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{max}} \left( P_{\omega,\omega+i}(1) \right) \Pi. \tag{C.6}$$

Finally, replacing (C.4) with (C.6) and applying the definition of $\Pi$ yield a system of linear equations with finite size as presented in the theorem.

## C.5 Proof of Corollary 6

We start with $\tau = 0$. In this case, $\omega = t_{max} + 1$ and the system of linear equations becomes to the following.

$$\pi_\Delta = \sum_{i=0}^{\infty} P_{i,\Delta}(1)\pi_i = \begin{cases} P_{0,0}(1)\pi_0 + P_{1,0}(1) \sum_{i=1}^{\infty} \pi_i & \Delta = 0, \\ \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \sum_{i=\Delta}^{\infty} \pi_i & 1 \leq \Delta \leq t_{max}. \end{cases} \tag{C.7}$$

$$\Pi = \sum_{i=1}^{t_{max}} \left( \sum_{k=1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{max}} P_{t_{max}+1,t_{max}+1+i}(1)\Pi. \tag{C.8}$$

$$ET \sum_{i=0}^{\infty} \pi_i = 1. \tag{C.9}$$

We first combine (C.7) and (C.9), which yields

$$\pi_\Delta = \begin{cases} P_{0,0}(1)\pi_0 + P_{1,0}(1)\left(\frac{1}{ET} - \pi_0\right) & \Delta = 0, \\ \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\left(\frac{1}{ET} - \sum_{i=0}^{\Delta-1}\pi_i\right) & 1 \le \Delta \le t_{max}. \end{cases}$$

Then, we have

$$\pi_0 = \frac{P_{1,0}(1)}{ET[1 - P_{0,0}(1) + P_{1,0}(1)]}.$$

According to (C.8), we obtain

$$\Pi = \frac{\sum_{i=1}^{t_{max}} \left( \sum_{k=1}^{i} P_{i,t_{max}+k}(1) \right) \pi_i}{1 - \sum_{i=1}^{t_{max}} P_{t_{max}+1,t_{max}+1+i}(1)}.$$

Then, we consider the case of $\tau = 1$. In this case, $\omega = t_{max} + 2$ and the system of linear equations reduces to the following.

$$\pi_0 = (1-p)\pi_0 + P_{1,0}(1) \sum_{i=1}^{\infty} \pi_i. \tag{C.10}$$

$$\pi_1 = p\pi_0 + P_{1,1}(1) \sum_{i=1}^{\infty} \pi_i.$$

$$\pi_\Delta = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\sum_{i=\Delta}^{\infty}\pi_i \quad 2 \le \Delta \le t_{max}-1.$$

$$\pi_\Delta = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i \quad t_{max} \le \Delta \le t_{max}+1. \tag{C.11}$$

$$\begin{aligned}\Pi = &\sum_{i=2}^{t_{max}+1}\left(\sum_{k=2}^{i} P_{i,t_{max}+k}(1)\right)\pi_i + \\ &\sum_{i=1}^{t_{max}} P_{t_{max}+2,t_{max}+2+i}(1)\Pi.\end{aligned} \tag{C.12}$$

$$\pi_0 + ET\sum_{i=1}^{\infty}\pi_i = 1. \tag{C.13}$$

We first combine (C.10) and (C.13), which yields

$$\pi_0 = (1-p)\pi_0 + P_{1,0}(1)\left(\frac{1-\pi_0}{ET}\right).$$

Hence, we have

$$\pi_0 = \frac{P_{1,0}(1)}{pET + P_{1,0}(1)}.$$

Similarly,

$$\pi_1 = \frac{pP_{1,0}(1) + pP_{1,1}(1)}{pET + P_{1,0}(1)}.$$

For each $2 \le \Delta \le t_{max}-1$,

$$\pi_\Delta = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\left(\frac{1-\pi_0}{ET} - \sum_{i=1}^{\Delta-1}\pi_i\right). \tag{C.14}$$

According to the property 3 in Lemma 4 and Lemma 5, we know that $P_{\Delta,\Delta}(1) = 0$ when $t_{max} \le \Delta \le t_{max}+1$. Hence, we can combine (C.11) and (C.14), for each $2 \le \Delta \le t_{max}+1$, which

yields

$$\pi_\Delta = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1)\left(\frac{1-\pi_0}{ET} - \sum_{i=1}^{\Delta-1}\pi_i\right).$$

Finally, according to (C.12), we obtain

$$\Pi = \frac{\sum_{i=2}^{t_{max}+1}\left(\sum_{k=2}^{i} P_{i,t_{max}+k}(1)\right)\pi_i}{1-\sum_{i=1}^{t_{max}} P_{t_{max}+2,t_{max}+2+i}(1)}.$$

## C.6 Proof of Lemma 6

We recall that $C^k(\Delta)$ is defined as the expected AoII $k$ time slots after the transmission starts at state $(\Delta, 0, -1)$, given that the transmission is still in progress. With this in mind, we start with the case of $\Delta = 0$. As AoII either increases by one or decreases to zero, we know $C^k(0) \in \{0, ..., k\}$. Then, we distinguish between the following cases.

- $C^k(0) = 0$ when the receiver's estimate is correct $k$ time slots after the transmission starts. Since $\Delta = 0$, we can easily conclude that $C^k(0) = 0$ happens with probability $p^{(k)}$.
- $C^k(0) = h$, where $1 \leq h \leq k$, happens when the receiver's estimate is correct at the $(k-h)$th time slot after the transmission starts, then, the source flips the state and stays in the same state for the remaining $h-1$ time slots. Hence, $C^k(0) = h$, where $1 \leq h \leq k$, happens with probability $p^{(k-h)}p(1-p)^{h-1}$.

Combining together, we obtain

$$C^k(0) = \sum_{h=1}^{k} hp^{(k-h)}p(1-p)^{h-1}.$$

Then, we consider the case of $\Delta > 0$. In this case, the transmission starts when the receiver's estimate is incorrect and $C^k(\Delta) \in \{0, 1, ...k-1, \Delta + k\}$. Then, we distinguish between the following cases.

- $C^k(\Delta) = 0$ when the receiver's estimate is correct at the $k$th time slot after the transmission starts, which happens with probability $(1 - p^{(k)})$.
- $C^k(\Delta) = h$, where $h \in \{1, 2, ..., k-1\}$, happens when the receiver's estimate is correct at the $(k-h)$th slot after the transmission starts. Then, the source flips the state and stays in the same state for the remaining $h-1$ time slots. Hence, $C^k(\Delta) = h$, where $h \in \{1, 2, ..., k-1\}$, happens with probability $(1 - p^{(k-h)})p(1-p)^{h-1}$.
- $C^k(\Delta) = \Delta + k$ when the estimate at the receiver side is always wrong for $k$ time slots after the transmission starts. Since $\Delta > 0$ and the receiver's estimate will not change, $C^k(\Delta) = \Delta + k$ happens with probability $(1-p)^k$.

Combining together, for $\Delta > 0$, we obtain

$$C^k(\Delta) = \sum_{h=1}^{k-1} h(1 - p^{(k-h)})p(1-p)^{h-1} + (\Delta + k)(1-p)^k.$$

## C.7 Proof of Theorem 7

We recall that when $\tau = \infty$, the transmitter will never initiate any transmissions. Hence, the receiver's estimate will never change. Without loss of generality, we assume the receiver's estimate $\hat{X}_k = 0$ for all $k$. The first step in calculating the expected AoII achieved by the threshold policy with $\tau = \infty$ is to calculate the stationary distribution of the induced DTMC. We know that

$\pi_\Delta$ satisfies the following equations.

$$\pi_0 = (1-p)\pi_0 + p\sum_{i=1}^{\infty}\pi_i. \tag{C.15}$$

$$\pi_1 = p\pi_0.$$

$$\pi_\Delta = (1-p)\pi_{\Delta-1} \quad \Delta \geq 2.$$

$$\sum_{i=0}^{\infty}\pi_i = 1. \tag{C.16}$$

Combining (C.15) and (C.16) yields

$$\pi_0 = (1-p)\pi_0 + p(1-\pi_0).$$

Hence, $\pi_0 = \frac{1}{2}$. Then, we can get

$$\pi_1 = \frac{p}{2},$$

$$\pi_\Delta = (1-p)^{\Delta-1}\pi_1 = \frac{p(1-p)^{\Delta-1}}{2} \quad \Delta \geq 2.$$

Combining together, we have

$$\pi_0 = \frac{1}{2}, \quad \pi_\Delta = \frac{p(1-p)^{\Delta-1}}{2} \quad \Delta \geq 1.$$

Since the transmitter will never make any transmission attempts, the cost for being at state $(\Delta, 0, -1)$ is nothing but $\Delta$ itself. Hence, the expected AoII is

$$\bar{\Delta}_\infty = \sum_{\Delta=1}^{\infty} \Delta \frac{p(1-p)^{\Delta-1}}{2} = \frac{1}{2p}.$$

## C.8 Proof of Theorem 8

We recall that, for $\Delta \geq \omega$, $\pi_\Delta$ satisfies

$$\pi_\Delta = \sum_{i=\Delta-t_{max}}^{\Delta-1} P_{i,\Delta}(1)\pi_i = \sum_{i=1}^{t_{max}} P_{i-t_{max}+\Delta-1,\Delta}(1)\pi_{i-t_{max}+\Delta-1} \quad \Delta \geq \omega.$$

We first focus on the system under **Assumption 1**. We know from by Lemma 4 that $P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1)$ where $t' = \Delta' - \Delta$ when $\Delta' \geq \omega$. Hence, for each $\Delta \geq \omega$,

$$\pi_\Delta = \sum_{i=1}^{t_{max}} p_{t_{max}+1-i} P^{t_{max}+1-i}_{i-t_{max}+\Delta-1,\Delta}(1)\pi_{i-t_{max}+\Delta-1}.$$

Renaming the variables yields

$$\pi_\Delta = \sum_{t=1}^{t_{max}} p_t P^t_{\Delta-t,\Delta}(1)\pi_{\Delta-t} \quad \Delta \geq \omega.$$

To proceed, we define, for each $1 \leq t \leq t_{max}$,

$$\pi_{\Delta,t} \triangleq p_t P^t_{\Delta-t,\Delta}(1)\pi_{\Delta-t} \quad \Delta \geq \omega. \tag{C.17}$$

Note that $\sum_{t=1}^{t_{max}} \pi_{\Delta,t} = \pi_\Delta$. Then, for a given $1 \leq t \leq t_{max}$, we multiple both side of (C.17) by $C(\Delta - t, 1)$ and sum over $\Delta$ from $\omega$ to $\infty$. Hence, we have

$$\sum_{i=\omega}^{\infty} C(i-t,1)\pi_{i,t} = \sum_{i=\omega}^{\infty} C(i-t,1)p_t P_{i-t,i}^t(1)\pi_{i-t}. \tag{C.18}$$

We define $\Delta_t' \triangleq C(\Delta,1) - C(\Delta - t, 1)$ where $\Delta > t$. Then, according to (4.10), we have

$$\Delta_t' = \sum_{i=1}^{t_{max}} p_i \Big( C^i(\Delta,1) - C^i(\Delta - t, 1) \Big).$$

According to Lemma 6, we have

$$C^i(\Delta - t, 1) = \Delta - t + \sum_{h=1}^{i-1} \left( \sum_{k=1}^{h-1} k(1-p^{(h-k)})p(1-p)^{k-1} + (\Delta - t + h)(1-p)^h \right).$$

$$C^i(\Delta, 1) = \Delta + \sum_{h=1}^{i-1} \left( \sum_{k=1}^{h-1} k(1-p^{(h-k)})p(1-p)^{k-1} + (\Delta + h)(1-p)^h \right).$$

Subtracting the two equations yields

$$C^i(\Delta,1) - C^i(\Delta - t, 1) = t + \sum_{h=1}^{i-1} \Big( t(1-p)^h \Big) = \frac{t - t(1-p)^i}{p}.$$

Then, we have

$$\Delta_t' = \sum_{i=1}^{t_{max}} p_i \left( \frac{t - t(1-p)^i}{p} \right).$$

We notice that $\Delta_t'$ is independent of $\Delta$ when $\Delta > t$. Hence, (C.18) can be rewritten as

$$\sum_{i=\omega}^{\infty} \Big( C(i,1) - \Delta_t' \Big) \pi_{i,t} = \sum_{i=\omega-t}^{\infty} C(i,1) p_t P_{i,i+t}^t(1) \pi_i.$$

Then, we define $\Pi_t \triangleq \sum_{i=\omega}^{\infty} \pi_{i,t}$ and $\Sigma_t \triangleq \sum_{i=\omega}^{\infty} C(i,1)\pi_{i,t}$. We notice that $P^t_{\Delta,\Delta+t}(1)$ is independent of $\Delta$ when $\Delta > 0$. Hence, we obtain

$$\sum_{i=\omega}^{\infty} C(i,1)\pi_{i,t} - \Delta'_t \sum_{i=\omega}^{\infty} \pi_{i,t} = p_t P^t_{1,1+t}(1) \sum_{i=\omega-t}^{\infty} C(i,1)\pi_i.$$

Plugging in the definitions yields

$$\Sigma_t - \Delta'_t \Pi_t = p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C(i,1)\pi_i + \Sigma \right).$$

Summing the above equation over $t$ from 1 to $t_{max}$ yields

$$\sum_{t=1}^{t_{max}} \left( \Sigma_t - \Delta'_t \Pi_t \right) = \sum_{t=1}^{t_{max}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C(i,1)\pi_i + \Sigma \right) \right].$$

Rearranging the above equation yields

$$\Sigma - \sum_{t=1}^{t_{max}} \Delta'_t \Pi_t = \sum_{t=1}^{t_{max}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C(i,1)\pi_i \right) \right] + \sum_{t=1}^{t_{max}} \left( p_t P^t_{1,1+t}(1) \right) \Sigma.$$

Hence, the closed-form expression of $\Sigma$ is

$$\Sigma = \frac{\sum_{t=1}^{t_{max}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C(i,1)\pi_i \right) + \Delta'_t \Pi_t \right]}{1 - \sum_{t=1}^{t_{max}} \left( p_t P^t_{1,1+t}(1) \right)}.$$

In the following, we calculate $\Pi_t$. Combining the definition of $\Pi_t$ with (C.17), we have

$$\Pi_t \triangleq \sum_{i=\omega}^{\infty} \pi_{i,t} = \sum_{i=\omega}^{\infty} \left( p_t P^t_{i-t,i}(1) \pi_{i-t} \right) = \sum_{i=\omega-t}^{\infty} \left( p_t P^t_{i,i+t}(1) \pi_i \right).$$

Since $P^t_{\Delta,\Delta+t}(1)$ is independent of $\Delta$ when $\Delta > 0$, we have

$$\Pi_t = p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} \pi_i + \Pi \right).$$

Combining together, we recover the results for **Assumption 1** as presented in the first part of the theorem.

In the sequel, we focus on **Assumption 2**. To this end, we follow similar steps as detailed above. We recall from Lemma 5, $P_{\Delta,\Delta'}(1) = p_{t'} P^{t'}_{\Delta,\Delta'}(1) + p_{t^+} P^{t^+}_{\Delta,\Delta'}(1)$ where $t' = \Delta' - \Delta$ when $\Delta' \geq \omega$. Then, for each $\Delta \geq \omega$,

$$\pi_\Delta = \sum_{i=1}^{t_{max}} \left( p_{t_{max}+1-i} P^{t_{max}+1-i}_{\Delta-t_{max}+i-1,\Delta}(1) + p_{t^+} P^{t^+}_{\Delta-t_{max}+i-1,\Delta}(1) \right) \pi_{\Delta-t_{max}-1+i}.$$

Renaming the variables yields

$$\pi_\Delta = \sum_{t=1}^{t_{max}} \left( p_t P^t_{\Delta-t,\Delta}(1) + p_{t^+} P^{t^+}_{\Delta-t,\Delta}(1) \right) \pi_{\Delta-t} = \sum_{t=1}^{t_{max}} \Upsilon(\Delta,t) \pi_{\Delta-t} \quad \Delta \geq \omega,$$

where $\Upsilon(\Delta,t) \triangleq p_t P^t_{\Delta-t,\Delta}(1) + p_{t^+} P^{t^+}_{\Delta-t,\Delta}(1)$. We notice that $\Upsilon(\Delta,t)$ is independent of $\Delta$ when $\Delta \geq \omega$. To proceed, we define, for each $1 \leq t \leq t_{max}$,

$$\pi_{\Delta,t} \triangleq \Upsilon(\Delta,t) \pi_{\Delta-t} \quad \Delta \geq \omega.$$

Note that $\sum_{t=1}^{t_{max}} \pi_{\Delta,t} = \pi_{\Delta}$. Then, for a given $1 \leq t \leq t_{max}$, we have

$$\sum_{i=\omega}^{\infty} C(i-t,1)\pi_{i,t} = \sum_{i=\omega}^{\infty} C(i-t,1)\Upsilon(i,t)\pi_{i-t}. \tag{C.19}$$

We define $\Delta'_t \triangleq C(\Delta,1) - C(\Delta-t,1)$ where $\Delta > t$. Then, according to (4.11), we have

$$\Delta'_t = \sum_{i=1}^{t_{max}} p_i \Big( C^i(\Delta,1) - C^i(\Delta-t,1) \Big) + p_{t^+} \Big( C^{t_{max}}(\Delta,1) - C^{t_{max}}(\Delta-t,1) \Big).$$

By Lemma 6, we have

$$C^i(\Delta-t,1) = \Delta - t + \sum_{h=1}^{i-1} \left( \sum_{k=1}^{h-1} k(1-p^{(h-k)})p(1-p)^{k-1} + (\Delta - t + h)(1-p)^h \right).$$

$$C^i(\Delta,1) = \Delta + \sum_{h=1}^{i-1} \left( \sum_{k=1}^{h-1} k(1-p^{(h-k)})p(1-p)^{k-1} + (\Delta + h)(1-p)^h \right).$$

Subtracting the two equations yields

$$C^i(\Delta,1) - C^i(\Delta-t,1) = t + \sum_{h=1}^{i-1} \Big( t(1-p)^h \Big) = \frac{t - t(1-p)^i}{p}.$$

Then, for each $1 \leq t \leq t_{max}$, we have

$$\Delta'_t = \sum_{i=1}^{t_{max}} p_i \left( \frac{t - t(1-p)^i}{p} \right) + p_{t^+} \left( \frac{t - t(1-p)^{t_{max}}}{p} \right).$$

We notice that $\Delta'_t = C(\Delta,1) - C(\Delta-t,1)$ is independent of $\Delta$ when $\Delta > t$. Hence, equation

(C.19) can be written as

$$\sum_{i=\omega}^{\infty}\left(C(i,1)-\Delta_t'\right)\pi_{i,t}=\sum_{i=\omega-t}^{\infty}C(i,1)\Upsilon(i+t,t)\pi_i.$$

Then, we define $\Pi_t \triangleq \sum_{i=\omega}^{\infty}\pi_{i,t}$ and $\Sigma_t \triangleq \sum_{i=\omega}^{\infty}C(i,1)\pi_{i,t}$. We recall that $\Upsilon(\Delta,t)$ is independent of $\Delta$ when $\Delta \geq \omega$. Hence, plugging in the definitions yields

$$\Sigma_t-\Delta_t'\Pi_t=\sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)C(i,1)\pi_i+\Upsilon(\omega+t,t)\Sigma.$$

Summing the above equation over $t$ from 1 to $t_{max}$ yields

$$\sum_{t=1}^{t_{max}}\left(\Sigma_t-\Delta_t'\Pi_t\right)=\sum_{t=1}^{t_{max}}\left(\sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)C(i,1)\pi_i+\Upsilon(\omega+t,t)\Sigma\right).$$

Rearranging the above equation yields

$$\Sigma-\sum_{t=1}^{t_{max}}\Delta_t'\Pi_t=\sum_{t=1}^{t_{max}}\left(\sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)C(i,1)\pi_i\right)+\sum_{t=1}^{t_{max}}\Upsilon(\omega+t,t)\Sigma.$$

Then, the closed-form expression of $\Sigma$ is

$$\Sigma=\frac{\sum_{t=1}^{t_{max}}\left[\left(\sum_{i=\omega-t}^{\omega-1}\Upsilon(i+t,t)C(i,1)\pi_i\right)+\Delta_t'\Pi_t\right]}{1-\sum_{t=1}^{t_{max}}\Upsilon(\omega+t,t)}.$$

In the following, we calculate $\Pi_t$. We have

$$\Pi_t \triangleq \sum_{i=\omega}^{\infty} \pi_{i,t} = \sum_{i=\omega}^{\infty} \Upsilon(i,t)\pi_{i-t} = \sum_{i=\omega-t}^{\infty} \Upsilon(i+t,t)\pi_i.$$

Since $\Upsilon(\Delta, t)$ is independent of $\Delta$ if $\Delta \geq \omega$, we have

$$\Pi_t = \sum_{i=\omega-t}^{\omega-1} \Upsilon(i+t,t)\pi_i + \Upsilon(\omega+t,t)\Pi \quad 1 \leq t \leq t_{max}.$$

Combining together, we recover the results for the system under **Assumption 2** as presented in the second half of the theorem.

## C.9 Proof of Lemma 8

Leveraging Lemma 7, the result can be proved using mathematical induction. To start with, we initialize $V_{\gamma,0}(s) = 0$ for all $s$. Hence, the base case (i.e., $\nu = 0$) is true. Then, we assume the monotonicity holds at iteration $\nu$, and check whether the monotonicity still holds at iteration $\nu + 1$. We recall that the estimated value function $V_{\gamma,\nu+1}(s)$ is updated using (4.13). Hence, the structural property is embedded in the state transition probability $P_{s,s'}(a)$. Using the state transition probabilities in Appendix C.1, equation (4.13) for the state with $\Delta > 0$ can be written

as

$$V_{\gamma,\nu+1}(\Delta,t,i) = \min_{a\in\{0,1\}}\left\{\Delta+\gamma\sum_{\Delta',t',i'} Pr[(\Delta',t',i') \mid (\Delta,t,i),a]V_{\gamma,\nu}(\Delta',t',i')\right\}$$
$$= \min_{a\in\{0,1\}}\left\{\Delta+\gamma\sum_{t',i'}\Big[Pr[(\Delta+1,t',i') \mid (\Delta,t,i),a]V_{\gamma,\nu}(\Delta+1,t',i')+ Pr[(0,t',i') \mid (\Delta,t,i),a]V_{\gamma,\nu}(0,t',i')\Big]\right\}.$$

Moreover, for any $\Delta_1 > 0$ and $\Delta_2 > 0$,

$$Pr[(\Delta_1+1,t',i') \mid (\Delta_1,t,i),a] = Pr[(\Delta_2+1,t',i') \mid (\Delta_2,t,i),a].$$

$$Pr[(0,t',i') \mid (\Delta_1,t,i),a] = Pr[(0,t',i') \mid (\Delta_2,t,i),a].$$

Let $V^a_{\gamma,\nu+1}(\Delta,t,i)$ be the resulting $V_{\gamma,\nu+1}(\Delta,t,i)$ when action $a$ is chosen. Then, we have

$$V^a_{\gamma,\nu+1}(\Delta+1,t,i) - V^a_{\gamma,\nu+1}(\Delta,t,i) =$$
$$1+\gamma\sum_{t',i'}\left\{Pr[(\Delta+1,t',i') \mid (\Delta,t,i),a]\Big(V_{\gamma,\nu}(\Delta+2,t',i') - V_{\gamma,\nu}(\Delta+1,t',i')\Big)\right\}.$$

Combining with the assumption for iteration $\nu$, we can easily conclude that $V^a_{\gamma,\nu+1}(\Delta+1,t,i) \geq V^a_{\gamma,\nu+1}(\Delta,t,i)$ when $\Delta > 0$ for $a \in \{0,1\}$. Since, $V_{\gamma,\nu+1}(\Delta,t,i) = \min_{a\in\{0,1\}}\{V^a_{\gamma,\nu+1}(\Delta,t,i)\}$, we know that $V_{\gamma,\nu+1}(\Delta+1,t,i) \geq V_{\gamma,\nu+1}(\Delta,t,i)$ when $\Delta > 0$. Finally, by mathematical induction, we can conclude that Lemma 8 is true.

## C.10 Proof of Theorem 9

We first define $h_\gamma(s) \triangleq V_\gamma(s) - V_\gamma(s^{ref})$ as the relative value function and choose the reference state $s^{ref} = (0, 0, -1)$. For simplicity, we abbreviate the reference state $s^{ref}$ as 0 for the remainder of this proof. Then, we show that $\mathcal{M}$ verifies the two conditions given in [80]. As a result, the existence of the optimal policy is guaranteed.

1. *There exists a non-negative $N$ such that $-N \leq h_\gamma(s)$ for all $s$ and $\gamma$*: Leveraging Lemma 8, we can easily conclude that $h_\gamma(s)$ is also non-decreasing in $\Delta$ when $\Delta > 0$. In the following, we consider the policy $\phi$ being the threshold policy with $\tau = 0$. Then, we know that policy $\phi$ induces an irreducible ergodic Markov chain and the expected cost is finite. Let $c_{s,s'}(\phi)$ be the expected cost of a first passage from $s \in \mathcal{S}$ to $s' \in \mathcal{S}$ when policy $\phi$ is adopted. Then, by [80, Proposition 4], we know that $c_{s,0}(\phi)$ is finite. Meanwhile, $h_\gamma(s) \leq c_{s,0}(\phi)$ as is given in the proof of [80, Proposition 5]. Hence, we have $V_\gamma(0) - V_\gamma(s) \leq c_{0,s}(\phi)$ and $V_\gamma(0) - V_\gamma(s) = -h_\gamma(s)$. Hence, we have $h_\gamma(s) \geq -c_{0,s}(\phi)$. Combining with the monotonicity proved in Lemma 8, we can choose $-N = \min_{s \in G}\{c_{0,s}(\phi)\}$, where $G = \{s : \Delta \in \{0, 1\}\}$. This condition indicates that [80, Assumption 2] holds.

2. *$\mathcal{M}$ has a stationary policy $\phi$ inducing an irreducible, ergodic Markov chain. Moreover, the resulting expected cost is finite*: We consider the policy $\phi$ being the threshold policy with $\tau = 0$. Then, according to Section 4.4, it induces an irreducible, ergodic Markov chain and the resulting expected cost is finite. Then, according to [80, Proposition 5], we can conclude that [80, Assumptions 1 and 3] hold.

As the two conditions are verified, the existence of the optimal policy is guaranteed by [80,

Theorem]. Moreover, the minimum expected cost is independent of the initial state.

## C.11 Proof of Theorem 10

We inherit the definitions and notations introduced in Section 4.5.1. We further define $v_{\gamma,n}(\cdot)$ as the minimum expected $\gamma$-discounted cost for operating the system from time 0 to time $n-1$. It is known that $\lim_{n\to\infty} v_{\gamma,n}(s) = V_\gamma(s)$, for all $s \in \mathcal{S}$. We also define the expected cost under policy $\phi$ as

$$J_\phi(s) = \limsup_{K\to\infty} \frac{1}{K} \mathbb{E}_\phi \left( \sum_{k=0}^{K-1} C(s_k) \mid s_0 \right),$$

and $J(s) \triangleq \inf_\phi J_\phi(s)$ is the best that can be achieved. $V_{\phi,\gamma}^{(m)}(s)$, $V_\gamma^{(m)}(s)$, $v_{\gamma,n}^{(m)}(s)$, $J_\phi^{(m)}(s)$, $J^{(m)}(s)$, and $h_\gamma^{(m)}(s)$ are defined analogously for $\mathcal{M}^{(m)}$. With the above definitions in mind, we show that our system verifies the two assumptions given in [52].

- *Assumption 1*: There exists a non-negative (finite) constant $L$, a non-negative (finite) function $M(\cdot)$ on $\mathcal{S}$, and constants $m_0$ and $\gamma_0 \in [0,1)$, such that $-L \le h_\gamma^{(m)}(s) \le M(s)$, for $s \in \mathcal{S}^{(m)}$, $m \ge m_0$, and $\gamma \in (\gamma_0, 1)$: $L$ can be chosen in the same way as presented in the proof of Theorem 10. More precisely, $-L = \min_{s\in G}\{h_\gamma^{(m)}(s)\}$, where $G = \{s : \Delta \in \{0,1\}\}$. Let $c_{s,0}(\phi)$ be the expected cost of a first passage from $s \in \mathcal{S}$ to the reference state 0 when policy $\phi$ is adopted and $c_{x,0}^{(m)}(\phi)$ is defined analogously for $\mathcal{M}^{(m)}$. In the following, we consider the policy $\phi$ being the threshold policy with $\tau = \infty$. We recall from Section 4.4 that the policy $\phi$ induces an irreducible ergodic Markov chain, and the expected cost is finite. Hence, $h_\gamma^{(m)}(s) \le c_{s,0}^{(m)}(\phi)$ by [80, Proposition 5] and $c_{x,0}(\phi)$ is finite by [80, Proposition 4]. We also know from the proof of [52, Corollary 4.3] that

$c_{s,0}(\phi)$ satisfies the following equation.

$$c_{s,0}(\phi) = C(s) + \sum_{s' \in \mathcal{S} - \{0\}} P^{\phi}_{ss'} c_{s',0}(\phi), \tag{C.20}$$

where $P^{\phi}_{ss'}$ is the state transition probability from state $s$ to $s'$ under policy $\phi$ for $\mathcal{M}$. $P^{(m),\phi}_{ss'}$ is defined analogously for $\mathcal{M}^{(m)}$. We can verify in a similar way to the proof of Lemma 8 that $c_{s,0}(\phi)$ is non-decreasing in $\Delta > 0$. The proof is omitted here for the sake of space. Then, we have

$$\begin{aligned}
\sum_{y \in \mathcal{S}^{(m)}_{-1}} P^{(m),\phi}_{sy} c_{y,0}(\phi) &= \sum_{y \in \mathcal{S}^{(m)}_{-1}} P^{\phi}_{sy} c_{y,0}(\phi) + \sum_{y \in \mathcal{S}^{(m)}_{-1}} \left( \sum_{z \in \mathcal{S} \setminus \mathcal{S}^{(m)}} P^{\phi}_{sz} q_z(y) \right) c_{y,0}(\phi) \\
&= \sum_{y \in \mathcal{S}^{(m)}_{-1}} P^{\phi}_{sy} c_{y,0}(\phi) + \sum_{z \in \mathcal{S} \setminus \mathcal{S}^{(m)}} P^{\phi}_{sz} \left( \sum_{y \in \mathcal{S}^{(m)}_{-1}} q_z(y) c_{y,0}(\phi) \right) \\
&\leq \sum_{y \in \mathcal{S}^{(m)}_{-1}} P^{\phi}_{sy} c_{y,0}(\phi) + \sum_{z \in \mathcal{S} \setminus \mathcal{S}^{(m)}} P^{\phi}_{sz} c_{z,0}(\phi) \\
&= \sum_{y \in \mathcal{S} \setminus \{0\}} P^{\phi}_{sy} c_{y,0}(\phi),
\end{aligned} \tag{C.21}$$

where $\mathcal{S}^{(m)}_{-1} = \mathcal{S}^{(m)} \setminus \{0\}$ and $q_{s'}(s) = \mathbb{1}\{t' = t; i' = i\}$, which is an indicator function with value 1 when the transitions to state $s'$ are redirected to state $s$. Otherwise, $q_{s'}(s) = 0$. Moreover, $\sum_{s \in \mathcal{S}^{(m)}_{-1}} q_{s'}(s) = 1$. Applying (C.21) to (C.20) yields

$$c_{s,0}(\phi) \geq C(s) + \sum_{y \in \mathcal{S}^{(m)} - \{0\}} P^{(m),\phi}_{sy} c_{y,0}(\phi).$$

Bearing in mind that $c_{s,0}^{(m)}(\phi)$ satisfies the following.

$$c_{s,0}^{(m)}(\phi) = C(s) + \sum_{y \in \mathcal{S}^{(m)} - \{0\}} P_{sy}^{(m),\phi} c_{y,0}^{(m)}(\phi).$$

Hence, we can conclude that $c_{s,0}^{(m)}(\phi) \leq c_{s,0}(\phi)$. Then, we can choose $M(s) = c_{s,0}(\phi) < \infty$.

- *Assumption 2*: $\limsup_{m\to\infty} J^{(m)} \triangleq J^* < \infty$ and $J^* \leq J(s)$ for all $s \in \mathcal{S}$: We first show that [52, Proposition 5.1] is true. Since we redistribute the transitions in a way such that, for each $s' \in \mathcal{S} \setminus \mathcal{S}^{(m)}$,

$$\sum_{y \in \mathcal{S}^{(m)}} q_{s'}(y) v_{\gamma,n}(y) = v_{\gamma,n}(s),$$

where $s = (m, t', i')$. Hence, we only need to verify that, for each $s' \in \mathcal{S} \setminus \mathcal{S}^{(m)}$ and $s = (m, t', i')$,

$$v_{\gamma,n}(s) \leq v_{\gamma,n}(s'). \tag{C.22}$$

To this end, we notice that $v_{\gamma,n}(s)$ satisfies the following inductive form [52].

$$v_{\gamma,n+1}(s) = \min_a \left\{ C(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}(a) v_{\gamma,n}(s') \right\}.$$

By following similar steps to those in the proof of Lemma 8, we can prove the monotonicity of $v_{\gamma,n}(s)$ for $\Delta > 0$ and $n \geq 0$. The proof is omitted for the sake of space. Hence, (C.22) is true since $\Delta' > m > 0$. Apparently, $J(s)$ is finite for $s \in \mathcal{S}$. Then, according to [52, Corollary 5.2], assumption 2 is true.

Consequently, by [52, Theorem 2.2], we know

- There exists an average cost optimal stationary policy for $\mathcal{M}^{(m)}$.
- Any limit point of the sequence of optimal policies for $\mathcal{M}^{(m)}$ is optimal for $\mathcal{M}$.

## C.12 Proof of Theorem 11

The proof is based on [81, pp. 42-43]. We consider a generic MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C})$. Let $C(s, A)$ be the instant cost for being at state $s \in \mathcal{S}$ under policy $A$. We also define $P^A_{s,s'}$ as the probability that applying policy $A$ at state $s$ will lead to state $s'$. Finally, $V^A(s)$ is defined as the value function resulting from the operation of policy $A$. Since $B$ is chosen over $A$, we have

$$C(s, B) + \sum_{s' \in \mathcal{S}} P^B_{s,s'} V^A(s') \leq C(s, A) + \sum_{s' \in \mathcal{S}} P^A_{s,s'} V^A(s').$$

Then, for each $s \in \mathcal{S}$, we define

$$\gamma_s \triangleq C(s, B) + \sum_{s' \in \mathcal{S}} P^B_{s,s'} V^A(s') - C(s, A) - \sum_{s' \in \mathcal{S}} P^A_{s,s'} V^A(s') \leq 0.$$

Meanwhile, both policies satisfy their own Bellman equation.

$$V^A(s) + \theta^A = C(s, A) + \sum_{s' \in \mathcal{S}} P^A_{s,s'} V^A(s') \quad s \in \mathcal{S},$$

$$V^B(s) + \theta^B = C(s, B) + \sum_{s' \in \mathcal{S}} P^B_{s,s'} V^B(s') \quad s \in \mathcal{S},$$

where $\theta^A$ and $\theta^B$ are the expected costs resulting from the operation of policy $A$ and policy $B$, respectively. Then, subtracting the two expressions and bringing in the expression for $\gamma_s$ yield

$$V^B(s) - V^A(s) + \theta^B - \theta^A = \gamma_s + \sum_{s' \in \mathcal{S}} P^B_{s,s'}(V^B(s') - V^A(s')).$$

Let $V^\Delta(s) \triangleq V^B(s) - V^A(s)$ and $\theta^\Delta \triangleq \theta^B - \theta^A$. Then, we have

$$V^\Delta(s) + \theta^\Delta = \gamma_s + \sum_{s' \in \mathcal{S}} P^B_{s,s'} V^\Delta(s') \quad s \in \mathcal{S}.$$

We know that

$$\theta^\Delta = \sum_{s \in \mathcal{S}} \pi^B_s \gamma_s,$$

where $\pi^B_s$ is the steady-state probability of state $s$ under policy $B$. Since $\pi^B_s$ is non-negative and $\gamma_s$ is non-positive, we can conclude that $\theta^\Delta \leq 0$. Consequently, $\theta^B \leq \theta^A$.

Then, we prove that the resulting policy is optimal when the policy improvement step converges. We prove this by contradiction. We assume that there are two policies $A$ and $B$ that satisfy $\theta^B < \theta^A$. Meanwhile, the policy improvement step has converged to policy $A$. Since the policy has converged, we know that $\gamma_s \geq 0$ for all $s \in \mathcal{S}$. Hence, $\theta^\Delta \geq 0$. Then, according to the definition of $\theta^\Delta$, we have $\theta^B \geq \theta^A$, which contradicts the assumption. Hence, superior policies cannot remain undiscovered. Then, we can conclude that the resulting policy is optimal when the policy iteration algorithm converges.

## C.13 Proof of Theorem 12

The general procedure for the optimality proof can be summarized as follows.

1. *Policy Evaluation*: We calculate the value function resulting from the adoption of the threshold policy with $\tau = 1$.

2. *Policy Improvement*: We obtain a new policy using the value function obtained in the previous step and verify that the new policy is the threshold policy with $\tau = 1$.

In the following, we elaborate on these two steps.

Policy Evaluation We first calculate the value function under the threshold policy with $\tau = 1$. For simplicity of notation, we denote the policy as $\phi$. Let $V^{\phi}(\Delta)$ be the value function of state $(\Delta, 0, -1)$ under the policy $\phi$. Then, combining (4.16) with the expression of $P_{\Delta,\Delta'}(a)$ in Lemma 4 and Lemma 5, $V^{\phi}(\Delta)$ satisfies the following system of linear equations.

$$V^{\phi}(0) = -\theta^{\phi} + pV^{\phi}(1) + (1-p)V^{\phi}(0). \tag{C.23}$$

For **Assumption 1** and each $\Delta \geq 1$,

$$V^{\phi}(\Delta) = C(\Delta, 1) - ET\theta^{\phi} + \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P^t_{\Delta,k}(1) V^{\phi}(k) + P^t_{\Delta,\Delta+t}(1) V^{\phi}(\Delta + t) \right) \right],$$

and, for **Assumption 2** and each $\Delta \geq 1$, we have

$$V^{\phi}(\Delta) = C(\Delta, 1) - ET\theta^{\phi} + \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P^t_{\Delta,k}(1) V^{\phi}(k) + P^t_{\Delta,\Delta+t}(1) V^{\phi}(\Delta + t) \right) \right] + p_{t^+} \left( \sum_{k=0}^{t_{max}-1} P^{t^+}_{\Delta,k}(1) V^{\phi}(k) + P^{t^+}_{\Delta,\Delta+t_{max}}(1) V^{\phi}(\Delta + t_{max}) \right),$$

where $\theta^\phi$ is the expected AoII resulting from the adoption of $\phi$. It is difficult to solve the above system of linear equations directly for the exact solution. However, as we will see later, some structural properties of the value function are sufficient. These properties are summarized in the following lemma.

**Lemma 16.** *$V^\phi(\Delta)$ satisfies the following equations.*

$$V^\phi(1) - V^\phi(0) = \frac{\theta^\phi}{p},$$

$$V^\phi(\Delta+1) - V^\phi(\Delta) = \sigma \quad \Delta \geq 1,$$

*where for **Assumption 1**,*

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right)}{1 - \sum_{t=1}^{t_{max}} p p_t (1-p)^{t-1}},$$

*and, for **Assumption 2**,*

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right) + p_{t^+} \left( \frac{1-(1-p)^{t_{max}}}{p} \right)}{1 - \left( \sum_{t=1}^{t_{max}} p p_t (1-p)^{t-1} + p_{t^+} (1-p)^{t_{max}} \right)}.$$

*Proof.* First of all, from (C.23), we can easily obtain

$$V^\phi(1) - V^\phi(0) = \frac{\theta^\phi}{p}.$$

Then, we show that $V^\phi(\Delta+1) - V^\phi(\Delta)$ is constant for $\Delta \geq 1$. We start with **Assumption**

**1**. According to Theorem 9, the optimal policy exists. Hence, the iterative policy evaluation algorithm [78, pp.74] can be used to solve the system of linear equations for $V^{\phi}(\Delta)$. Let $V_{\nu}^{\phi}(\Delta)$ be the estimated value function at iteration $\nu$ of the iterative policy evaluation algorithm. Without loss of generality, we initialize $V_0^{\phi}(\Delta) = 0$ for all $\Delta$. Then, for each $\Delta \geq 1$, the value function is updated in the following way.

$$V_{\nu+1}^{\phi}(\Delta) = C(\Delta, 1) - ET\theta^{\phi} + \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta,k}^{t}(1) V_{\nu}^{\phi}(k) + P_{\Delta,\Delta+t}^{t}(1) V_{\nu}^{\phi}(\Delta + t) \right) \right].$$

Then, we have $\lim_{\nu \to \infty} V_{\nu}^{\phi}(\Delta) = V^{\phi}(\Delta)$. Hence, we can prove the desired results using mathematical induction. The base case $\nu = 0$ is true by initialization. Then, we assume $V_{\nu}^{\phi}(\Delta+1) - V_{\nu}^{\phi}(\Delta) = \sigma_{\nu}$ where $\sigma_{\nu}$ is independent of $\Delta \geq 1$. Then, we will exam whether $V_{\nu+1}^{\phi}(\Delta + 1) - V_{\nu+1}^{\phi}(\Delta)$ is independent of $\Delta \geq 1$. Leveraging the properties in Lemma 4, we have

$$\begin{aligned}
&V_{\nu+1}^{\phi}(\Delta + 1) - V_{\nu+1}^{\phi}(\Delta) \\
=&C(\Delta + 1, 1) - ET\theta^{\phi} + \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta+1,k}^{t}(1) V_{\nu}^{\phi}(k) + P_{\Delta+1,\Delta+1+t}^{t}(1) V_{\nu}^{\phi}(\Delta + t + 1) \right) \right] - \\
&C(\Delta, 1) + ET\theta^{\phi} - \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta,k}^{t}(1) V_{\nu}^{\phi}(k) + P_{\Delta,\Delta+t}^{t}(1) V_{\nu}^{\phi}(\Delta + t) \right) \right] \\
=&C(\Delta + 1, 1) - C(\Delta, 1) + \sum_{t=1}^{t_{max}} \left( p_t P_{\Delta,\Delta+t}^{t}(1) \sigma_{\nu} \right).
\end{aligned}$$

According to Lemma 6, we have

$$C(\Delta + 1, 1) - C(\Delta, 1) = \sum_{t=1}^{t_{max}} \left( C^t(\Delta + 1, 1) - C^t(\Delta, 1) \right).$$

In the case of $\Delta \geq 1$, we have

$$\begin{aligned} C^t(\Delta+1,1) - C^t(\Delta,1) =& 1 + \sum_{k=1}^{t-1} \left( (k+\Delta+1)(1-p)^k - (k+\Delta)(1-p)^k \right) \\ =& \frac{1-(1-p)^t}{p} \quad 1 \leq t \leq t_{max}. \end{aligned}$$

Combining together, we obtain

$$C(\Delta+1,1) - C(\Delta,1) = \sum_{t=1}^{t_{max}} \left( p_t \frac{1-(1-p)^t}{p} \right).$$

Hence, we can conclude that $V^{\phi}_{\nu+1}(\Delta+1) - V^{\phi}_{\nu+1}(\Delta)$ is independent of $\Delta$ when $\Delta \geq 1$. Then, by mathematical induction, $V^{\phi}(\Delta) - V^{\phi}(\Delta+1)$ is independent of $\Delta$ when $\Delta \geq 1$. We denote by $\sigma$ the constant. Then, $\sigma$ satisfies the following equation.

$$\begin{aligned} \sigma =& V^{\phi}(\Delta) - V^{\phi}(\Delta+1) \\ =& \sum_{t=1}^{t_{max}} \left( \frac{p_t - p_t(1-p)^t}{p} + p_t p (1-p)^{t-1} \sigma \right). \end{aligned}$$

After some algebraic manipulations, we obtain

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right)}{1 - \sum_{t=1}^{t_{max}} p p_t (1-p)^{t-1}}.$$

Then, we show that $V^{\phi}(\Delta+1) - V^{\phi}(\Delta)$ is independent of $\Delta \geq 1$ under **Assumption 2**. Following the same steps, we can prove the desired results by mathematical induction. We first notice that,

for each $\Delta \geq 1$, the estimated value function is updated following

$$V_{\nu+1}^{\phi}(\Delta) = C(\Delta,1) - ET\theta^{\phi} + \sum_{t=1}^{t_{max}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta,k}^{t}(1) V_{\nu}^{\phi}(k) + P_{\Delta,\Delta+t}^{t}(1) V_{\nu}^{\phi}(\Delta+t) \right) \right] + p_{t^+} \left( \sum_{k=0}^{t_{max}-1} P_{\Delta,k}^{t^+}(1) V_{\nu}^{\phi}(k) + P_{\Delta,\Delta+t_{max}}^{t^+}(1) V_{\nu}^{\phi}(\Delta+t_{max}) \right).$$

Meanwhile, the base case $\nu = 0$ is true by initialization. Then, we assume $V_{\nu}^{\phi}(\Delta+1) - V_{\nu}^{\phi}(\Delta) = \sigma_{\nu}$ where $\sigma_{\nu}$ is independent of $\Delta \geq 1$, and exam whether $V_{\nu+1}^{\phi}(\Delta+1) - V_{\nu+1}^{\phi}(\Delta)$ is independent of $\Delta \geq 1$. Leveraging the properties in Lemma 5, we have

$$V_{\nu+1}^{\phi}(\Delta+1) - V_{\nu+1}^{\phi}(\Delta) = C(\Delta+1,1) - C(\Delta,1) + \left( \sum_{t=1}^{t_{max}} p_t P_{\Delta,\Delta+t}^{t}(1) + p_{t^+} P_{\Delta,\Delta+t_{max}}^{t^+}(1) \right) \sigma_{\nu}^{\phi}.$$

Moreover, according to the expressions in Lemma 5, we obtain

$$\sum_{t=1}^{t_{max}} p_t P_{\Delta,\Delta+t}^{t}(1) + p_{t^+} P_{\Delta,\Delta+t_{max}}^{t^+}(1) = \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1} + p_{t^+} (1-p)^{t_{max}},$$

which is independent of $\Delta \geq 1$. Leveraging the expression of $C(\Delta,1)$ in Lemma 6, we obtain

$$C(\Delta,1) - C(\Delta-1,1) = \sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right) + p_{t^+} \left( \frac{1-(1-p)^{t_{max}}}{p} \right).$$

We notice that $C(\Delta,1) - C(\Delta-1,1)$ is also independent of $\Delta \geq 1$. Consequently, we can conclude that $V_{\nu+1}^{\phi}(\Delta+1) - V_{\nu+1}^{\phi}(\Delta)$ is independent of $\Delta \geq 1$. Then, by mathematical induction, $V^{\phi}(\Delta+1) - V^{\phi}(\Delta)$ is independent of $\Delta \geq 1$. We denote the constant by $\sigma$, which satisfies the

following equation.

$$\sigma = \sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right) + p_{t^+} \left( \frac{1-(1-p)^{t_{max}}}{p} \right) + \left( \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1} + p_{t^+} (1-p)^{t_{max}} \right) \sigma.$$

After some algebraic manipulations, we obtain

$$\sigma = \frac{\sum_{t=1}^{t_{max}} p_t \left( \frac{1-(1-p)^t}{p} \right) + p_{t^+} \left( \frac{1-(1-p)^{t_{max}}}{p} \right)}{1 - \left( \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1} + p_{t^+} (1-p)^{t_{max}} \right)}.$$

□

With Lemma 16 in mind, we can continue to the next step.

Policy Improvement Here, we show that the new policy induced from the $V^{\phi}(\Delta)$ obtained in the previous step and $\theta^{\phi}$ is the threshold policy with $\tau = 1$. To this end, we define $\delta V^{\phi}(\Delta) \triangleq V^{\phi,0}(\Delta) - V^{\phi,1}(\Delta)$, where $V^{\phi,a}(\Delta)$ is the value function resulting from taking action $a$ at state $(\Delta, 0, -1)$. Then, the suggested action at state $(\Delta, 0, -1)$ is $a = 1$ if $\delta V^{\phi}(\Delta) \geq 0$. Otherwise, $a = 0$ is suggested. In the following, we investigate the expression of $\delta V^{\phi}(\Delta)$. We first notice that, for $\Delta \geq 1$, $V^{\phi}(\Delta) = V^{\phi,1}(\Delta)$. Then, using Lemma 16, we obtain

$$\begin{aligned}
\delta V^{\phi}(\Delta) =& \Delta - \theta^{\phi} + (1-p) V^{\phi}(\Delta + 1) + p V(0) - V^{\phi,1}(\Delta) \\
=& \Delta - \theta^{\phi} + (1-p) V^{\phi}(\Delta + 1) + p V(0) - V^{\phi}(\Delta) \\
=& \Delta - 2\theta^{\phi} + [(1-p) - p(\Delta - 1)]\sigma,
\end{aligned}$$

where $\Delta \geq 1$. We notice that

$$\delta V^{\phi}(\Delta + 1) - \delta V^{\phi}(\Delta) = 1 - p\sigma.$$

For **Assumption 1**, plugging in the expression of $\sigma$ yields

$$\begin{aligned}
1 - p\sigma =& 1 - \frac{\sum_{t=1}^{t_{max}} (p_t - p_t(1-p)^t)}{1 - \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1}} \\
=& \frac{1 - \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1} - \sum_{t=1}^{t_{max}} (p_t - p_t(1-p)^t)}{1 - \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1}} \\
=& \frac{(1-2p) \sum_{t=1}^{t_{max}} p_t (1-p)^{t-1}}{1 - \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1}} \geq 0.
\end{aligned}$$

For **Assumption 2**, we have

$$\begin{aligned}
1 - p\sigma =& 1 - \frac{\sum_{t=1}^{t_{max}} p_t (1 - (1-p)^t) + p_{t^+}(1 - (1-p)^{t_{max}})}{1 - \left( \sum_{t=1}^{t_{max}} p_t p (1-p)^{t-1} + p_{t^+}(1-p)^{t_{max}} \right)} \\
\geq& 1 - \frac{\sum_{t=1}^{t_{max}} p_t (1 - (1-p)^t) + p_{t^+}(1 - (1-p)^{t_{max}})}{1 - \left( \sum_{t=1}^{t_{max}} p_t (1-p)^{t} + p_{t^+}(1-p)^{t_{max}} \right)} \\
=& 0.
\end{aligned}$$

Consequently, when $\Delta \geq 1$, $\delta V^{\phi}(\Delta+1) \geq \delta V^{\phi}(\Delta)$ for both assumptions. We notice that $\delta V^{\phi}(1) = 1 - 2\theta^{\phi} + (1-p)\sigma$. According to Condition 1, $\theta^{\phi} = \bar{\Delta}_1 \leq \frac{1+(1-p)\sigma}{2}$. Hence, we have

$$\delta V^{\phi}(1) = 1 - 2\bar{\Delta}_1 + (1-p)\sigma \geq 0.$$

Combining together, we have

$$\delta V^{\phi}(\Delta) \geq \delta V^{\phi}(1) \geq 0 \quad \Delta \geq 1.$$

Hence, the suggested action at state $(\Delta, 0, -1)$ where $\Delta \geq 1$ is to initiate the transmission (i.e., $a = 1$). Now, the only missing part is the action at state $(0, 0, -1)$. To determine the action, we recall from Theorem 11 that the new policy will always be no worse than the old one. Meanwhile, by Condition 1, $\bar{\Delta}_1 \leq \bar{\Delta}_0$. Hence, the suggested action at state $(0, 0, -1)$ is to stay idle (i.e., $a = 0$). Combining with the suggested actions at other states, we can conclude that the policy improvement step yields the threshold policy with $\tau = 1$.

Consequently, the policy iteration algorithm converges. Then, according to Theorem 11, we can conclude that the threshold policy with $\tau = 1$ is optimal.

# Appendix D: Freshness with Preemption

## D.1 Details of State Transition Probability

For a clearer presentation, we write $P_{s,s'}(a)$ as $Pr[s' \mid s, a]$. Then, we distinguish between different states.

- $s = (0, 0, -1)$. In this case, the channel is idle. We start with the case where the transmitter initiates a new transmission (i.e., $a = 1$). We know that the update is delivered with probability $q_1$. In this case, $t' = 0$ and $i' = -1$ by definition. Moreover, the receiver's estimate will not change. Hence, according to (5.2), we have

$$Pr[(0, 0, -1) \mid (0, 0, -1), a = 1] = q_1(1 - p).$$

$$Pr[(1, 0, -1) \mid (0, 0, -1), a = 1] = q_1 p.$$

The update will still be in transmission with probability $1 - q_1$. In this case, $t' = t + 1$ as the transmission continues. $i' = 0$ since the transmitted update is the same as the receiver's estimate. Meanwhile, the receiver's estimate will not change. Hence, according to (5.2), we have

$$Pr[(0, 1, 0) \mid (0, 0, -1), a = 1] = (1 - q_1)(1 - p).$$

$$Pr[(1,1,0) \mid (0,0,-1), a=1] = (1-q_1)p.$$

Then, we consider the case where the transmitter chooses to stay idle (i.e., $a=0$). In this case, $t'=t$ and $i'=i$. Meanwhile, the receiver's estimate will remain the same as no update is delivered. Hence, according to (5.2), we have

$$Pr[(0,0,-1) \mid (0,0,-1), a=0] = 1-p.$$

$$Pr[(1,0,-1) \mid (0,0,-1), a=0] = p.$$

- $s=(0,t,0)$ where $t \geq 1$. In this case, the channel is busy. When the transmitter chooses to terminate the current transmission and initiate a new one (i.e., $a=1$), the update is delivered with probability $q_1$. In this case, $t'=0$ and $i'=-1$ by definition. Meanwhile, the receiver's estimate will not change. Hence, according to (5.2), we have

$$Pr[(0,0,-1) \mid (0,t,0), a=1] = q_1(1-p).$$

$$Pr[(1,0,-1) \mid (0,t,0), a=1] = q_1 p.$$

The update will still be in transmission with probability $1-q_1$. In this case, $t'=1$ because a new transmission starts. $i'=0$ because the transmitted update is the same as the receiver's estimate. Also, the receiver's estimate will not change. Hence, according to (5.2), we have

$$Pr[(0,1,0) \mid (0,t,0), a=1] = (1-q_1)(1-p).$$

$$Pr[(1,1,0) \mid (0,t,0), a=1] = (1-q_1)p.$$

When the transmitter chooses $a = 0$, the transmitted update will be delivered with probability $q_{t+1}$. In this case, $t' = 0$ and $i' = -1$ by definition. Meanwhile, the receiver's estimate will not change as $i = 0$ indicates that the newly arrived update brings no new information to the receiver. Hence, according to (5.2), we have

$$Pr[(0,0,-1) \mid (0,t,0), a=0] = q_{t+1}(1-p).$$

$$Pr[(1,0,-1) \mid (0,t,0), a=0] = q_{t+1}p.$$

The transmitted update will still be in transmission with probability $1 - q_{t+1}$. In this case, $t' = t + 1$ as the transmission continues. $i' = i$, and the receiver's estimate will stay the same. Hence, according to (5.2), we have

$$Pr[(0,t+1,0) \mid (0,t,0), a=0] = (1-q_{t+1})(1-p).$$

$$Pr[(1,t+1,0) \mid (0,t,0), a=0] = (1-q_{t+1})p.$$

- $s = (0,t,1)$ where $t \geq 1$. The analysis is similar to the case of $s = (0,t,0)$ except that when the update is not preempted and is delivered, the receiver's estimate will flip. Hence, we present the results directly.

$$Pr[(0,0,-1) \mid (0,t,1), a=1] = q_1(1-p).$$

$$Pr[(1,0,-1) \mid (0,t,1), a=1] = q_1 p.$$

$$Pr[(0,1,0) \mid (0,t,1), a=1] = (1-q_1)(1-p).$$

$$Pr[(1,1,0) \mid (0,t,1), a=1] = (1-q_1)p.$$

$$Pr[(0,0,-1) \mid (0,t,1), a=0] = q_{t+1} p.$$

$$Pr[(1,0,-1) \mid (0,t,1), a=0] = q_{t+1}(1-p).$$

$$Pr[(0,t+1,1) \mid (0,t,1), a=0] = (1-q_{t+1})(1-p).$$

$$Pr[(1,t+1,1) \mid (0,t,1), a=0] = (1-q_{t+1})p.$$

- $s = (\Delta, t, i)$ where $\Delta > 0$. In this case, the analysis is similar to the case of $s = (0, t, i)$ except for the following.
  - $i' = 1$ with probability $1 - q_1$ when the transmitter chooses $a = 1$.
  - When the receiver's estimate changes, $\Gamma = 0$. Otherwise, $\Gamma = 1$. Then, the dynamics of $\Delta'$ can be determined using (5.2).

  Hence, we omit the discussion and present the results directly.

$$Pr[(0,0,-1) \mid (\Delta,0,-1), a=1] = q_1(1-p).$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,0,-1), a=1] = q_1 p.$$

$$Pr[(0,1,1) \mid (\Delta,0,-1), a=1] = (1-q_1)p.$$

$$Pr[(\Delta+1,1,1) \mid (\Delta,0,-1), a=1] = (1-q_1)(1-p).$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,0,-1), a=0] = 1-p.$$

$$Pr[(0,0,-1) \mid (\Delta,0,-1), a=0] = p.$$

For each $t \geq 1$,

$$Pr[(0,0,-1) \mid (\Delta,t,0), a=1] = q_1(1-p).$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,t,0), a=1] = q_1 p.$$

$$Pr[(0,1,1) \mid (\Delta,t,0), a=1] = (1-q_1)p.$$

$$Pr[(\Delta+1,1,1) \mid (\Delta,t,0), a=1] = (1-q_1)(1-p).$$

$$Pr[(0,0,-1) \mid (\Delta,t,0), a=0] = q_{t+1}p.$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,t,0), a=0] = q_{t+1}(1-p).$$

$$Pr[(0,t+1,0) \mid (\Delta,t,0), a=0] = (1-q_{t+1})p.$$

$$Pr[(\Delta+1,t+1,0) \mid (\Delta,t,0), a=0] = (1-q_{t+1})(1-p).$$

$$Pr[(0,0,-1) \mid (\Delta,t,1), a=1] = q_1(1-p).$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,t,1), a=1] = q_1 p.$$

$$Pr[(0,1,1) \mid (\Delta,t,1), a=1] = (1-q_1)p.$$

$$Pr[(\Delta+1,1,1) \mid (\Delta,t,1), a=1] = (1-q_1)(1-p).$$

$$Pr[(0,0,-1) \mid (\Delta,t,1), a=0] = q_{t+1}(1-p).$$

$$Pr[(\Delta+1,0,-1) \mid (\Delta,t,1), a=0] = q_{t+1}p.$$

$$Pr[(0,t+1,1) \mid (\Delta,t,1), a=0] = (1-q_{t+1})p.$$

$$Pr[(\Delta+1,t+1,1) \mid (\Delta,t,1), a=0] = (1-q_{t+1})(1-p).$$

Combining the cases together, we fully characterized the state transition probability $P_{s,s'}(a)$.

## D.2 Proof of Lemma 9

Combining with the system dynamics, the steady state probabilities satisfy the following balance equations.

$$\pi_{-1}(0) = q_1(1-p)\sum_{i=0}^{\infty}\pi(i). \tag{D.1}$$

$$\pi_{-1}(\Delta) = q_1 p \pi(\Delta-1) \quad \Delta \geq 1.$$

$$\pi_0(0) = (1-q_1)(1-p)\pi(0).$$

$$\pi_0(1) = (1-q_1)p\pi(0).$$

$$\pi_0(\Delta) = 0 \quad \Delta \geq 2.$$

$$\pi_1(0) = (1-q_1)p\sum_{i=1}^{\infty}\pi(i).$$

$$\pi_1(1) = 0.$$

$$\pi_1(\Delta) = (1-q_1)(1-p)\pi(\Delta-1) \quad \Delta \geq 2.$$

$$\sum_{i=0}^{\infty} \pi(i) = 1. \tag{D.2}$$

Combining (D.1) and (D.2) yields

$$\pi_{-1}(0) = q_1(1-p).$$

According to the definition of $\pi(0)$, we have

$$\pi(0) = q_1(1-p) + (1-q_1)(1-p)\pi(0) + (1-q_1)p(1-\pi(0)).$$

Then, we obtain

$$\pi(0) = \frac{p + q_1 - 2q_1p}{1-(1-q_1)(1-2p)}.$$

Likewise, we can obtain

$$\begin{aligned}
\pi(1) =& q_1p\pi(0) + (1-q_1)p\pi(0) = p\pi(0) \\
=& \frac{p^2 + q_1p - 2q_1p^2}{1-(1-q_1)(1-2p)}.
\end{aligned}$$

$$\begin{aligned}
\pi(\Delta) =& (q_1p + (1-q_1)(1-p))\pi(\Delta-1) \\
=& (1-q_1-p+2q_1p)\pi(\Delta-1) \quad \Delta \geq 2.
\end{aligned} \tag{D.3}$$

After some algebraic manipulation, for each $\Delta \geq 1$, we have

$$\begin{aligned}
\pi(\Delta) =& (1-q_1-p+2q_1p)^{\Delta-1}\pi(1) \\
=& \frac{(1-q_1-p+2q_1p)^{\Delta-1}(p^2+q_1p-2q_1p^2)}{1-(1-q_1)(1-2p)}.
\end{aligned}$$

## D.3 Proof of Corollary 7

We derive the closed-form expression based on Lemma 9 and its proof. We define $\Pi \triangleq \sum_{\Delta=2}^{\infty} \pi(\Delta)$. Then, we sum (D.3) from 2 to $\infty$ and apply the definition of $\Pi$, which yield

$$\Pi = (1 - q_1 - p + 2q_1p)(\Pi + \pi(1)).$$

After some algebraic manipulations, we have

$$\Pi = \frac{1 - q_1 - p + 2q_1p}{q_1 + p - 2q_1p}\pi(1).$$

We also define $\Sigma \triangleq \sum_{\Delta=2}^{\infty} \Delta\pi(\Delta)$. Then, the expected AoII achieved by the strong preemptive policy is

$$\bar{\Delta}_{sp} = \alpha\left(\pi(1) + \Sigma\right) + \beta.$$

To obtain $\Sigma$, we multiply both size of (D.3) by $\Delta - 1$.

$$(\Delta - 1)\pi(\Delta) = (1 - q_1 - p + 2q_1p)(\Delta - 1)\pi(\Delta - 1) \quad \Delta \geq 2.$$

Then, we sum the above equation from 2 to $\infty$ and apply the definitions of $\Pi$ and $\Sigma$, which yield

$$\Sigma - \Pi = (1 - q_1 - p + 2q_1p)(\Sigma + \pi(1)).$$

Then, we obtain

$$\Sigma = \frac{(1 - q_1 - p + 2q_1p)\pi(1) + \Pi}{q_1 + p - 2q_1p}.$$

Plugging in the expressions of $\pi(1)$ and $\Pi$, we obtain

$$\Sigma = \frac{p - p(q_1 + p - 2q_1p)^2}{(p + q_1 - 2q_1p)(q_1 + 2p - 2q_1p)}.$$

Combining together, we have

$$\bar{\Delta}_{sp} = \frac{\alpha p}{(p + q_1 - 2q_1p)(q_1 + 2p - 2q_1p)} + \beta.$$

## D.4 Proof of Lemma 11

Given that the value function can be computed iteratively, we use mathematical induction to prove the desired result. First, the base case $\nu = 0$ is true by initialization. Then, we assume that the monotonicity holds at iteration $\nu$ and check whether the monotonicity still holds at iteration $\nu + 1$. To this end, we first revisit how the estimated value function is updated by incorporating the structural properties of the state transition probability. From Appendix D.1, we know that, within a single transition, $\Delta$ either increases by one or decreases to zero. More precisely,

$$Pr[(\Delta', t', i') \mid (\Delta, t, i), a] = 0 \quad \Delta' \notin \{0, \Delta + 1\}.$$

Applying the structural property to (5.5) yields

$$V_{\gamma,\nu+1}(s) = \min_{a\in\mathcal{A}}\Bigg\{C(s) + \gamma\sum_{t',i'}\Bigg(Pr[(\Delta+1,t',i') \mid (\Delta,t,i),a]V_{\gamma,\nu}(\Delta+1,t',i') + \\ Pr[(0,t',i') \mid (\Delta,t,i),a]V_{\gamma,\nu}(0,t',i')\Bigg)\Bigg\} \quad s\in\mathcal{S}.$$

Moreover, the state transition probability is independent of $\Delta$ when $\Delta > 0$. Specifically, for any $\Delta_1 > 0$, $\Delta_2 > 0$, $t$, and $i$,

$$Pr[(0,t'i') \mid (\Delta_1,t,i),a] = Pr[(0,t'i') \mid (\Delta_2,t,i),a].$$

$$Pr[(\Delta_1+1,t',i') \mid (\Delta_1,t,i),a] = Pr[(\Delta_2+1,t',i') \mid (\Delta_2,t,i),a].$$

Let $V^a_{\gamma,\nu+1}(s)$ be the value function resulting from the adoption of action $a$, $s_1 = (\Delta_1,t,i)$, and $s_2 = (\Delta_2,t,i)$ where $\Delta_1 \geq \Delta_2 > 0$. We notice that the immediate cost $C(s)$ depends only on $\Delta$ and is non-decreasing in $\Delta$. Combined with the assumption on the estimated value function at iteration $\nu$, we can conclude that

$$V^a_{\gamma,\nu+1}(s_1) \geq V^a_{\gamma,\nu+1}(s_2) \quad a\in\{0,1\}.$$

Since $V_{\gamma,\nu+1}(s) = \min_{a\in\{0,1\}}\{V^a_{\gamma,\nu+1}(s)\}$, we can conclude that the monotonicity still holds at iteration $\nu+1$. Then, by mathematical induction, the lemma is true.

## D.5 Proof of Theorem 14

The proof is based on [81, pp.42-43]. We consider a generic MDP $\mathcal{M}^G$. We first clarify the notations we will use in the proof.

- The state space and the state is denoted by $\mathcal{S}^G$ and $s$, respectively. The action suggested by policy $A$ at state $s$ is denoted by $A(s)$.
- The probability that operating policy $A$ at state $s$ leads to state $s'$ is denoted by $P^A_{s,s'}$. Likewise, the probability that action $a$ at state $s$ leads to state $s'$ is denoted by $P_{s,s'}(a)$.
- The immediate cost for operating policy $A$ at state $s$ is denoted by $C(s, A)$. Similarly, the immediate cost for operating action $a$ at state $s$ is denoted by $C(s, a)$
- The value function of state $s$ resulting from the adoption of policy $A$ is denoted by $V^A(s)$. The expected cost achieved by policy $A$ is denoted by $\theta^A$.

With the notations in mind, we prove the optimality by contradiction. We assume that the policy improvement function has converged to policy $A$ and there exists a policy $B$ such that $\theta^A > \theta^B$.

We recall that the policy improvement function procedures a new policy $\psi'$ based on the old policy $\psi$ using the following equation.

$$\psi'(s) = \underset{a \in \mathcal{A}}{\operatorname{argmin}} \left\{ C(s,a) + \sum_{s' \in \mathcal{S}^G} P_{s,s'}(a) V^{\psi}(s') \right\}.$$

Since the policy improvement function has converged to policy $A$, we have the following inequality

holds for any policy $B$.

$$C(s,A) + \sum_{s' \in \mathcal{S}^G} P^A_{s,s'} V^A(s') \leq C(s,B) + \sum_{s' \in \mathcal{S}^G} P^B_{s,s'} V^A(s').$$

We define

$$\delta(s) \triangleq C(s,A) - C(s,B) + \sum_{s' \in \mathcal{S}^G} (P^A_{s,s'} - P^B_{s,s'}) V^A(s') \leq 0.$$

Meanwhile, $V^A(s)$ and $V^B(s)$ satisfy there own Bellman equations. More precisely,

$$V^A(s) + \theta^A = C(s,A) + \sum_{s' \in \mathcal{S}^G} P^A_{s,s'} V^A(s').$$

$$V^B(s) + \theta^B = C(s,B) + \sum_{s' \in \mathcal{S}^G} P^B_{s,s'} V^B(s').$$

Subtracting the above two equations and bringing in $\delta(s)$ yield

$$V^A(s) - V^B(s) + \theta^A - \theta^B = \delta(s) + \sum_{s' \in \mathcal{S}^G} P^B_{s,s'} (V^A(s') - V^B(s')).$$

Let $V^\delta(s) \triangleq V^A(s) - V^B(s)$ and $\theta^\delta \triangleq \theta^A - \theta^B$. Plugging in the definitions yields

$$V^\delta(s) + \theta^\delta = \delta(s) + \sum_{s' \in \mathcal{S}^G} P^B_{s,s'} V^\delta(s').$$

As is mentioned in Section 5.4, each policy induces a DTMC and the expected cost $\theta^\delta = \sum_{s \in \mathcal{S}^G} \delta(s) \pi^B(s)$ where $\pi^B(s)$ is the stationary distribution of the DTMC induced by policy $B$. Since the stationary distribution is non-negative and $\delta(s) \leq 0$ for all $s \in \mathcal{M}^G$, we can conclude that $\theta^\delta \leq 0$. In other words, $\theta^A \leq \theta^B$, which contradicts the assumption that $\theta^A > \theta^B$.

Therefore, the contradiction occurs, and the converged policy $A$ is optimal.

## D.6 Proof of Theorem 15

We first calculate the value function $V^{\psi}(s)$ resulting from adopting the strong preemptive policy $\psi$. Combining the strong preemptive policy definition and (5.7), we know that the value function satisfies the following system of linear equations.

$$V^{\psi}(0,0,-1) = f(0) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(0,1,0) + (1-p_s)pV^{\psi}(1,1,0) + p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(1,0,-1).$$

$$V^{\psi}(0,t,0) = f(0) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(0,1,0) + (1-p_s)pV^{\psi}(1,1,0) + p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(1,0,-1) \quad t \geq 1.$$

$$V^{\psi}(0,t,1) = f(0) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(0,1,0) + (1-p_s)pV^{\psi}(1,1,0) + p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(1,0,-1) \quad t \geq 1.$$

For each $\Delta \geq 1$,

$$V^{\psi}(\Delta,0,-1) = f(\Delta) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(\Delta+1,1,1) + (1-p_s)pV^{\psi}(0,1,1) + p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(\Delta+1,0,-1).$$

$$V^{\psi}(\Delta,t,0) = f(\Delta) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(\Delta+1,1,1) + (1-p_s)pV^{\psi}(0,1,1) +$$
$$p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(\Delta+1,0,-1) \quad t \geq 1.$$

$$V^{\psi}(\Delta,t,1) = f(\Delta) - \theta^{\psi} + (1-p_s)(1-p)V^{\psi}(\Delta+1,1,1) + (1-p_s)pV^{\psi}(0,1,1) +$$
$$p_s(1-p)V^{\psi}(0,0,-1) + p_s pV^{\psi}(\Delta+1,0,-1) \quad t \geq 1.$$

We notice that the size of the system of linear equations is infinite, so it is difficult to obtain the solution by solving directly. However, some structural properties of $V^{\psi}(s)$ are sufficient for us to complete the proof. These structural properties are summarized in the following lemma.

**Lemma 17.** *$V^{\psi}(s)$ possesses the following structural properties.*

1. $V^{\psi}(\Delta,0,-1) = V^{\psi}(\Delta,t,0) = V^{\psi}(\Delta,t,1) \triangleq V^{\psi}(\Delta)$ *for* $\Delta \geq 0$ *and* $t \geq 1$.

2. $V^{\psi}(\Delta)$ *is non-decreasing in* $\Delta$.

*Proof.* The first property can be verified easily by comparing the linear equations they satisfy. Hence, we will focus on the second property. The system of linear equations can be solved iteratively [78]. More precisely,

$$V^{\psi}_{\nu+1}(s) = f(\Delta) - \theta^{\psi} + \sum_{s' \in \mathcal{S}} P^{\psi}_{s,s'} V^{\psi}_{\nu}(s') \quad s \in \mathcal{S},$$

where $V^{\psi}_{\nu}(s)$ is the estimated value function at iteration $\nu$ and $P^{\psi}_{s,s'}$ is the probability that operating policy $\psi$ at state $s$ leads to state $s'$. We know that $\lim_{\nu\to\infty} V^{\psi}_{\nu}(s) = V^{\psi}(s)$. Then, leveraging the iterative nature, we can use mathematical induction to prove the desired results. We initialize

$V_0^{\psi}(s) = 0$ for $s \in \mathcal{S}$. Then, the base case $\nu = 0$ is true by initialization. We assume the monotonicity is true at iteration $\nu$. Then, we check if it holds at iteration $\nu + 1$. Using the first property, we have the following holds for state $s$ with $\Delta \geq 1$.

$$\begin{aligned} V_{\nu+1}^{\psi}(\Delta+1) - V_{\nu+1}^{\psi}(\Delta) = f(\Delta+1) - f(\Delta) + \\ (1-p_s)(1-p)[V_{\nu}^{\psi}(\Delta+2) - V_{\nu}^{\psi}(\Delta+1)] + \\ p_s p[V_{\nu}^{\psi}(\Delta+2) - V_{\nu}^{\psi}(\Delta+1)]. \end{aligned}$$

Applying the assumption for iteration $\nu$ and the monotonicity of the time penalty function, we can easily conclude that $V_{\nu+1}^{\psi}(\Delta+1) \geq V_{\nu+1}^{\psi}(\Delta)$ when $\Delta \geq 1$. Then, we consider the case of $\Delta = 0$.

$$\begin{aligned} V_{\nu+1}^{\psi}(1) - V_{\nu+1}^{\psi}(0) =& f(1) - f(0) + (1-p_s)(1-p)[V_{\nu}^{\psi}(2) - V_{\nu}^{\psi}(0)] + \\ & (1-p_s)p[V_{\nu}^{\psi}(0) - V_{\nu}^{\psi}(1)] + p_s p[V_{\nu}^{\psi}(2) - V_{\nu}^{\psi}(1)] \\ \geq& (1-p_s)(1-p)[V_{\nu}^{\psi}(1) - V_{\nu}^{\psi}(0)] + p_s p[V_{\nu}^{\psi}(2) - V_{\nu}^{\psi}(1)] + \\ & (1-p_s)p[V_{\nu}^{\psi}(0) - V_{\nu}^{\psi}(1)] \\ =& (1-p_s)(1-2p)[V_{\nu}^{\psi}(1) - V_{\nu}^{\psi}(0)] + p_s p[V_{\nu}^{\psi}(2) - V_{\nu}^{\psi}(1)]. \end{aligned}$$

We recall that $0 < p < \frac{1}{2}$. Hence, $V_{\nu+1}^{\psi}(1) \geq V_{\nu+1}^{\psi}(0)$. Combining together, the property holds at iteration $\nu + 1$. Then, by mathematical induction, we can conclude that the second property is true. □

Equipped with Lemma 17, we can proceed to obtain the new policy induced by the value function $V^{\psi}(s)$. To this end, we define $V^{\psi,a}(s)$ as the expected cost resulting from taking action

$a$ at state $s$, which can be calculated using the following equation.

$$V^{\psi,a}(s) = f(\Delta) - \theta^{\psi} + \sum_{s' \in \mathcal{S}} P_{s,s'}(a) V^{\psi}(s').$$

Consequently, to determine the new action, we only need to determine the sign of $\delta V^{\psi}(s) \triangleq V^{\psi,0}(s) - V^{\psi,1}(s)$. When $\delta V^{\psi}(s) < 0$, the action suggested by the new policy is $a = 0$. Otherwise, $a = 1$ is suggested. Without loss of generality, we let $V^{\psi}(0) = 0$. Then, we distinguish between different states.

- $s = (0, 0, -1)$.

$$\delta V^{\psi}(0, 0, -1) = pV^{\psi}(1) - (1 - p_s)pV^{\psi}(1) - p_s pV^{\psi}(1) = 0.$$

- $s = (\Delta, 0, -1)$ where $\Delta \geq 1$.

$$\begin{aligned}\delta V^{\psi}(\Delta, 0, -1) =& (1-p)V^{\psi}(\Delta+1) - (1-p_s)(1-p)V^{\psi}(\Delta+1) - p_s pV^{\psi}(\Delta+1) \\ =& p_s(1-2p)V^{\psi}(\Delta+1) \\ \geq& 0.\end{aligned}$$

- $s = (0, t, 0)$ where $t \geq 1$.

$$\begin{aligned}\delta V^{\psi}(0, t, 0) =& (1-p_s)pV^{\psi}(1) + p_s pV^{\psi}(1) - (1-p_s)pV^{\psi}(1) - p_s pV^{\psi}(1) \\ =& 0.\end{aligned}$$

- $s = (0, t, 1)$ where $t \geq 1$.

$$\begin{aligned}\delta V^{\psi}(0,t,1) =&(1-p_s)pV^{\psi}(1) + p_s(1-p)V^{\psi}(1) - (1-p_s)pV^{\psi}(1) - p_spV^{\psi}(1)\\ =&p_s(1-2p)V^{\psi}(1)\\ \geq&0.\end{aligned}$$

- $s = (\Delta, t, 0)$ where $\Delta \geq 1$ and $t \geq 1$.

$$\begin{aligned}\delta V^{\psi}(\Delta,t,0) =&(1-p_s)(1-p)V^{\psi}(\Delta+1) + p_s(1-p)V^{\psi}(\Delta+1) -\\ &(1-p_s)(1-p)V^{\psi}(\Delta+1) - p_spV^{\psi}(\Delta+1)\\ =&p_s(1-2p)V^{\psi}(\Delta+1)\\ \geq&0.\end{aligned}$$

- $s = (\Delta, t, 1)$ where $\Delta \geq 1$ and $t \geq 1$.

$$\begin{aligned}\delta V^{\psi}(\Delta,t,1) =&(1-p_s)(1-p)V^{\psi}(\Delta+1) + p_spV^{\psi}(\Delta+1) -\\ &(1-p_s)(1-p)V^{\psi}(\Delta+1) - p_spV^{\psi}(\Delta+1)\\ =&0.\end{aligned}$$

Combing together, we know that $\delta V^{\psi}(s) \geq 0$ for all $s \in \mathcal{S}$, meaning that the new policy always suggests the cation $a = 1$. Hence, the new policy is still the strong preemptive policy. Then, by Theorem 14, we can conclude that the strong preemptive policy is optimal.

## D.7 Proof of Theorem 17

We follow the same methodology presented in the proof of Theorem 15. First, we calculate the value function $V^{\psi}(s)$ resulting from the adoption of the threshold preemptive policy $\psi$. The value function satisfies the following system of linear equations.

$$V^{\psi}(0,0,-1) = f(0) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(0,1,0) + (1-q_1)pV^{\psi}(1,1,0) + q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(1,0,-1).$$

$$V^{\psi}(0,t,0) = f(0) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(0,1,0) + (1-q_1)pV^{\psi}(1,1,0) + q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(1,0,-1) \quad 1 \leq t \leq t_{max} - 1.$$

$$V^{\psi}(0,t,1) = f(0) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(0,1,0) + (1-q_1)pV^{\psi}(1,1,0) + q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(1,0,-1) \quad 1 \leq t \leq t_{max} - 1.$$

For each $\Delta \geq 1$,

$$V^{\psi}(\Delta,0,-1) = f(\Delta) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(\Delta+1,1,1) + (1-q_1)pV^{\psi}(0,1,1) + q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(\Delta+1,0,-1).$$

$$V^{\psi}(\Delta,t,0) = f(\Delta) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(\Delta+1,1,1) + (1-q_1)pV^{\psi}(0,1,1) +$$
$$q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(\Delta+1,0,-1) \quad 1 \le t \le t_{max}-1.$$

$$V^{\psi}(\Delta,t,1) = f(\Delta) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(\Delta+1,1,1) + (1-q_1)pV^{\psi}(0,1,1) +$$
$$q_1(1-p)V^{\psi}(0,0,-1) + q_1pV^{\psi}(\Delta+1,0,-1) \quad 1 \le t \le t_{max}-2.$$

$$V^{\psi}(\Delta,t_{max}-1,1) = f(\Delta) - \theta^{\psi} + (1-p)V^{\psi}(0,0,-1) + pV^{\psi}(\Delta+1,0,-1). \tag{D.4}$$

Instead of solving the above system of linear equations, some structural properties of the solution will be sufficient for the following analysis.

**Lemma 18.** *$V^{\psi}(s)$ possesses the following properties.*

1. *$V^{\psi}(\Delta,0,-1) = V^{\psi}(\Delta,t,0) \triangleq V^{\psi}(\Delta)$ for $\Delta \ge 0$ and $1 \le t \le t_{max}-1$.*
2. *$V^{\psi}(0,t,1) = V^{\psi}(0)$ for $1 \le t \le t_{max}-1$. $V^{\psi}(\Delta,t,1) = V^{\psi}(\Delta)$ for $\Delta > 0$ and $1 \le t \le t_{max}-2$.*
3. *$V^{\psi}(\Delta)$ is non-decreasing in $\Delta$.*
4. *$V^{\psi}(1) - V^{\psi}(0) = \frac{\theta^{\psi}-f(0)}{p}$ and $V^{\psi}(\Delta+1) - V^{\psi}(\Delta) \triangleq \sigma$ is independent of $\Delta \ge 1$, where*

$$\sigma = \frac{\alpha}{q_1 + p - 2q_1p}.$$

*Proof.* The first two properties are obvious, as we can verify by comparing the corresponding linear equations. For the third property, the proof is based on the mathematical induction as

presented in the proof of Lemma 17. Hence, we omit the proof here for the sake of space. In the following, we focus on the last property. Applying the first two properties to the system of linear equations yields

$$V^{\psi}(0) = f(0) - \theta^{\psi} + (1-p)V^{\psi}(0) + pV^{\psi}(1). \tag{D.5}$$

$$\begin{aligned} V^{\psi}(\Delta) =& f(\Delta) - \theta^{\psi} + (1-q_1)(1-p)V^{\psi}(\Delta+1) + (1-q_1)pV^{\psi}(0) + q_1(1-p)V^{\psi}(0) + \\ & q_1 pV^{\psi}(\Delta+1) \quad \Delta \geq 1. \end{aligned}$$

From (D.5), we can easily conclude that

$$V^{\psi}(1) - V^{\psi}(0) = \frac{\theta^{\psi} - f(0)}{p}.$$

In the following, we prove that $V^{\psi}(\Delta+1) - V^{\psi}(\Delta)$ is independent of $\Delta \geq 1$. We recall that $V^{\psi}(\Delta)$ can be calculated using the iterative method. Let $V_{\nu}^{\psi}(\Delta)$ be the estimated value function at iteration $\nu$, which is updated in the following way.

$$V_{\nu+1}^{\psi}(0) = f(0) - \theta^{\psi} + (1-p)V_{\nu}^{\psi}(0) + pV_{\nu}^{\psi}(1).$$

$$\begin{aligned} V_{\nu+1}^{\psi}(\Delta) =& f(\Delta) - \theta^{\psi} + (1-q_1)(1-p)V_{\nu}^{\psi}(\Delta+1) + (1-q_1)pV_{\nu}^{\psi}(0) + q_1(1-p)V_{\nu}^{\psi}(0) + \\ & q_1 pV_{\nu}^{\psi}(\Delta+1) \quad \Delta \geq 1. \end{aligned}$$

Then, we know that $\lim_{\nu\to\infty} V_{\nu}^{\psi}(\Delta) = V^{\psi}(\Delta)$. Consequently, we can use mathematical induction

to prove the desired results. To this end, we initialize $V_0^{\psi}(\Delta) = 0$ for $\Delta \geq 0$. Then, the base case $\nu = 0$ is true by initialization. We assume the property holds at iteration $\nu$ and examine whether it still holds at iteration $\nu + 1$. We recall that $f(\Delta) = \alpha\Delta + \beta$. Hence, we have

$$V_{\nu+1}^{\psi}(\Delta+1) - V_{\nu+1}^{\psi}(\Delta) = \\ \alpha + (1-q_1)(1-p)[V_{\nu}^{\psi}(\Delta+2) - V_{\nu}^{\psi}(\Delta+1)] + \\ q_1 p[V_{\nu}^{\psi}(\Delta+2) - V_{\nu}^{\psi}(\Delta+1)] \quad \Delta \geq 1.$$

According to our assumption, $V_{\nu}^{\psi}(\Delta+1) - V_{\nu}^{\psi}(\Delta)$ is independent of $\Delta \geq 1$. Hence, we can conclude that $V_{\nu+1}^{\psi}(\Delta+1) - V_{\nu+1}^{\psi}(\Delta)$ is also independent of $\Delta \geq 1$. Then, by mathematical induction, we can conclude that $V^{\psi}(\Delta+1) - V^{\psi}(\Delta)$ is independent of $\Delta \geq 1$. To calculate the constant $\sigma$, we have

$$\begin{aligned} V^{\psi}(\Delta+1) - V^{\psi}(\Delta) =& \alpha + (1-q_1)(1-p)[V^{\psi}(\Delta+2) - V^{\psi}(\Delta+1)] + \\ & q_1 p[V^{\psi}(\Delta+2) - V^{\psi}(\Delta+1)] \\ =& \alpha + (1-q_1)(1-p)\sigma + q_1 p\sigma. \end{aligned}$$

Finally, we obtain

$$\sigma = \frac{\alpha}{q_1 + p - 2q_1 p}.$$

□

With Lemma 18 in mind, we can proceed with obtaining the policy induced by $V^{\psi}(s)$. Same as we did in the proof of Theorem 15, we define $V^{\psi,a}(s)$ as the expected cost resulting

from taking action $a$ at state $s$. To determine the induced policy, we only need to determine the sign of $\delta V^{\psi}(s) \triangleq V^{\psi,0}(s) - V^{\psi,1}(s)$ for $s \in \mathcal{S}$. When $\delta V^{\psi}(s) < 0$, the action suggested by the induced policy is $a = 0$. Otherwise, $a = 1$ is suggested. Without loss of generality, we let $V^{\psi}(0) = 0$. Then, we distinguish between the following states.

- $s = (0, 0, -1)$.

$$\delta V^{\psi}(0, 0, -1) = pV^{\psi}(1) - (1 - q_1)pV^{\psi}(1) - q_1 pV^{\psi}(1) = 0.$$

- $s = (\Delta, 0, -1)$ where $\Delta \geq 1$. When $t_{max} > 2$, we have

$$\begin{aligned}\delta V^{\psi}(\Delta, 0, -1) =& (1 - p)V^{\psi}(\Delta + 1) - (1 - q_1)(1 - p)V^{\psi}(\Delta + 1) - q_1 pV^{\psi}(\Delta + 1) \\ =& q_1(1 - 2p)V^{\psi}(\Delta + 1) \\ \geq& 0.\end{aligned}$$

When $t_{max} = 2$, we have

$$\delta V^{\psi}(\Delta, 0, -1) = (1 - p)V^{\psi}(\Delta + 1) - (1 - q_1)(1 - p)V^{\psi}(\Delta + 1, 1, 1) - q_1 pV^{\psi}(\Delta + 1).$$

Since $V^{\psi}(\Delta+1,1,1)$ satisfies (D.4), we have

$$\begin{aligned} V^{\psi}(\Delta+1,1,1)-V^{\psi}(\Delta+1) =& pV^{\psi}(\Delta+1)-(1-q_1)(1-p)V^{\psi}(\Delta+1)- \\ & q_1pV^{\psi}(\Delta+1) \\ =&(1-q_1)(2p-1)V^{\psi}(\Delta+1) \\ \leq& 0. \end{aligned} \tag{D.6}$$

Then, we know that $V^{\psi}(\Delta+1,1,1)\leq V^{\psi}(\Delta+1)$. Consequently,

$$\begin{aligned} \delta V^{\psi}(\Delta,0,-1) \geq&(1-p)V^{\psi}(\Delta+1)-(1-q_1)(1-p)V^{\psi}(\Delta+1)-q_1pV^{\psi}(\Delta+1) \\ =&q_1(1-2p)V^{\psi}(\Delta+1) \\ \geq& 0. \end{aligned}$$

- $s=(0,t,0)$ where $1\leq t\leq t_{max}-1$.

$$\delta V^{\psi}(0,t,0)=(1-q_{t+1})pV^{\psi}(1)+q_{t+1}pV^{\psi}(1)-(1-q_1)pV^{\psi}(1)-q_1pV^{\psi}(1)=0.$$

- $s=(0,t,1)$ where $1\leq t\leq t_{max}-3$ and $t_{max}\geq 4$.

$$\begin{aligned} \delta V^{\psi}(0,t,1) =&(1-q_{t+1})pV^{\psi}(1)+q_{t+1}(1-p)V^{\psi}(1)-(1-q_1)pV^{\psi}(1)-q_1pV^{\psi}(1) \\ =&q_{t+1}(1-2p)V^{\psi}(1) \\ \geq& 0. \end{aligned}$$

- $s = (0, t_{max} - 2, 1)$ where $t_{max} \geq 3$.

$$\begin{aligned}\delta V^{\psi}(0, t_{max} - 2, 1) =&(1 - q_{t_{max}-1})pV^{\psi}(1, t_{max} - 1, 1)+\\ &q_{t_{max}-1}(1-p)V^{\psi}(1) - pV^{\psi}(1).\end{aligned}$$

We notice that $V^{\psi}(1, t_{max} - 1, 1)$ satisfies (D.4). Hence, replacing $V^{\psi}(1, t_{max} - 1, 1)$ with the corresponding expression yields

$$\begin{aligned}\delta V^{\psi}(0, t_{max} - 2, 1) =&(1 - q_{t_{max}-1})p[f(1) - \theta^{\psi} + pV^{\psi}(2)]+\\ &(q_{t_{max}-1} - q_{t_{max}-1}p - p)V^{\psi}(1).\end{aligned}$$

According to Lemma 18, $V^{\psi}(2) = V^{\psi}(1) + \sigma$. Hence, we have

$$\begin{aligned}\delta V^{\psi}(0, t_{max} - 2, 1) =&(1 - q_{t_{max}-1})p[f(1) - \theta^{\psi} + p(V^{\psi}(1) + \sigma)]+\\ &(q_{t_{max}-1} - q_{t_{max}-1}p - p)V^{\psi}(1)\\ =&[(p-1)(p - q_{t_{max}-1}) - q_{t_{max}-1}p^2]V^{\psi}(1)+\\ &(1 - q_{t_{max}-1})p(f(1) - \theta^{\psi} + p\sigma).\end{aligned}$$

We recall that the expected AoII $\theta^{\psi}$ is given by Theorem 16. Plugging the expressions for $\sigma$, $\theta^{\psi}$, and using the property that $V^{\psi}(1) = \frac{\theta^{\psi} - f(0)}{p}$ yield

$$\delta V^{\psi}(0, t_{max} - 2, 1) = \alpha \frac{(q_{t_{max}-1} - q_{t_{max}-1}p - p) + (1 - q_{t_{max}-1})p(q_1 + 2p - 2q_1p)^2}{(q_1 + 2p - 2q_1p)(q_1 + p - 2q_1p)}. \tag{D.7}$$

We notice that $\alpha > 0$ and the denominator of (D.7) is positive. Hence, examining the sign

of the numerator of (D.7) is sufficient to determine the sign of $\delta V^{\psi}(0, t_{max} - 2, 1)$. To this end, we define

$$\mathcal{Q}_1 \triangleq (q_{t_{max}-1} - q_{t_{max}-1}p - p) + (1 - q_{t_{max}-1})p(q_1 + 2p - 2q_1p)^2.$$

Then, by Condition 2, we know that $\mathcal{Q}_1 \geq 0$. Consequently, $\delta V^{\psi}(0, t_{max} - 2, 1) \geq 0$.

- $s = (0, t_{max} - 1, 1)$.

$$\begin{aligned}\delta V^{\psi}(0, t_{max} - 1, 1) =& (1-p)V^{\psi}(1) - (1-q_1)pV^{\psi}(1) - q_1pV^{\psi}(1)\\ =& (1-2p)V^{\psi}(1)\\ \geq& 0.\end{aligned}$$

- $s = (\Delta, t, 0)$ where $\Delta \geq 1$ and $1 \leq t \leq t_{max} - 1$. When $t_{max} > 2$, we have

$$\begin{aligned}\delta V^{\psi}(\Delta, t, 0) =& (1-q_{t+1})(1-p)V^{\psi}(\Delta+1) + q_{t+1}(1-p)V^{\psi}(\Delta+1) -\\ & (1-q_1)(1-p)V^{\psi}(\Delta+1) - q_1pV^{\psi}(\Delta+1)\\ =& (1-p)V^{\psi}(\Delta+1) - (1-q_1)(1-p)V^{\psi}(\Delta+1) - q_1pV^{\psi}(\Delta+1)\\ =& q_1(1-2p)V^{\psi}(\Delta+1)\\ \geq& 0.\end{aligned}$$

When $t_{max} = 2$, we recall that $V^{\psi}(\Delta + 1, 1, 1) \leq V^{\psi}(\Delta + 1)$. Hence, we have

$$\begin{aligned}\delta V^{\psi}(\Delta, t, 0) =&(1 - q_{t+1})(1 - p)V^{\psi}(\Delta + 1) + q_{t+1}(1 - p)V^{\psi}(\Delta + 1)- \\ &(1 - q_1)(1 - p)V^{\psi}(\Delta + 1, 1, 1) - q_1 pV^{\psi}(\Delta + 1) \\ \geq&(1 - p)V^{\psi}(\Delta + 1) - (1 - q_1)(1 - p)V^{\psi}(\Delta + 1) - q_1 pV^{\psi}(\Delta + 1) \\ \geq&0.\end{aligned}$$

- $s = (\Delta, t, 1)$ where $\Delta \geq 1$, $1 \leq t \leq t_{max} - 3$, and $t_{max} \geq 4$.

$$\begin{aligned}\delta V^{\psi}(\Delta, t, 1) =&(1 - q_{t+1})(1 - p)V^{\psi}(\Delta + 1) + q_{t+1}pV^{\psi}(\Delta + 1)- \\ &(1 - q_1)(1 - p)V^{\psi}(\Delta + 1) - q_1 pV^{\psi}(\Delta + 1) \\ =&(q_1 - q_{t+1})(1 - 2p)V^{\psi}(\Delta + 1).\end{aligned}$$

Since we assume that Condition 2 holds, we know that $\delta V^{\psi}(\Delta, t, 1) \geq 0$.

- $s = (\Delta, t_{max} - 2, 1)$ where $\Delta \geq 1$ and $t_{max} \geq 3$.

$$\begin{aligned}\delta V^{\psi}(\Delta, t_{max} - 2, 1) = &(1 - q_{t_{max}-1})(1 - p)V^{\psi}(\Delta + 1, t_{max-1} - 1, 1)+ \\ &q_{t_{max}-1}pV^{\psi}(\Delta + 1) - (1 - q_1)(1 - p)V^{\psi}(\Delta + 1) - q_1 pV^{\psi}(\Delta + 1).\end{aligned}$$

We recall that $V^{\psi}(\Delta+1, t_{max}-1, 1)$ satisfies (D.4). Hence, replacing $V^{\psi}(\Delta+1, t_{max}-1, 1)$

with corresponding expression yields

$$\begin{aligned}\delta V^{\psi}(\Delta, t_{max}-2, 1) =& (1-q_{t_{max}-1})(1-p)(f(\Delta+1)-\theta^{\psi}+ \\ & (1-p)V^{\psi}(0)+pV^{\psi}(\Delta+2))+q_{t_{max}-1}pV^{\psi}(\Delta+1)- \\ & (1-q_1)(1-p)V^{\psi}(\Delta+1)-q_1pV^{\psi}(\Delta+1).\end{aligned}$$

We recall that $V^{\psi}(\Delta+1)-V^{\psi}(\Delta)=\sigma$ for $\Delta \geq 1$. Then,

$$\begin{aligned}\delta V^{\psi}(\Delta, t_{max}-2, 1) =& [(1-q_1)(2p-1)-p^2(1-q_{t_{max}-1})]V^{\psi}(1)+ \\ & (1-q_{t_{max}-1})(1-p)(f(\Delta+1)-\theta^{\psi})+ \\ & \{[(1-q_1)(2p-1)-p^2(1-q_{t_{max}-1})]\Delta+ \\ & (1-q_{t_{max}-1})(1-p)p\}\sigma.\end{aligned}$$

Meanwhile, $V^{\psi}(1)=\frac{\theta^{\psi}-f(0)}{p}$. Hence, we have

$$\begin{aligned}\frac{\delta V^{\psi}(\Delta, t_{max}-2, 1)}{\alpha} = & \frac{(1-q_1)(2p-1)-p(1-q_{t_{max}-1})}{p}\theta^{\psi}+ \\ & (1-q_{t_{max}-1})(1-p)(\Delta+1)+ \\ & \{[(1-q_1)(2p-1)-p^2(1-q_{t_{max}-1})]\Delta+ \\ & (1-q_{t_{max}-1})(1-p)p\}\sigma.\end{aligned}$$

We define the coefficient before $\Delta$ as $\mathcal{Q}_2$. Then, we have

$$\begin{aligned}\mathcal{Q}_2 =&[(1-q_1)(2p-1)-p^2(1-q_{t_{max}-1})]\sigma+(1-q_{t_{max}-1})(1-p)\\ =&\frac{(1-q_1)(2p-1)-p^2(1-q_{t_{max}-1})+(1-q_{t_{max}-1})(1-p)(q_1+p-2q_1p)}{q_1+p-2q_1p}\\ =&\frac{(1-2p)\{q_1-1+(1-q_{t_{max}-1})[p+q_1(1-p)]\}}{q_1+p-2q_1p}.\end{aligned}$$

Then, under Condition 2, we know that $\mathcal{Q}_2 \geq 0$. Hence, $\frac{\delta V^{\psi}(\Delta,t_{max}-2,1)}{\alpha}$ is non-decreasing in $\Delta$. Then, when $\Delta = 1$, we have

$$\begin{aligned}\frac{\delta V^{\psi}(1,t_{max}-2,1)}{\alpha} =&\frac{(1-q_1)(2p-1)-p(1-q_{t_{max}-1})}{(2p+q_1-2q_1p)(p+q_1-2q_1p)}+\frac{(1-q_{t_{max}-1})(1-p)p}{q_1+p-2q_1p}+\\ &(1-q_{t_{max}-1})(1-p)+\mathcal{Q}_2\\ \triangleq&\mathcal{Q}_3.\end{aligned}$$

Again, under Condition 2, we know that $\mathcal{Q}_3 \geq 0$. Combing with the fact that $\alpha > 0$, we can conclude that $\delta V^{\psi}(1,t_{max}-2,1) \geq 0$. Consequently, $\delta V^{\psi}(\Delta,t_{max}-2,1) \geq \delta V^{\psi}(1,t_{max}-2,1) \geq 0$ for $\Delta \geq 1$.

- $s = (\Delta, t_{max}-1, 1)$ where $\Delta \geq 1$. When $t_{max} > 2$, we have

$$\begin{aligned}\delta V^{\psi}(\Delta,t_{max}-1,1) =&pV^{\psi}(\Delta+1)-(1-q_1)(1-p)V^{\psi}(\Delta+1)-q_1pV^{\psi}(\Delta+1)\\ =&(1-q_1)(2p-1)V^{\psi}(\Delta+1)\\ \leq&0.\end{aligned}$$

When $t_{max} = 2$, we have

$$\delta V^{\psi}(\Delta, 1, 1) = pV^{\psi}(\Delta + 1) - (1 - q_1)(1 - p)V^{\psi}(\Delta + 1, 1, 1) - q_1 pV^{\psi}(\Delta + 1).$$

From (D.6), we know that $V^{\psi}(\Delta + 1, 1, 1) = (1 - q_1)(2p - 1)V^{\psi}(\Delta + 1) + V^{\psi}(\Delta + 1)$.

Bring in the expression yields

$$\delta V^{\psi}(\Delta, 1, 1) = (1 - q_1)[(2 - 2q_1)p^2 + (3q_1 - 1)p - q_1]V^{\psi}(\Delta + 1).$$

We recall that $0 < p < \frac{1}{2}$ and $0 \leq q_1 \leq 1$. Hence, $\delta V^{\psi}(\Delta, 1, 1) \leq 0$.

Combining the above cases, we can conclude that the policy iteration algorithm converges to the threshold preemptive policy. Then, by Theorem 14, the threshold preemptive policy is optimal.

# Bibliography


[1] Jianhua Shi, Jiafu Wan, Hehua Yan, and Hui Suo. A survey of cyber-physical systems. In *2011 international conference on wireless communications and signal processing (WCSP)*, pages 1–6. IEEE, 2011.

[2] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376, 2015.

[3] Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE internet of things journal*, 4(5):1125–1142, 2017.

[4] Joachim Sachs, Petar Popovski, Andreas Höglund, David Gozalvez-Serrano, Peter Fertl, Mischa Dohler, and Takehiro Nakamura. *Machine-type communications*, page 77–106. Cambridge University Press, 2016.

[5] Elif Uysal, Onur Kaya, Anthony Ephremides, James Gross, Marian Codreanu, Petar Popovski, Mohamad Assaad, Gianluigi Liva, Andrea Munari, Beatriz Soret, Touraj Soleymani, and Karl Henrik Johansson. Semantic communications in networked systems: A data significance perspective. *IEEE Network*, 36(4):233–240, Jul 2022.

[6] Sanjit Kaul, Roy Yates, and Marco Gruteser. Real-time status: How often should one update? In *2012 Proceedings IEEE INFOCOM*, pages 2731–2735. IEEE, 2012.

[7] Roy D Yates, Yin Sun, D Richard Brown, Sanjit K Kaul, Eytan Modiano, and Sennur Ulukus. Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, 39(5):1183–1210, 2021.

[8] Yin Sun, Igor Kadota, Rajat Talak, and Eytan Modiano. Age of information: A new metric for information freshness. *Synthesis Lectures on Communication Networks*, 12(2):1–224, 2019.

[9] Antzela Kosta, Nikolaos Pappas, and Vangelis Angelakis. Age of information: A new concept, metric, and tool. *Foundations and Trends in Networking*, 12(3):162–259, 2017.

[10] Maice Costa, Marian Codreanu, and Anthony Ephremides. Age of information with packet management. In *2014 IEEE International Symposium on Information Theory*, pages 1583–1587, 2014.

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

[12] Maice Costa, Marian Codreanu, and Anthony Ephremides. On the age of information in status update systems with packet management. *IEEE Transactions on Information Theory*, 62(4):1897–1910, 2016.

[13] Yoshiaki Inoue, Hiroyuki Masuyama, Tetsuya Takine, and Toshiyuki Tanaka. A general formula for the stationary distribution of the age of information and its application to single-server queues. *IEEE Transactions on Information Theory*, 65(12):8305–8324, 2019.

[14] Clement Kam, Sastry Kompella, and Anthony Ephremides. Age of information under random updates. In *2013 IEEE International Symposium on Information Theory*, pages 66–70. IEEE, 2013.

[15] Clement Kam, Sastry Kompella, Gam D Nguyen, and Anthony Ephremides. Effect of message transmission path diversity on status age. *IEEE Transactions on Information Theory*, 62(3):1360–1374, 2015.

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

[17] Yin Sun, Eiif Uysal-Biyikoglu, and Sastry Kompella. Age-optimal updates of multiple information flows. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 136–141. IEEE, 2018.

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

[19] Baran Tan Bacinoglu and Elif Uysal-Biyikoglu. Scheduling status updates to minimize age of information with an energy harvesting sensor. In *2017 IEEE international symposium on information theory (ISIT)*, pages 1122–1126. IEEE, 2017.

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

[21] Ahmed Arafa, Jing Yang, Sennur Ulukus, and H Vincent Poor. Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies. *IEEE Transactions on Information Theory*, 66(1):534–556, 2019.

[22] Shahab Farazi, Andrew G Klein, and D Richard Brown. Average age of information for status update systems with an energy harvesting server. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 112–117. IEEE, 2018.

[23] Tasmeen Zaman Ornee and Yin Sun. Sampling and remote estimation for the ornstein-uhlenbeck process through queues: Age of information and beyond. *IEEE/ACM Transactions on Networking*, 29(5):1962–1975, 2021.

[24] Yin Sun and Benjamin Cyr. Sampling for data freshness optimization: Non-linear age functions. *Journal of Communications and Networks*, 21(3):204–219, 2019.

[25] Yin Sun, Yury Polyanskiy, and Elif Uysal. Sampling of the wiener process for remote estimation over a channel with random delay. *IEEE Transactions on Information Theory*, 66(2):1118–1135, 2019.

[26] Parimal Parag, Austin Taghavi, and Jean-Francois Chamberland. On real-time status updates over symbol erasure channels. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2017.

[27] Qing He, Di Yuan, and Anthony Ephremides. Optimal link scheduling for age minimization in wireless systems. *IEEE Transactions on Information Theory*, 64(7):5381–5394, 2017.

[28] Ali Maatouk, Mohamad Assaad, and Anthony Ephremides. On the age of information in a csma environment. *IEEE/ACM Transactions on Networking*, 28(2):818–831, 2020.

[29] Igor Kadota, Abhishek Sinha, and Eytan Modiano. Optimizing age of information in wireless networks with throughput constraints. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 1844–1852. IEEE, 2018.

[30] Changhee Joo and Atilla Eryilmaz. Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis. *IEEE/ACM transactions on networking*, 26(6):2556–2568, 2018.

[31] Vishrant Tripathi, Nicholas Jones, and Eytan Modiano. Fresh-csma: A distributed protocol for minimizing age of information. *arXiv preprint arXiv:2212.03087*, 2022.

[32] Tanya Shreedhar, Sanjit K Kaul, and Roy D Yates. An age control transport protocol for delivering fresh updates in the internet-of-things. In *2019 IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pages 1–7. IEEE, 2019.

[33] Tanya Shreedhar, Sanjit K Kaul, and Roy D Yates. Acp+: An age control protocol for the internet. *arXiv preprint arXiv:2210.12539*, 2022.

[34] Vishrant Tripathi, Igor Kadota, Ezra Tal, Muhammad Shahir Rahman, Alexander Warren, Sertac Karaman, and Eytan Modiano. Wiswarm: Age-of-information-based wireless networking for collaborative teams of uavs. *arXiv preprint arXiv:2212.03298*, 2022.

[35] Igor Kadota, Muhammad Shahir Rahman, and Eytan Modiano. Wifresh: Age-of-information from theory to implementation. In *2021 International Conference on Computer Communications and Networks (ICCCN)*, pages 1–11. IEEE, 2021.

[36] Ali Maatouk, Saad Kriouile, Mohamad Assaad, and Anthony Ephremides. The age of incorrect information: A new performance metric for status updates. *IEEE/ACM Transactions on Networking*, 28(5):2215–2228, 2020.

[37] Ali Maatouk, Mohamad Assaad, and Anthony Ephremides. The age of incorrect information: An enabler of semantics-empowered communication. *IEEE Transactions on Wireless Communications*, 2022.

[38] Clement Kam, Sastry Kompella, and Anthony Ephremides. Age of incorrect information for remote estimation of a binary markov source. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6. IEEE, 2020.

[39] Saad Kriouile and Mohamad Assaad. Minimizing the age of incorrect information for real-time tracking of markov remote sources. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2978–2983. IEEE, 2021.

[40] Saad Kriouile and Mohamad Assaad. Minimizing the age of incorrect information for unknown markovian source. *arXiv preprint arXiv:2210.09681*, 2022.

[41] Subham Saha, Harkirat Singh Makkar, Vineeth Bala Sukumaran, and Chandra R. Murthy. On the relationship between mean absolute error and age of incorrect information in the estimation of a piecewise linear signal over noisy channels. *IEEE Communications Letters*, 26(11):2576–2580, 2022.

[42] Bhavya Joshi, Rajshekhar V Bhat, B. N. Bharath, and Rahul Vaze. Minimization of age of incorrect estimates of autoregressive markov processes. In *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, pages 1–8, 2021.

[43] Qingyu Liu, Chengzhang Li, Y. Thomas Hou, Wenjing Lou, Jeffrey H. Reed, and Sastry Kompella. Ao$^2$I: Minimizing age of outdated information to improve freshness in data collection. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, page 1359–1368, London, United Kingdom, May 2022. IEEE.

[44] Jhelum Chakravorty and Aditya Mahajan. Fundamental limits of remote estimation of autoregressive markov processes under communication constraints. *IEEE Transactions on Automatic Control*, 62(3):1109–1124, 2016.

[45] Ashutosh Nayyar, Tamer Başar, Demosthenis Teneketzis, and Venugopal V. Veeravalli. Optimal strategies for communication and remote estimation with an energy harvesting sensor. *IEEE Transactions on Automatic Control*, 58(9):2246–2260, 2013.

[46] Gabriel M Lipsa and Nuno C Martins. Remote state estimation with communication costs for first-order lti systems. *IEEE Transactions on Automatic Control*, 56(9):2013–2025, 2011.

[47] Orhan C Imer and Tamer Basar. Optimal estimation with limited measurements. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 1029–1034. IEEE, 2005.

[48] Yutao Chen and Anthony Ephremides. Minimizing age of incorrect information for unreliable channel with power constraint. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2021.

[49] Baran Tan Bacinoglu, Yin Sun, Elif Uysal, and Volkan Mutlu. Optimal status updating with a finite-battery energy harvesting source. *Journal of Communications and Networks*, 21(3):280–294, 2019.

[50] Linn I Sennott. Constrained average cost markov decision chains. *Probability in the Engineering and Informational Sciences*, 7(1):69–83, 1993.

[51] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, USA, 3rd edition, 2009.

[52] Linn I Sennott. On computing average cost optimal policies with application to routing to parallel queues. *Mathematical methods of operations research*, 45(1):45–62, 1997.

[53] Jack J Dongarra, Iain S Duff, Danny C Sorensen, and Henk A Van der Vorst. *Numerical linear algebra for high-performance computers*. SIAM, 1998.

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

[55] Yu-Pin Hsu. Age of information: Whittle index for scheduling stochastic arrivals. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 2634–2638. IEEE, 2018.

[56] Vishrant Tripathi and Eytan Modiano. A whittle index approach to minimizing functions of age of information. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1160–1167. IEEE, 2019.

[57] Ali Maatouk, Saad Kriouile, Mohamad Assad, and Anthony Ephremides. On the optimality of the whittle's index policy for minimizing the age of information. *IEEE Transactions on Wireless Communications*, 20(2):1263–1277, 2020.

[58] Jingzhou Sun, Zhiyuan Jiang, Bhaskar Krishnamachari, Sheng Zhou, and Zhisheng Niu. Closed-form whittle's index-enabled random access for timely status update. *IEEE Transactions on Communications*, 68(3):1538–1551, 2019.

[59] Gam D Nguyen, Sastry Kompella, Clement Kam, and Jeffrey E Wieselthier. Information freshness over a markov channel: The effect of channel state information. *Ad Hoc Networks*, 86:63–71, 2019.

[60] Rajat Talak, Sertac Karaman, and Eytan Modiano. Optimizing age of information in wireless networks with perfect channel state information. In *2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 1–8. IEEE, 2018.

[61] Ling Shi, Peng Cheng, and Jiming Chen. Optimal periodic sensor scheduling with limited resources. *IEEE Transactions on Automatic Control*, 56(9):2190–2195, 2011.

[62] Alex S Leong, Subhrakanti Dey, and Daniel E Quevedo. Sensor scheduling in variance based event triggered estimation with packet drops. *IEEE Transactions on Automatic Control*, 62(4):1880–1895, 2016.

[63] Yilin Mo, Emanuele Garone, Alessandro Casavola, and Bruno Sinopoli. Stochastic sensor scheduling for energy constrained estimation in multi-hop wireless sensor networks. *IEEE Transactions on Automatic Control*, 56(10):2489–2495, 2011.

[64] Yutao Chen and Anthony Ephremides. Scheduling to minimize age of incorrect information with imperfect channel state information. *Entropy*, 23(12):1572, 2021.

[65] John Gittins, Kevin Glazebrook, and Richard Weber. *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

[66] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25(A):287–298, 1988.

[67] Richard R Weber and Gideon Weiss. On an index policy for restless bandits. *Journal of applied probability*, 27(3):637–648, 1990.

[68] Kevin D Glazebrook, Diego Ruiz-Hernandez, and Christopher Kirkbride. Some indexable families of restless bandit problems. *Advances in Applied Probability*, 38(3):643–672, 2006.

[69] Maialen Larrañaga. *Dynamic control of stochastic and fluid resource-sharing systems*. PhD thesis, INP Toulouse, September 2015.

[70] Dimitris Bertsimas and José Niño-Mora. Restless bandits, linear programming relaxations, and a primal-dual index heuristic. *Operations Research*, 48(1):80–90, 2000.

[71] Michael L. Littman, Thomas L. Dean, and Leslie Pack Kaelbling. On the complexity of solving markov decision problems. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 394–402, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

[72] Ina Maria Verloop. Asymptotically optimal priority policies for indexable and nonindexable restless bandits. *The Annals of Applied Probability*, 26(4):1947–1995, 2016.

[73] Yin Sun, Yury Polyanskiy, and Elif Uysal-Biyikoglu. Remote estimation of the wiener process over a channel with random delay. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 321–325, 2017.

[74] Tasmeen Zaman Ornee and Yin Sun. Sampling for remote estimation through queues: Age of information and beyond. In *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pages 1–8. IEEE, 2019.

[75] Clement Kam, Sastry Kompella, Gam D Nguyen, Jeffrey E Wieselthier, and Anthony Ephremides. Towards an effective age of information: Remote estimation of a markov source. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 367–372. IEEE, 2018.

[76] Yutao Chen and Anthony Ephremides. Analysis of age of incorrect information under generic transmission delay. *arXiv preprint arXiv:2212.14381*, 2022.

[77] Yutao Chen and Anthony Ephremides. Minimizing age of incorrect information over a channel with random delay. *arXiv preprint arXiv:2301.06150*, 2023.

[78] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[79] Jon Postel. Internet protocol. *RFC*, 791:1–51, 1981.

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

[81] Ronald A Howard. *Dynamic programming and markov processes*. John Wiley, 1960.

[82] Yutao Chen and Anthony Ephremides. Preempting to minimize age of incorrect information under random delay. *arXiv preprint arXiv:2209.14254*, 2022.

[83] Roy D Yates, Philippe Ciblat, Aylin Yener, and Micheie Wigger. Age-optimal constrained cache updating. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 141–145. IEEE, 2017.