Minimizing Age of Incorrect Information in the Presence of Timeout

Yutao Chen and Anthony Ephremides
Department of Electrical and Computer Engineering, University of Maryland

Abstract

We consider a slotted-time system with a transmitter-receiver pair. In the system, a transmitter observes a dynamic source and sends updates to a remote receiver through a communication channel. We assume that the channel is error-free but suffers a random delay. Moreover, when an update has been transmitted for too long, the transmission will be terminated immediately, and the update will be discarded. We assume the maximum transmission time is predetermined and is not controlled by the transmitter. The receiver will maintain estimates of the current state of the dynamic source using the received updates. In this paper, we adopt the Age of Incorrect Information (AoII) as the performance metric and investigate the problem of optimizing the transmitter’s action in each time slot to minimize AoII. We first characterize the optimization problem using Markov Decision Process and evaluate the performance of some canonical transmission policies. Then, by leveraging the policy improvement theorem, we prove that, under a simple and easy-to-verify condition, the optimal policy for the transmitter is the one that initiates a transmission whenever the channel is idle and AoII is not zero. Lastly, we take the case where the transmission time is geometrically distributed as an example. For this example, we verify the condition numerically and provide numerical results that highlight the performance of the optimal policy.

I. INTRODUCTION

Communication systems are used in all aspects of our lives and play an increasingly important role. Consequently, communication systems are being asked to play more roles than just the dissemination of words, sounds, and images. With the widespread deployment of communication systems and the continuous expansion of their communication purposes, we have to demand higher performance from the communications systems. Meanwhile, we wonder whether traditional metrics such as throughput and latency could continue to meet such demands. One of the major drawbacks of such traditional metrics is that they treat each update equally and
ignore that not every update can provide the recipient with equally important information for communication purposes. In view of this, researchers seek to reconsider existing communication paradigms and look for new ones, among which semantic communication is an important attempt. The semantics of information is formally defined in [1] as the significance of the messages relative to the purpose of the data exchange. Then, semantic communication is regarded as "the provisioning of the right piece of information to the right point of computation (or actuation) at the right point in time". Different from the classical metrics in data communication, semantic metrics should incorporate one or more of the following attributes: Freshness which captures the freshness of information, Relevance which captures the amount of change in the process since the previous sample, and Value which captures the difference between the benefit of having this sample and the cost of its transmission. One of the most successful examples is the Age of Information (AoI) introduced in [2]. AoI captures the freshness of information by tracking the time that has elapsed since the last received update was generated, which results in different treatments for different updates. For example, when the update can significantly reduce the age on the recipient side, it will become important and worth the extra resources to transmit. After the introduction, AoI has attracted extensive attention [3]–[5]. 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 [6]. Noting the shortcomings of AoI in remote estimation, researchers in [7] introduced the Age of Incorrect Information (AoII).

AoII considers both the timeliness and accuracy of information. More specifically, AoII combines the degree of information mismatch between the recipient and the source and the aging process of the mismatched information. According to the definition given in [7], AoII captures the aging process of the conflicting information through a time penalty function, which quantifies the time elapsed since the receiver last obtained the perfect information about the source. The mismatch between the recipient’s information and the source is captured by the information penalty function, which quantifies the degree of information mismatch between the two. Through the introduction of the penalty functions, AoII can adapt the various system and communication objectives by choosing different penalty functions.

Several works have been done since the introduction of AoII. In [7], the authors investigate the minimization of AoII when there is a limit on the average number of transfers allowed. Then, the authors extend the results to the case of the generic time penalty function in [8]. In [9], the
authors investigate a similar system setting, but the AoII considers the quantified information mismatch between the source and the recipient. AoII in the context of scheduling is another critical problem. In scheduling problems, a base station observes multiple events and needs to select a part of the users to update. Under these general settings, [10] investigates the problem of minimizing AoII when the channel state information is available and the time penalty function is generic. The authors of [11] consider a similar system, but the base station cannot know the states of the events before the transmission. In real-life applications, communication channel usually suffers random delays due to various influences. Under this system setup, the authors of [12] compare three performance metrics: AoII, AoI, and real-time error through extensive numerical results. [13] consider the problem of minimizing AoII is such system and assume that the update will always be delivered. In this paper, we consider a similar system setup, but we assume that the transmission will be terminated when it takes too long. The system with a random delay communication channel has also been studied in the context of remote estimation and AoI [14]–[17]. However, the problem considered in this paper is very different as AoII is a combination of age-based metrics framework and error-based metrics framework.

The main contributions of this paper can be summarized as follows. 1) We study the system where the communication channel suffers a random delay and a maximum transmission time is set to prevent the updates from occupying the channel for too long. 2) We investigate the problem of minimizing the Age of Incorrect Information under generic delay distribution. 3) We theoretically calculate the performance of several canonical policies in the considered system. 4) We find the optimal policy and prove its optimality theoretically.

The remainder of this paper is organized in the following way. In Section II, we introduce the system settings and formulate the optimization problem. Then, in Section III, we characterize the problem using Markov Decision Process and show that the optimal policy exists. Section IV evaluates the performance of three canonical policies, laying the foundations for the optimality proof detailed in Section V. Finally, in Section VI we consider a specific instance of the system and provide the numerical results that corroborate our theoretical results and highlight the performance of the optimal policy.
II. SYSTEM OVERVIEW

A. System Model

We consider a transmitter-receiver pair in a slotted-time system. In the system, a transmitter observes a dynamic source and needs to decide when to send updates to a remote receiver so that the receiver can maintain a good estimate 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 Markov chain is illustrated in Fig. 1. 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 will be discarded upon the arrival of a new one. Then, the transmitter will decide 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 must wait until the current transmission finishes. The updates will be transmitted through a communication channel that is error-free but suffers a random delay. More precisely, the update will not be corrupted during the transmission, but each transmission will take a random amount of time $T \in \mathbb{N}^*$. To prevent an update from occupying the channel for too long, we preset a maximum transmission time $t_{\text{max}}$. When an update has been transmitted for $t_{\text{max}}$ time slot (i.e., $T = t_{\text{max}}$), the transmission will be terminated immediately, and the update will be discarded. Then, the transmitter can decide whether to transmit a new update in the next time slot. We assume $t_{\text{max}}$ is predetermined and is not controlled by the transmitter. We denote by $p_t \triangleq \Pr(T = t)$ the probability distribution of the transmission time of an update. We assume $T$ is independent and identically distributed and $p_t$ satisfies the following two equations.

$$p_t \geq 0, \quad t \geq 1,$$

$$\sum_{t=1}^{\infty} p_t = 1.$$
The receiver maintains an estimate of the current state of the dynamic source, and it will modify its estimate every time a new update is received. We denote by $\hat{X}_k$ the receiver’s estimate at time slot $k$. According to [17], the best estimator when $p \leq \frac{1}{2}$ is simply the last received update. When $p > \frac{1}{2}$, the optimal estimator depends on the transmission time. This paper considers the case of $p \leq \frac{1}{2}$ and the corresponding best estimator. The results can be extended to the case of $p > \frac{1}{2}$ by adopting the corresponding best estimator. The receiver will use ACK/NACK packets to inform the transmitter of its reception of the new update. When ACK is received, the transmitter knows that the receiver’s estimate changed to the last sent update. When NACK is received, the transmitter knows that the receiver’s estimate did not change. As in [7], we assume that the transmitter will receive the ACK/NACK packets reliably and instantaneously. In this way, the transmitter always knows the current estimate at the receiver side. The system model is illustrated in Fig. 2.

In a typical time slot, the transmitter receives an update 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, nothing happens. Otherwise, the update will be transmitted through the communication channel. The transmission of the update takes a random amount of time. When the transmission time of the current update reaches $t_{\text{max}}$, the transmission will be terminated immediately, and the update will be discarded. When an update is delivered, the receiver will modify its estimation based on the received update and send a ACK packet to inform the transmitter of its reception of the update.

B. Age of Incorrect Information

The system adopts the Age of Incorrect Information (AoII) as the performance metric. We first define

$$U_k \triangleq \max \{ h : h \leq k, X_h = \hat{X}_h \}.$$
We notice that $U_k$ is simply the last time instant before time $k$ (including $k$) that the receiver’s estimate is correct. Leveraging the definition of $U_k$, AoII at time slot $k$ can be written as

$$\Delta_{AoII}(X_k, \hat{X}_k, k) = \sum_{h=U_k+1}^{k} \left( g(X_h, \hat{X}_h) \cdot F(h - U_k) \right),$$  \hspace{1cm} (1)$$

where $g(X_k, \hat{X}_k)$ is the information penalty function. $F(k) \triangleq f(k) - f(k-1)$ where $f(k)$ is the time penalty function. In this paper, we choose $g(X_k, \hat{X}_k) = |X_k - \hat{X}_k|$ and $f(k) = k$. Hence, $F(k) = 1$ and $g(X_k, \hat{X}_k) \in \{0, 1\}$, given that the dynamic source has only two states. Then, equation (1) can be simplified as

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

As we will see later, the dynamic of $\Delta_k$ is critical. More precisely, we need to characterize the value of $\Delta_{k+1}$ using $\Delta_k$. To this end, we capture the dynamic of $\Delta_k$ with the following two cases.

- When the receiver’s estimate is correct at time $k + 1$, we have $U_{k+1} = k + 1$. Then, by definition, $\Delta_{k+1} = 0$.
- When the receiver’s estimate is incorrect at time $k + 1$, we have $U_{k+1} = U_k$. Then, by definition, $\Delta_{k+1} = k + 1 - U_k = \Delta_k + 1$.

Combining together, we have

$$\Delta_{k+1} = \mathbb{1}\{U_{k+1} \neq k + 1\} \cdot (\Delta_k + 1),$$ \hspace{1cm} (2)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. A sample path of $\Delta_k$ is shown in Fig. 3.

C. Problem Formulation

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

$$\arg \min_{\phi \in \Phi} \lim_{K \to \infty} \frac{1}{K} \mathbb{E}_\phi \left( \sum_{k=0}^{K-1} \Delta_k \right),$$ \hspace{1cm} (3)$$

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

Definition 1 (Optimal policy). A policy is said to be optimal if it yields the minimal expected AoII.
Fig. 3: A sample path of $\Delta_k$. In the figure, $T_i$ and $D_i$ are the transmission time and the delivery time of the $i$-th update, respectively. At $T_1$, the transmitted update is $X_3$. The estimate at time slot 6 (i.e., $\hat{X}_6$) changes due to the reception of the update transmitted at $T_2$. Note that the transmission decisions in the sample path are taken randomly.

In the next section, we will characterize the problem reported in (3) using a Markov Decision Process (MDP) and show that the optimal policy exists.

III. MARKOV DECISION PROCESS CHARACTERIZATION

We notice that the problem reported in (3) can be characterized by an infinite horizon with average cost Markov Decision Process (MDP) $\mathcal{M}$, which consists of the following components.

- The state space $\mathcal{S}$. The state can be represented by the triplet $s = (\Delta, t, i)$ where $\Delta \in \mathbb{N}^0$ is the current AoII. $t \in \{0, 1, ..., t_{\text{max}} - 1\}$ indicates the time the transmission has been in progress. We define $t = 0$ if there is no transmission in progress. The last element $i \in \{-1, 0, 1\}$ indicates the state of the channel. $i = -1$ if the channel is idle. $i = 0$ if the channel is busy, and the transmitting update is the same as receiver’s estimate. Otherwise, $i = 1$. For the remainder of this paper, we will use $s$ and $(\Delta, t, i)$ to represent the state interchangeably.

- The feasible action $\mathcal{A}$. When $i = -1$, the feasible actions are $a \in \{0, 1\}$ where $a = 0$ if the transmitter decides not to send a new update and $a = 1$ otherwise. When $i \neq -1$, the feasible action is $a = 0$.

- The state transition probabilities $\mathcal{P}$. The probability that action $a$ in state $s$ leads to state $s'$ is denoted by $P[s' | s, a]$. The dynamics of $P[s' | s, a]$ are detailed in the following section.
The instant cost $C$. The instant cost for being at state $s$ is $C(s) = \Delta$.

**Remark 1.** Note that, according to the definitions of $t$ and $i$, $i = -1$ if and only if $t = 0$. In this case, the channel is idle.

Let $V(s)$ be the value function associated with state $s \in S$. It is well known that value function satisfies the following Bellman equation.

$$V(s) + \theta = \min_{a \in A} \left\{ C(s) + \sum_{s' \in S} P[s'| s, a] V(s') \right\}, \tag{4}$$

where $\theta$ is the expected AoII achieved by the optimal policy.

**A. State Transition Probability**

In the following, we tackle the state transition probability $P[s'| s, a]$. We notice that it is sufficient to characterize the system’s state in the next time slot and the corresponding probability using the current state and the transmitter’s action. Before diving deep into the probabilities, we first calculate the key probability $Pr(T > t + 1 \mid t)$, which is the probability that the update will not be delivered in the next time slot, given that the transmission has been in progress for $t$ time slots. Then, $Pr(T > t + 1 \mid t)$ is given by

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

where $P_t \triangleq Pr(T \leq t)$. Leveraging this probability, we can proceed with deriving the state transition probabilities. We first note that $\Delta$ will evolve according to (2). Hence, in the following, we will omit the discussion on the evolution of $\Delta$. Then, we distinguish between the following cases.

- $s = (0, 0, -1)$. In this case, the channel is idle. Hence, the feasible action is $a \in \{0, 1\}$. We start with the case of $a = 0$. In this case, the transmitter decides to stay idle. Hence, $i$, $t$, and the receiver’s estimate will not change.

  $$Pr[(0, 0, -1) \mid (0, 0, -1), a = 0] = 1 - p.$$  

  $$Pr[(1, 0, -1) \mid (0, 0, -1), a = 0] = p.$$  

Then, we consider the case of $a = 1$. In this case, the transmitter decides to initiate a new transmission, and the update will be delivered after a random amount of time $T$. When $T > 1$, the update will not be delivered in the next time slot, and the channel will be busy.
in the next time slot. Since the transmission started when \( \Delta = 0, i = 0 \). As the transmission continues, \( t \) will increase by one, and the receiver’s estimate will not change.

\[
Pr[(0, 1, 0) \mid (0, 0, -1), a = 1] = Pr(T > 1 \mid 0)(1 - p).
\]

\[
Pr[(1, 1, 0) \mid (0, 0, -1), a = 1] = Pr(T > 1 \mid 0)p.
\]

When \( T = 1 \), the update will be delivered at the next time slot. Hence, \( t \) and \( i \) will reset to 0 and \(-1\), respectively. Since the transmission started when \( \Delta = 0 \), the transmitting update brings no new information to the receiver. Hence, the receiver’s estimate will not change.

\[
Pr[(0, 0, -1) \mid (0, 0, -1), a = 1] = Pr(T = 1 \mid 0)(1 - p).
\]

\[
Pr[(1, 0, -1) \mid (0, 0, -1), a = 1] = Pr(T = 1 \mid t)p.
\]

• \( s = (0, t, 0) \) where \( 1 \leq t \leq t_{\text{max}} - 2 \). In this case, the channel is busy. Hence, the transmitter must stay idle. When the update does not arrive at the next time slot, which happens with probability \( Pr(T > t + 1 \mid t) \), \( i \) will not change since both the transmitting update and the receiver’s estimate remain the same. \( t \) will increase by one as the transmission continues.

\[
Pr[(0, t + 1, 0) \mid (0, t, 0)] = Pr(T > t + 1 \mid t)(1 - p).
\]

\[
Pr[(1, t + 1, 0) \mid (0, t, 0)] = Pr(T > t + 1 \mid t)p.
\]

When the update arrives at the next time slot, \( t \) and \( i \) will reset to 0 and \(-1\), respectively. Since the transmitting update is the same as the receiver’s estimate, it brings no new information to the receiver. Hence, the receiver’s estimate will not change.

\[
Pr[(0, 0, -1) \mid (0, t, 0)] = Pr(T = t + 1 \mid t)(1 - p).
\]

\[
Pr[(1, 0, -1) \mid (0, t, 0)] = Pr(T = t + 1 \mid t)p.
\]

• \( s = (0, t, 1) \) where \( 1 \leq t \leq t_{\text{max}} - 2 \). The analysis is very similar to the above case, except that the receiver’s estimate will flip when the update arrives.

\[
Pr[(0, t + 1, 1) \mid (0, t, 1)] = Pr(T > t + 1 \mid t)(1 - p).
\]

\[
Pr[(1, t + 1, 1) \mid (0, t, 1)] = Pr(T > t + 1 \mid t)p.
\]

\[
Pr[(0, 0, -1) \mid (0, t, 1)] = Pr(T = t + 1 \mid t)p.
\]

\[
Pr[(1, 0, -1) \mid (0, t, 1)] = Pr(T = t + 1 \mid t)(1 - p).
\]
\( s = (0, t_{\text{max}} - 1, 0) \). In this case, the transmission will be timed out with probability \( Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1) \), in which case the estimate at the receiver side will not change. When the update arrives, which happens with probability \( Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1) \), the estimate will also not change since the received update is the same as the estimate. Moreover, in both cases, \( t \) and \( i \) will reset to 0 and \(-1\), respectively.

\[
Pr[(0, 0, -1) \mid (0, t_{\text{max}} - 1, 0)] = 1 - p.
\]

\[
Pr[(1, 0, -1) \mid (0, t_{\text{max}} - 1, 0)] = p.
\]

- \( s = (0, t_{\text{max}} - 1, 1) \). The transmission will be timed out with probability \( Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1) \), in which case the estimate at the receiver side will not change. However, when the update arrives, which happens with probability \( Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1) \), the estimate will flip since the estimate is different from the received update. Moreover, in both case, \( t \) and \( i \) will reset to 0 and \(-1\), respectively.

\[
Pr[(0, 0, -1) \mid (0, t_{\text{max}} - 1, 1)] = Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1)p + Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1)(1 - p).
\]

\[
Pr[(1, 0, -1) \mid (0, t_{\text{max}} - 1, 1)] = Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1)(1 - p) + Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1)p.
\]

- \( s = (\Delta, 0, -1) \) where \( \Delta > 0 \). In this case, the channel is idle. Hence, the feasible action is \( a \in \{0, 1\} \). We first consider the case of \( a = 0 \). In this case, the transmitter decides not to initiate a new transmission. Hence, \( i, t \), and the receiver’s estimate will not change.

\[
Pr[(\Delta + 1, 0, -1) \mid (\Delta, 0, -1), a = 0] = 1 - p.
\]

\[
Pr[(0, 0, -1) \mid (\Delta, 0, -1), a = 0] = p.
\]

Then, we consider the case of \( a = 1 \). In this case, the transmitter decides to start a new transmission, and the update will be delivered after a random amount of time \( T \). When \( T > 1 \), the update will not be delivered in the next time slot. As the transmission continues, \( t \) will increase by one, and the receiver’s estimate will not change. Since the transmission happens when \( \Delta > 0 \), \( i = 1 \).

\[
Pr[(\Delta + 1, 1, 1) \mid (\Delta, 0, -1), a = 1] = Pr(T > 1 \mid 0)(1 - p).
\]

\[
Pr[(0, 1, 1) \mid (\Delta, 0, -1), a = 1] = Pr(T > 1 \mid 0)p.
\]
When $T = 1$, the update will be delivered in the next time slot. Hence, $t$ and $i$ will reset to 0 and $-1$, respectively. Since the transmission started when $\Delta > 0$, the receiver’s estimate will flip due to the reception of the new update.

$$
Pr[(\Delta + 1, 0, -1) \mid (\Delta, 0, -1), a = 1] = Pr(T = 1 \mid 0)p.
$$

$$
Pr[(0, 0, -1) \mid (\Delta, 0, -1), a = 1] = Pr(T = 1 \mid 0)(1 - p).
$$

- $s = (\Delta, t, 0)$ where $\Delta > 0$ and $1 \leq t \leq t_{max} - 2$. In this case, the channel is busy. Hence, the transmitter must stay idle. When the update does not arrive in the next time slot, which happens with probability $Pr(T > t + 1 \mid t)$, $i$ will not change since both the transmitting update and the receiver’s estimate remain the same. $t$ will increase by one as the transmission continues.

$$
Pr[(\Delta + 1, t + 1, 0) \mid (\Delta, t, 0)] = Pr(T > t + 1 \mid t) (1 - p).
$$

$$
Pr[(0, t + 1, 0) \mid (\Delta, t, 0)] = Pr(T > t + 1 \mid t)p.
$$

When the update arrives in the next time slot, which happens with probability $Pr(T = t + 1 \mid t)$, $t$ and $i$ will reset to 0 and $-1$, respectively. Meanwhile, the transmitting update brings no new information to the receiver. Hence, the receiver’s estimate will not change.

$$
Pr[(\Delta + 1, t, 1) \mid (\Delta, t, 0)] = Pr(T = t + 1 \mid t)(1 - p).
$$

$$
Pr[(0, 0, t, 1) \mid (\Delta, t, 0)] = Pr(T = t + 1 \mid t)p.
$$

- $s = (\Delta, t, 1)$ where $\Delta > 0$ and $1 \leq t \leq t_{max} - 2$. The analysis is very similar to the above case, except that the receiver’s estimate will flip when the update arrives.

$$
Pr[(\Delta + 1, t + 1, 1) \mid (\Delta, t, 1)] = Pr(T > t + 1 \mid t) (1 - p).
$$

$$
Pr[(0, t + 1, 1) \mid (\Delta, t, 1)] = Pr(T > t + 1 \mid t)p.
$$

$$
Pr[(\Delta + 1, 0, -1) \mid (\Delta, t, 1)] = Pr(T = t + 1 \mid t)p.
$$

$$
Pr[(0, 0, -1) \mid (\Delta, t, 1)] = Pr(T = t + 1 \mid t)(1 - p).
$$

- $s = (\Delta, t_{max} - 1, 0)$ where $\Delta > 0$. The transmission will be timed out with probability $P(T > t_{max} \mid t_{max} - 1)$, in which case the receiver’s estimate will not change. When the update arrives, which happens with probability $P(T = t_{max} \mid t_{max} - 1)$, the estimate will
also not change since the received update is the same as the receiver’s estimate. Moreover, in both cases, \( t \) and \( i \) will reset to 0 and \(-1\), respectively.

\[
Pr[(0, 0, -1) \mid (\Delta, t_{\text{max}} - 1, 0)] = p.
\]
\[
Pr[(\Delta + 1, 0, -1) \mid (\Delta, t_{\text{max}} - 1, 0)] = 1 - p.
\]

- \( s = (\Delta, t_{\text{max}} - 1, 1) \) where \( \Delta > 0 \). The analysis is very similar to the above case, except that the receiver’s estimate will flip when the update arrives.

\[
Pr[(0, 0, -1) \mid (\Delta, t_{\text{max}} - 1, 1)] = Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1)(1 - p) + \Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1)p.
\]
\[
Pr[(\Delta + 1, 0, -1) \mid (\Delta, t_{\text{max}} - 1, 1)] = Pr(T = t_{\text{max}} \mid t_{\text{max}} - 1)p + \Pr(T > t_{\text{max}} \mid t_{\text{max}} - 1)(1 - p).
\]

Note that transitions not discussed above have a probability of 0. Then, combing together, we fully characterized the state transition probabilities \( \mathcal{P} \).

**B. Existence of Optimal Policy**

In this section, we will prove that the optimal policy for \( \mathcal{M} \) exists. To start with, we introduce the infinite horizon \( \gamma \)-discounted cost of \( \mathcal{M} \), where \( 0 < \gamma < 1 \) is a discount factor. Then, the expected \( \gamma \)-discounted cost under policy \( \phi \) can be calculated as

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

where \( s_t \) is the state of system at time \( t \). The quantity \( V_{\gamma}(. \mid s) \triangleq \inf_{\phi} V_{\phi, \gamma}(s) \) is defined as the best that can be achieved. Equivalently, \( V_{\gamma}(\cdot) \) is the value function associated with the infinite horizon \( \gamma \)-discounted MDP. Hence, \( V_{\gamma}(\cdot) \) satisfies the following Bellman equation.

\[
V_{\gamma}(s) = \min_{a \in A} \left\{ C(s) + \gamma \sum_{s' \in S} P[s' \mid s, a] V_{\gamma}(s') \right\}.
\]

We also define \( h_{\gamma}(s) \equiv V_{\gamma}(s) - V_{\gamma}(0) \) as the relative value function and choose the reference state \( 0 = (0, 0, -1) \).

We recall that \( V_{\gamma}(s) \) is the value function associated with the infinite horizon \( \gamma \)-discounted MDP. Hence, value iteration algorithm is a canonical algorithm to calculate \( V_{\gamma}(\cdot) \). Let \( V_{\gamma, \nu}(\cdot) \) be
the estimated value function at iteration \( \nu \) of value iteration. Then, the estimated value function is updated in the following way.

\[
V_{\gamma,\nu+1}(s) = \min_{a \in \mathcal{A}} \left\{ C(s) + \gamma \sum_{s' \in \mathcal{S}} P[s' \mid s, a] V_{\gamma,\nu}(s') \right\}.
\] (6)

**Lemma 1** (Convergence of value iteration). The estimated value function will converge to value function as the iteration reported in (6) goes to infinity. More precisely, \( \lim_{\nu \to \infty} V_{\gamma,\nu}(s) = V_\gamma(s) \) for all \( \gamma \) and \( s \).

**Proof.** According to Propositions 1 and 3 of [18], it is sufficient to show that \( V_\gamma(s) \) is finite for all \( \gamma \) and \( s \). The complete proof can be found in Appendix A.

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

**Lemma 2** (Monotonicity). \( V_\gamma(s) \) is increasing in \( \Delta \) when \( \Delta > 0 \).

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

Now, we proceed with showing the existence of the optimal policy for \( \mathcal{M} \). To this end, we first define the stationary policy.

**Definition 2** (Stationary policy). A stationary policy specifies a single action in each time slot.

**Theorem 1** (Existence of optimal policy). There exists a stationary policy \( \phi \) that is optimal for \( \mathcal{M} \). Moreover, the minimum expected AoII is independent of the initial state.

**Proof.** We show that \( \mathcal{M} \) satisfies the two conditions given in [18]. Then, the existence of the optimal policy is guaranteed by [18]. The complete proof can be found in Appendix C.

**Remark 2.** The proof of Theorem 1 is very similar to the one presented in [13].

IV. POLICY EVALUATION

In this section, we evaluate the performance of several canonical policies. We notice that the feasible action is \( a = 0 \) when the channel is busy. Therefore, a policy only needs to specify the action when the channel is idle. Moreover, the system’s dynamic under a policy can be
characterized by a discrete-time Markov chain. Let the state of the Markov chain be $s$. We notice that the transitions between states $s = (\Delta, t, i)$ with $t > 0$ are independent of the transmitter’s action. Hence, we define $\pi_\Delta$ being the steady-state probability of state $s = (\Delta, 0, -1)$, which is all we need to evaluate the performance. In the following subsection, we will characterize the transition probabilities between states $s = (\Delta, 0, -1)$.

A. Compact State Transition Probabilities

We denote by $P_{\Delta,\Delta'}(a)$ the probability that action $a$ at state $(\Delta, 0, -1)$ will lead to state $(\Delta', 0, -1)$, $P_{\Delta,\Delta'}^t(a)$ the probability that action $a$ at state $(\Delta, 0, -1)$ will lead to state $(\Delta', 0, -1)$ when the transmission takes $t$ time slot where $t \leq t_{\text{max}}$, and $P_{\Delta,\Delta'}^{t+}(a)$ the probability that action $a$ at state $(\Delta, 0, -1)$ will lead to state $(\Delta', 0, -1)$ when the update is discarded due to time out.

Note that when $a = 0$, the probabilities can be obtained directly from the results in Section III-A. Hence, in the following, we focus on the case of $a = 1$. To get the corresponding closed-form expressions, we first define $p(t)$ as the probability that the Markovian source will be in the same state after $t$ time slots. Since the Markov chain has two states and is symmetric, $p(t)$ is independent of the state and can be calculated by

$$p(t) = \left( \begin{array}{cc} 1 - p & p \\ p & 1 - p \end{array} \right)^t_{11}.$$

We define that $p(0) \triangleq 1$. Then, we can prove the following lemma.

**Lemma 3** (State transition probabilities). *The expressions of $P_{\Delta,\Delta'}^t(1)$ where $1 \leq t \leq t_{\text{max}}$, and $P_{\Delta,\Delta'}^{t+}(1)$ can be summarized as the following.*

$$P_{0,\Delta'}^t(1) = \begin{cases} p(t) & \Delta' = 0, \\ p(t-k)p(1-p)^{k-1} & \Delta' = k \in \{1, 2, ..., t\}, \\ 0 & \text{otherwise}. \end{cases}$$

$$P_{0,\Delta'}^{t+}(1) = \begin{cases} p(t_{\text{max}}) & \Delta' = 0, \\ p(t_{\text{max}}-k)p(1-p)^{k-1} & \Delta' = k \in \{1, 2, ..., t_{\text{max}}\}, \\ 0 & \text{otherwise}. \end{cases}$$
For $\Delta \geq 1$,

$$P_{\Delta,\Delta'}^{t}(1) = \begin{cases} 
    p(t) & \Delta' = 0, \\
    (1 - p^{(t-1)})(1 - p) & \Delta' = 1, \\
    (1 - p^{(t-k)})p^{2}(1 - p)^{k-2} & \Delta' = k \in \{2, 3, \ldots, t - 1\}, \\
    p(1 - p)^{t-1} & \Delta' = \Delta + t, \\
    0 & \text{otherwise}.
\end{cases}$$

$$P_{\Delta,\Delta'}^{t+}(1) = \begin{cases} 
    1 - p^{(t_{\text{max}})} & \Delta' = 0, \\
    (1 - p^{(t_{\text{max}}-k)})p(1 - p)^{k-1} & \Delta' = k \in \{1, 2, \ldots, t_{\text{max}} - 1\}, \\
    (1 - p)^{t_{\text{max}}} & \Delta' = \Delta + t_{\text{max}}, \\
    0 & \text{otherwise}.
\end{cases}$$

The above probabilities possess the following properties.

1) For each $t$, $0 \leq \Delta' \leq t - 1$, and $\Delta \geq 1$, $P_{\Delta,\Delta'}^{t}(1)$ and $P_{\Delta,\Delta'}^{t+}(1)$ are independent of $\Delta$.

2) For each $t$ and $\Delta \geq 1$, $P_{\Delta,\Delta+t}(1)$ and $P_{\Delta,\Delta+t_{\text{max}}}(1)$ are independent of $\Delta$.

**Proof.** The complete proof can be found in Appendix D. 

Then, $P_{\Delta,\Delta'}(1)$ can be calculated by

$$P_{\Delta,\Delta'}(1) = \sum_{t=1}^{t_{\text{max}}} p_{t}P_{\Delta,\Delta'}^{t}(1) + p_{t+}P_{\Delta,\Delta'}^{t+}(1),$$

Equation (7) can be expressed equivalently as the following.

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

Proposition 1 (Compact state transition probabilities).
where \( t' = \Delta' - \Delta \). Meanwhile, \( P_{\Delta, \Delta'}(1) \) possesses the following properties.

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

Proof. The complete proof can be found in Appendix E.

B. Expected Cost During Transmission

In this section, we investigate another important quantity \( C(\Delta, a) \), which is defined as the expected cost during the period from (including) the time when action \( a \) is taken at state \((\Delta, 0, -1)\) to (excluding) when the channel becomes idle for the first time after the action. We notice that \( C(\Delta, 0) = \Delta \). Hence, in the following, we focus on the case of \( a = 1 \).

**Proposition 2** (Expected cost). \( C(\Delta, 1) \) is given by the following expressions.

\[
C(\Delta, 1) = \sum_{t=1}^{t_{\text{max}}} p_t C^t(\Delta, 1) + p_{t_{\text{max}}} C^{t_{\text{max}}}(\Delta, 1),
\]

where

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

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

Proof. The complete proof can be found in Appendix F.

C. Considered Policies

Leveraging the results detailed in the previous sections, we can proceed with evaluating the performance of several canonical policies by calculating the resulting expected AoII. To start with, we first define three canonical policies.

**Definition 3** (Lazy policy). Under the lazy policy, the transmitter will never transmit any updates.

**Definition 4** (Zero wait policy). Under the zero wait policy, the transmitter will transmit a new update whenever the channel is idle.
**Definition 5** (Threshold policy). *Under the threshold policy, the transmitter will initiate a new transmission only when the current AoII is no less than threshold $\tau \in \mathbb{N}_0$ and the channel is idle. Let $a_\tau(s)$ be the action suggested by threshold policy with threshold $\tau$. Then, the action is determined by the following expression.*

$$a_\tau(s) = 1 \{\Delta \geq \tau, i = -1\}.$$  

**Remark 3.** When $\tau = 0$, the threshold policy reduces to the zero wait policy. When $\tau = \infty$, the threshold policy reduces to the lazy policy.

We notice that threshold policy can be fully characterized by the threshold $\tau$. In the following, we will use $\tau$ to represent threshold policy and $\bar{\Delta}_\tau$ to represent the expected AoII achieved by threshold policy $\tau$.

**Proposition 3** (Performance of lazy policy). *The expected AoII resulting from the adoption of the lazy policy is given by*  

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

*Proof.* We notice that a discrete-time Markov chain can characterize the dynamic of AoII under the lazy policy. Hence, we first calculate the stationary distribution of the induced Markov chain. Then, the expected AoII can be calculated using the stationary distribution. The complete proof can be found in Appendix G.

Let $ET$ be the expected transmission time of an update. Hence, $ET$ can be calculated as  

$$ET \triangleq \sum_{t=1}^{t_{\text{max}}} t p_t + t_{\text{max}} p_{t+}.$$  

We also define an auxiliary quantity $\Upsilon(\Delta, t) \triangleq p_t P_{t-\Delta,t-\Delta}(1) + p_{t+} P_{t+\Delta,t-\Delta}(1)$.

We notice from property 2 of Proposition 1 that $\Upsilon(\Delta, t)$ is independent of $\Delta$ when $\Delta \geq t_{\text{max}} + 1$.

Then, we have the following proposition.

**Proposition 4** (Performance of zero wait policy). *The expected AoII resulting from the adoption of the zero wait policy is given by*  

$$\bar{\Delta}_0 = \sum_{i=0}^{t_{\text{max}}} C(i, 1) \pi_i + \Sigma,$$  

where  

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

$$\pi_{\Delta} = \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1) \pi_i + P_{\Delta,\Delta}(1) \left(\frac{1}{ET} - \sum_{i=0}^{\Delta-1} \pi_i\right), \quad 1 \leq \Delta \leq t_{\text{max}} - 1.$$
\[ \pi_{t_{\text{max}}} = \sum_{i=0}^{t_{\text{max}}-1} P_{i,t_{\text{max}}}(1) \pi_i, \]

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

\[ \Pi_t = \sum_{i=t_{\text{max}}+1-t}^{t_{\text{max}}-1} \Upsilon(i + t, t)\pi_i + \Upsilon(t_{\text{max}} + t, t)\Pi, \quad 1 \leq t \leq t_{\text{max}}, \]

\[ \Pi = \sum_{i=0}^{t_{\text{max}}-1} \left( \sum_{k=0}^i P_{i,t_{\text{max}}+k}(1) \right) \pi_i, \]

\[ \Delta'_t = \sum_{i=1}^{t_{\text{max}}} p_i \left( \frac{t - t(1 - p)^i}{p} \right) + p_t \left( \frac{t - t(1 - p)^{t_{\text{max}}}}{p} \right), \quad 1 \leq t \leq t_{\text{max}}. \]

**Proof.** Similar to the proof of Proposition 3, we first calculate the stationary distribution of the Markov chain induced by the zero wait policy. Then, the expected AoII can be obtained using the stationary distribution. We notice that the induced Markov chain has infinitely many states. To overcome the infinity, we introduce two auxiliary quantities \( \Pi \triangleq \sum_{i=1}^{t_{\text{max}}} \pi_i \) and \( \Sigma \triangleq \sum_{i=t_{\text{max}}+1}^{t_{\text{max}}+1} C(i, 1)\pi_i \). Leveraging these two quantities, we can obtain the closed-form expression of the expected AoII achieved by the zero wait policy. The complete proof can be found in Appendix H.

Next, we calculate the expected AoII achieved by the threshold policy defined by Definition 5. To this end, we define \( \omega \triangleq t_{\text{max}} + \tau \). Then, we have the following proposition.

**Proposition 5 (Performance of threshold policy).** The expected AoII resulting from the adoption of threshold policy \( \tau \) is given by

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

where

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

\[ 1 - \sum_{t=1}^{t_{\text{max}}} \Upsilon(\omega + t, t) \]
\[ \Pi_t = \sum_{i=t}^{\omega-1} \Upsilon(i+t,t)\pi_i + \Upsilon(\omega+t,t)\Pi, \quad 1 \leq t \leq t_{\text{max}}, \]

\[ \Pi = \sum_{i=\tau}^{\omega-1} \left( \sum_{k=\tau}^{i} P_{i,t_{\text{max}}+k}(1) \right) \pi_i \bigg/ \left( 1 - \sum_{i=1}^{t_{\text{max}}} P_{\omega,\omega+i}(1) \right), \]

\[ \Delta'_t = \sum_{i=1}^{t_{\text{max}}} p_i \left( \frac{t - t(1-p)}{p} \right) + p_{t^+} \left( \frac{t - t(1-p)t_{\text{max}}}{p} \right), \quad 1 \leq t \leq t_{\text{max}}. \]

\( \pi_{\Delta}'s \) for \( 1 \leq \Delta \leq \omega - 1 \) are the solutions to the following system of linear equations.

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

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

For each \( 2 \leq \Delta \leq t_{\text{max}} - 1, \)

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

For each \( t_{\text{max}} \leq \Delta \leq \omega - 1, \)

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

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

**Proof.** We adopt a similar methodology as presented in the proof of Proposition 4. The complete proof can be found in Appendix 1. \( \square \)

For later analysis, we consider a special case of \( \tau = 1. \) More precisely, we have the following corollary.
Corollary 1 (Proposition 5 when τ = 1). The expected AoII resulting from the adoption of threshold policy with τ = 1 is given by

\[ \bar{\Delta}_1 = \sum_{i=1}^{\omega-1} C(i, 1) \pi_i + \Sigma, \]

where

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

\[ \pi_{\Delta} = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1) \pi_i + P_{\Delta,\Delta}(1) \left( \frac{1 - \pi_0}{ET} - \sum_{i=1}^{\Delta-1} \pi_i \right), \quad 2 \leq \Delta \leq t_{max} - 1, \]

\[ \pi_{t_{max}} = \sum_{i=1}^{t_{max}-1} P_{i,t_{max}}(1) \pi_i, \]

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

\[ 1 - \sum_{t=1}^{t_{max}} \Upsilon(\omega + t, t) \]

\[ \Pi_t = \sum_{i=\omega-t}^{\omega-1} \Upsilon(i + t, t) \pi_i + \Upsilon(\omega + t, t) \Pi, \quad 1 \leq t \leq t_{max}, \]

\[ \Pi = \sum_{t=1}^{t_{max}} \left( \sum_{k=1}^{t} P_{t,t_{max}+k}(1) \right) \pi_t \]

\[ 1 - \sum_{t=1}^{t_{max}} P_{t_{max}+1,t_{max}+1+t}(1) \]

\[ \Delta'_t = \sum_{i=1}^{t_{max}} p_i \left( \frac{t - t(1 - p)_i}{p} \right) + p_{t+} \left( \frac{t - t(1 - p)^{t_{max}}}{p} \right), \quad 1 \leq t \leq t_{max}. \]

Proof. The results can be obtained by setting τ = 1 in Proposition 5. The complete proof can be found in Appendix J.

Remark 4. We notice that \( \bar{\Delta}_0 \) and \( \bar{\Delta}_1 \) are given as closed-form expressions, so they can be easily computed.

V. OPTIMAL POLICY

In this section, we find the optimal policy for the problem reported in (5). To begin with, we introduce the policy improvement theorem.
A. Policy Improvement Theorem

We start with introducing the policy iteration algorithm, which is an iterative algorithm that iterates between two main steps.

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

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

where $P^{\phi}_{s,s'}$ is the state transition probabilities from $s$ to $s'$ resulting from the adoption of policy $\phi$. Note that (8) forms a underdetermined system. Hence, we can select any state $s$ as a reference state and set the corresponding value function as 0. In this way, we can obtain a unique solution.

2) The second step is policy improvement step. In this step, we obtain a new policy $\phi'$ by applying $V^\phi(\cdot)$ and $\theta^\phi$ obtained in the first step to the Bellman equation. More precisely, the action suggested by $\phi'$ at state $s$ is determined in the following way.

$$\phi'(s) = \arg \max_{a \in A} \left\{ C(s) + \sum_{s' \in S} P[s' \mid s, a] V^\phi(s') \right\}.$$

The pseudocode for policy iteration algorithm can be found in Appendix M. With policy iteration algorithm in mind, we can proceed with presenting the policy improvement theorem.

**Theorem 2** (Policy improvement theorem). *Suppose that we have obtained the value function resulting from the operation of a policy $A$ and that the policy improvement step has produced a policy $B$.*

- **If** $B$ **is different from** $A$, **then** $\theta^A \geq \theta^B$.
- **When policy improvement step converges** (i.e., $B$ **is the same as** $A$), **the converged policy is optimal.**

**Proof.** The proof follows the steps presented in [19]. The complete proof can be found in Appendix K. 

[Refer to the original paper for the complete proof.]
B. The Optimality Proof

Leveraging the policy improvement theorem detailed in Theorem 2, we are able to find the optimal policy for the problem reported in (3). To this end, we first introduce a 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

\[
\sigma = \frac{\sum_{i=1}^{t_{max}} p_i \left( \frac{1 - (1 - p)^{t_i}}{p} \right) + p_{t+} \left( \frac{1 - (1 - p)t_{max}}{p} \right)}{1 - \left( \sum_{t=1}^{t_{max}} p_t (1 - p)^{t-1} + p_{t+} (1 - p)t_{max} \right)}.
\]  \hspace{1cm} (9)

**Remark 5.** The closed-form expressions of \( \bar{\Delta}_0 \) and \( \bar{\Delta}_1 \) are given in Proposition 4 and Corollary 7, respectively. Hence, the inequality in Condition 1 is easy to verify.

**Theorem 3** (Optimal policy). Under Condition 1, the optimal policy for \( \mathcal{M} \) is threshold policy with \( \tau = 1 \).

**Proof.** 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 threshold policy with \( \tau = 1 \).

2) **Policy Improvement:** We apply the value functions obtained in the previous step to Bellman equation and verify that the resulting policy remains the same.

The complete proof can be found in Appendix L.

\[\square\]

VI. Numerical Results

In this section, we present the numerical results that verify Condition 1 and highlight the performance of the optimal policy. We vary the system parameters and plot the corresponding numerical results.
A. Verification of Condition 1

We start with the verification of Condition 1. We notice that verifying Condition 1 is equivalent to verifying the following two inequalities.

\[
\begin{align*}
\bar{\Delta}_1 - \frac{1+(1-p)\sigma}{2} &\leq 0, \\
\bar{\Delta}_1 - \bar{\Delta}_0 &\leq 0.
\end{align*}
\]

With this in mind, we vary the system parameters and verify the two inequalities accordingly. More precisely, in Fig. 4, we laid out the results for different values of \( t_{\text{max}} \). For each specific \( t_{\text{max}} \), we consider the case where \( T \) is geometrically distributed with success probability \( p_s = 0.6 \). Then, we vary the value of \( p \) and plot the results. The figure shows that both inequalities are verified for \( p \in [0.025, 0.5] \) for all \( t_{\text{max}} \) considered. Furthermore, in Fig. 5 and for each \( t_{\text{max}} \), we set \( p = 0.35 \) and assume that \( T \) is geometrically distributed. We vary the value of success probability \( p_s \) and plot the corresponding results. We can conclude from the plot that both inequalities are verified for \( p_s \in [0, 1] \) for all \( t_{\text{max}} \) considered.

Remark 6. When \( T \) follows other probability distribution, Condition 1 can be verified in a very similar way.
(a) Verification for $\bar{\Delta}_1 - \frac{1+(1-p)\sigma}{2} \leq 0$.

(b) Verification for $\bar{\Delta}_1 - \bar{\Delta}_0 \leq 0$.

Fig. 5: Condition 1 verification. Different lines represent different values of $t_{\text{max}}$. In this case, we fix $p = 0.35$ and consider the case where $T$ is geometrically distributed with parameter $p_s$.

B. Performance of the Optimal Policy

In the following, we provide the numerical results to highlight the performance of the optimal policy under various system parameters.

Remark 7. As Condition 1 is verified for the cases considered in Section VI-A, we can conclude that, for these cases, the optimal policy is threshold policy with $\tau = 1$. Hence, we consider the same system settings as considered in Section VI-A.

Now, we show the performance of the threshold policy with $\tau = 1$. Note that the expected AoIIs can be calculated using Corollary 1. In Fig. 6a, we fix the success probability in geometric distribution $p_s = 0.6$ and vary the value of $p$. As we can see, the expected AoII increases as $p$ increases. The reason behind this is simple. When the transmission time distribution is fixed and $p$ increases, the transmitted updates will likely be incorrect by the time they arrive. Mathematically, we can show that $C(\Delta, 1)$ is increasing in $p$. Hence, the expected AoII will increase. Likewise, the expected AoII increases as $t_{\text{max}}$ increases. To explain this, we notice that when $p$ is fixed and $t_{\text{max}}$ increases, the transmitted updates will likely be incorrect by the time they arrive. Again, we can verify that $C(\Delta, 1)$ is increasing in $t_{\text{max}}$. Hence, the expected AoII will increase. We also notice that, under geometric distribution, there is a high probability that the transmission will be completed in the first few time slots. Hence, $t_{\text{max}}$ will have mild impact on the expected AoII,
(a) When $p_s = 0.6$ is fixed. (b) When $p = 0.35$ is fixed.

Fig. 6: The performance of the optimal policy when the transmission time $T$ is geometrically distributed. Different lines represent different values of $t_{max}$.

especially when $t_{max}$ is large. In Fig. 6(b), we fix $p = 0.6$ and vary the value of $p_s$. The figure shows that the expected AoII is decreasing in $p_s$. This is because the expected transmission time decreases as $p_s$ increases. In this case, when $p$ is fixed, the transmitted update will likely be correct when they arrive. We also notice that the expected AoII increases as $t_{max}$ increases, and the impact of $t_{max}$ is weak. The reason is similar to the one stated for Fig. 6(a). Finally, the performance gap between different $t_{max}$ is getting smaller when $p_s$ enlarges. This is because the increment in $p_s$ increases the probability of completing the transmission earlier. Therefore, when $p_s$ is large, $t_{max}$ has almost negligible impact on the performance.

VII. Conclusion

In this paper, we considered the problem of minimizing the Age of Incorrect Information in a slotted-time system with a transmitter-receiver pair. In the system, a transmitter observes a dynamic source and sends updates to a remote receiver through a communication channel with random delay. We preset a maximum transmission time to prevent an update from occupying the channel for too long. When the transmission time of an update reaches the maximum transmission time, the update is immediately discarded, and the communication channel is cleared for subsequent updates. The goal is to find when the transmitter should initiate transmissions to minimize the AoII of the system. By leveraging the notation of the Markov Decision Process,
we characterized the problem and proved that the optimal policy exists. Then, by exploring the unique structure of the Markov Decision Process, we theoretically calculated the expected AoIIIs achieved by several canonical transmission policies, which are critical in proving the optimal policy. Then, we proved that, under Condition 1, the optimal policy is the one that initiates transmissions whenever the channel is idle and AoII is not zero. Finally, we take the case where the transmission time is geometrically distributed as an example. For this example, we numerically verified Condition 1 and laid out the optimal policy’s performance to highlight the characteristics of the considered problem.

REFERENCES

[1] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, T. Soleymani et al., “Semantic communications in networked systems,” arXiv preprint arXiv:2103.05391, 2021.

[2] S. Kaul, R. Yates, and M. Gruteser, “Real-time status: How often should one update?” in 2012 Proceedings IEEE INFOCOM. IEEE, 2012, pp. 2731–2735.

[3] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, “Age of information: An introduction and survey,” IEEE Journal on Selected Areas in Communications, vol. 39, no. 5, pp. 1183–1210, 2021.

[4] Y. Sun, I. Kadota, R. Talak, and E. Modiano, “Age of information: A new metric for information freshness,” Synthesis Lectures on Communication Networks, vol. 12, no. 2, pp. 1–224, 2019.

[5] A. Kosta, N. Pappas, V. Angelakis et al., “Age of information: A new concept, metric, and tool,” Foundations and Trends® in Networking, vol. 12, no. 3, pp. 162–259, 2017.

[6] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, “Update or wait: How to keep your data fresh,” IEEE Transactions on Information Theory, vol. 63, no. 11, pp. 7492–7508, 2017.

[7] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: A new performance metric for status updates,” IEEE/ACM Transactions on Networking, vol. 28, no. 5, pp. 2215–2228, 2020.

[8] A. Maatouk, M. Assaad, and A. Ephremides, “The age of incorrect information: An enabler of semantics-empowered communication,” arXiv preprint arXiv:2012.13214, 2020.

[9] Y. Chen and A. Ephremides, “Minimizing age of incorrect information for unreliable channel with power constraint,” in 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021, pp. 1–6.

[10] ——, “Scheduling to minimize age of incorrect information with imperfect channel state information,” Entropy, vol. 23, no. 12, p. 1572, 2021.

[11] S. Kriouile and M. Assaad, “Minimizing the age of incorrect information for real-time tracking of markov remote sources,” in 2021 IEEE International Symposium on Information Theory (ISIT). IEEE, 2021, pp. 2978–2983.

[12] C. Kam, S. Kompella, and A. 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). IEEE, 2020, pp. 1–6.

[13] Y. Chen and A. Ephremides, “Age of incorrect information under delay,” arXiv preprint arXiv:2203.02979, 2022.

[14] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, “Remote estimation of the wiener process over a channel with random delay,” in 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017, pp. 321–325.
According to Propositions 1 and 3 of [18], it is sufficient to show that $V_{\gamma}(s)$ is finite for all $\gamma$ and $s$. To this end, we consider the policy $\phi$ being the one that never initiate any transmissions. According to (5), we have
\[
V_{\phi,\gamma}(s) = \mathbb{E}_{\phi} \left[ \sum_{t=0}^{\infty} \gamma^t C(s_t) \mid s \right] \leq \sum_{t=0}^{\infty} \gamma^t (\Delta + t) = \frac{\Delta}{1 - \gamma} + \frac{\gamma}{1 - \gamma^2} < \infty.
\]
Then, by definition, we have $V_{\gamma}(s) \leq V_{\gamma,\phi}(s) < \infty$ for all $\gamma$ and $s$. Then, we can conclude that the value iteration reported in (6) will converge to the value function.

Leveraging Lemma 1, Lemma 2 can be proved by 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 want to examine whether the monotonicity still holds at iteration $\nu + 1$. We notice that the state transitions corresponding to state $s = (\Delta, t, i)$ where $\Delta > 0$ share the same structure. Hence, combining with the assumption made for iteration $\nu$, we can easily show that the lemma also holds at iteration $\nu + 1$ of value iteration. The details are omitted here for the sake of space. Then, by mathematical induction, we can conclude that Lemma 2 is true.
APPENDIX C

PROOF OF THEOREM 1

We show that $\mathcal{M}$ verifies the two conditions given in [18]. Then, the existence of the optimal policy is guaranteed. In the following, we will verify the two conditions one by one.

- There exists a non-negative $N$ such that $-N \leq h_\gamma(s)$ for all $s$ and $\gamma$: Leveraging Lemma 2, we can easily conclude that $h_\gamma(s)$ is also increasing in $\Delta$ when $\Delta > 0$. To proceed, 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. We know from Proposition 4 of [18] that $c_{s,0}(\phi)$ is finite. In the following, we consider the policy $\phi$ being the one under which the transmitter initiate transmission whenever the channel is idle. As we will show in Section IV, policy $\phi$ induces an irreducible ergodic Markov chain and the expected cost is finite. Hence, $h_\gamma(s) \leq c_{s,0}(\phi)$ by Proposition 5 of [18]. Then, we notice that

$$V_\gamma(0) - V_\gamma(s) \leq c_{0,s}(\phi),$$

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 2, we can choose $N = \max_{s \in G} \{c_{0,s}(\phi)\}$, where $G = \{s = (\Delta, t, i) : \Delta \in \{0, 1\}\}$.

- $\mathcal{M}$ has a stationary policy $\phi$ inducing an irreducible, ergodic Markov chain. Moreover, the resulting expected AoII is finite: We consider the policy $\phi$ being the one under which the transmitter initiate transmission whenever the channel is idle. Clearly, it induces an irreducible, ergodic Markov chain. Moreover, the resulting expected AoII is finite, as we can conclude easily from the closed-form expression given in Section IV.

As the two conditions are verified, the existence of the optimal policy is guaranteed by the results presented in [18]. Moreover, as given in [18], the minimum expected AoII is independent of the initial state.

APPENDIX D

PROOF OF LEMMA 3

We start with the case of $\Delta = 0$. More precisely, we derive the expressions of $P^t_{0,\Delta'}(1)$ and $P^{t^+}_{0,\Delta}(1)$. To this end, we distinguish between different values of $t$. 

July 8, 2022 DRAFT
• When \( 1 \leq t < t_{\text{max}} \), the update will be delivered after \( t \) time slot. Then, we further distinguish between the following cases.
  
  - \( \Delta' = 0 \) when the receiver’s estimate is correct when the update arrives. Since the transmitting update is the same as the receiver’s estimate, \( \Delta' = 0 \) happens with probability \( p(t) \).
  
  - \( \Delta' = k \in \{1, 2, ..., t\} \) when the source remains in the same state \( t - k \) 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 \( t = t_{\text{max}} \), the update either arrives or be discarded. 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, \( P_{0,\Delta'}^t(1) = P_{0,\Delta'}^{t_{\text{max}}}(1) \), which can be obtained by setting the \( t \) in the above case to \( t_{\text{max}} \).

Combining together, for \( 1 \leq t \leq t_{\text{max}} \), we obtain the following.

\[
P_{0,\Delta'}^t(1) = \begin{cases} 
  p(t) & \Delta' = 0, \\
  p^{(t-k)}p(1-p)^{k-1} & \Delta' = k \in \{1, 2, ..., t\}, \\
  0 & \text{otherwise}.
\end{cases}
\]

\[
P_{0,\Delta'}^{t_{\text{max}}}(1) = \begin{cases} 
  p^{(t_{\text{max}})} & \Delta' = 0, \\
  p^{(t_{\text{max}}-k)}p(1-p)^{k-1} & \Delta' = k \in \{1, 2, ..., t_{\text{max}}\}, \\
  0 & \text{otherwise}.
\end{cases}
\]

Then, we consider the case where \( \Delta \geq 1 \). 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 \( 1 \leq t < t_{\text{max}} \), the update will be delivered after \( t \) time slots and the the receiver’s estimate will flip. Then, we further distinguish our discussion into the following cases.
  
  - \( \Delta' = 0 \) when the receiver’s estimate is correct when the update is received. We recall that the receiver uses what it received as the estimate. Hence, \( \Delta' = 0 \) happens with probability \( p^{(t)} \).
  
  - \( \Delta' = 1 \) when the receiver’s estimate is correct at \( (t-1) \)th time slot after the transmission and becomes incorrect when the update arrives. Hence, \( \Delta' = 1 \) happens with probability \( (1 - p^{(t-1)})(1 - p) \).
– \( \Delta' = k \in \{2, 3, \ldots, t - 1\} \) when the source is in a state different from the state when the transmission started at \((t - k)\)th time slot after the transmission started. 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, \( \Delta' = k \in \{2, 3, \ldots, t - 1\} \) 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 \( t = t_{\text{max}} \) and the transmitted update is delivered, the receiver’s estimate flips. Hence, \( P_{\Delta, \Delta'}^{t_{\text{max}}}(1) \) can be obtained by setting the \( t \) in the above case to \( t_{\text{max}} \).

- When \( t = t_{\text{max}} \) and the transmitted update is discarded, the receiver’s estimate remains the same. With this in mind, we further divide our discussion into the following cases.

– \( \Delta' = 0 \) when the source is in a state different from the state when the transmission started. Hence, \( \Delta' = 0 \) happens with probability \( 1 - p^{(t_{\text{max}})} \).

– \( \Delta' = k \in \{1, 2, \ldots, t_{\text{max}} - 1\} \) when the source is in a state different from the state when the transmission started at \((t_{\text{max}} - k)\)th time slot after the transmission started. Then, the source changes state and remains in the same state for the remainder of the transmission. Hence, \( \Delta' = k \in \{2, 3, \ldots, t_{\text{max}} - 1\} \) happens with probability \((1 - p^{(t_{\text{max}} - k)})p(1 - p)^{k-1}\).

– \( \Delta' = \Delta + t_{\text{max}} \) when the source remains in the same state throughout the transmission. Combining with the source dynamic, we can conclude that \( \Delta' = \Delta + t_{\text{max}} \) happens with probability \((1 - p)^{t_{\text{max}}} \).

Combining together, for \( \Delta \geq 1 \) and \( 1 \leq t \leq t_{\text{max}} \), we obtain

\[
P_{\Delta, \Delta'}^{t}(1) = \begin{cases} 
p^{(t)} & \Delta' = 0, \\
(1 - p^{(t-1)})(1 - p) & \Delta' = 1, \\
(1 - p^{(t-k)})p^2(1 - p)^{k-2} & \Delta' = k \in \{2, 3, \ldots, t - 1\}, \\
p(1 - p)^{t-1} & \Delta' = \Delta + t, \\
0 & otherwise. 
\end{cases}
\]
\[ P_{\Delta,\Delta'}^t(1) = \begin{cases} 
1 - p(t_{max}) & \Delta' = 0, \\
(1 - p(t_{max} - k))p(1 - p)^{k-1} & \Delta' = k \in \{1, 2, ..., t_{max} - 1\}, \\
(1 - p)^{t_{max}} & \Delta' = \Delta + t_{max}, \\
0 & \text{otherwise.}
\end{cases} \]

The proof of the property is obvious because we can easily verify it by checking the expressions. Therefore, details are omitted.

APPENDIX E

PROOF OF PROPOSITION 1

The equivalent expression can be obtain easily by examining (7). Hence, the details are omitted. In the following, we focus on showing the presented properties. To this end, we show the properties one by one.

1) When \(0 \leq \Delta' \leq t - 1\) and \(\Delta \geq \max\{1, \Delta'\}\), we recall from Lemma 3 that \(P_{\Delta,\Delta'}^t(1)\) and \(P_{\Delta,\Delta'}^+(1)\) are independent of \(\Delta\) for any feasible \(t\). Moreover, \(P_{\Delta,\Delta'}^t(1) = P_{\Delta,\Delta'}^+(1) = 0\) when \(t \leq \Delta' \leq t_{max} - 1\) for any feasible \(t\). Combining with (7), we can conclude that property 1 is true.

2) Since \(t' = \Delta' - \Delta\), \(P_{\Delta,\Delta'}^t = P_{\Delta,\Delta'+t'}^t(1)\). Then, according to Lemma 3, \(P_{\Delta,\Delta'+t'}^t(1)\) is independent of \(\Delta \geq 1\) and depends only on \(t'\). Moreover, we notice that, when \(\Delta' \geq t_{max}\) and \(\Delta' \neq \Delta + t_{max}\), \(P_{\Delta,\Delta'}^t(1) = 0\) for \(\Delta \geq 1\). Also, according to Lemma 3, \(P_{\Delta,\Delta'}^t(1)\) is independent of \(\Delta \geq 1\) when \(\Delta' = \Delta + t_{max}\). Combining together, we can conclude that \(P_{\Delta,\Delta'}^t(1)\) depends only on \(t'\) when \(\Delta \geq 1\). Consequently, property 2 is true.

3) When \(\Delta' > \Delta + t_{max}\), the property holds apparently. When \(t_{max} - 1 < \Delta' < \Delta + 1\), \(t' \leq 0\), \(P_{\Delta,\Delta'}^t(1) = 0\) by definition. Hence, property 3 is true.

APPENDIX F

PROOF OF PROPOSITION 2

We first notice that, when \(a = 1\), \(C(\Delta, 1)\) can be calculated by

\[ C(\Delta, 1) = \sum_{t=1}^{t_{max}} p_tC'(\Delta, 1) + p_{t+}C_{t_{max}}(\Delta, 1), \]
where $C^t(\Delta, 1)$ is the expected cost when the transmission finishes after $t$ time slots. Hence, $C^t(\Delta, 1)$ can be calculated by

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

where $C^k(\Delta)$ is the instant cost $k$ time slots after the transmission has started and the transmission has not yet completed. We notice that $C^k(\Delta)$ is a random variable. Hence, we investigate its probability distribution. To this end, we distinguish between the following cases.

- We start with considering the expression of $C^k(0)$. In this case, the transmission started when $\Delta = 0$.
  - When $k = 0$, $C^0(0) \triangleq 0$ by definition.
  - For $1 \leq k \leq t_{\text{max}} - 1$, $C^k(0) = h$ where $h \in \{0, 1, 2, \ldots, k\}$. Since $\Delta = 0$, we can easily conclude that $C^k(0) = 0$ with probability $p^{(k)}$. For $h > 0$, $C^k(0) = h$ when the source is in the same state as the state when the transmission started at the $(k-h)$th time slot after the transmission. Then, the source flips the state and remains in the same state for the remaining time slots. Hence, we can conclude that $C^k(0) = h$ with probability $p^{(k-h)}(1-p)^{h-1}$.

Combining together, we obtain the following.

$$C^k(0) = \sum_{h=1}^{k} hp^{(k-h)}(1-p)^{h-1}.$$

- Then, we consider the case of $C^k(\Delta)$, where $\Delta > 0$. In this case, the transmission started when the receiver’s estimate is incorrect.
  - When $k = 0$, $C^0(\Delta) \triangleq \Delta$ by definition.
  - For $1 \leq k \leq t_{\text{max}} - 1$, $C^k(\Delta) = h$, where $h \in \{1, 2, \ldots, k-1\}$, when the source is in the state different from the state when the transmission started at the $(k-h)$th slot after the transmission. Then, the source flips the state and remains in the same state for the remaining time slots. Hence, $C^k(\Delta) = h$ 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 wrong for $k$ time slots after the transmission started. Since $\Delta > 0$ and the receiver’s estimate did not change, $C^k(\Delta) = \Delta + k$ happens with probability $(1-p)^k$. Lastly, $C^k(\Delta) = 0$ when the state of the source at $k$ time slots after the start of the transmission is different from that when the transmission started, which happens with probability $(1 - p^{(k)})$. 


Hence, we obtain the following.

\[ C^k(\Delta) = \sum_{h=1}^{k-1} h(1 - p^{(k-h)})p(1 - p)^{h-1} + (\Delta + k)(1 - p)^k. \]

Combining together, we have

\[ C(\Delta, 1) = \sum_{t=1}^{t_{max}} ptC^t(\Delta, 1) + ptC_{max}(\Delta, 1), \]

where

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

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

**APPENDIX G**

**PROOF OF PROPOSITION 3**

We recall that, under the lazy policy, the transmitter will never transmit any updates. Hence, the receiver’s estimate will never change. Without loss of generality, we assume the receiver’s estimate \( \hat{X}_k = 0 \) for all \( k \). Then, combining with the state transition probabilities detailed in Section III-A, \( \pi_{\Delta} \)'s satisfy the following equations.

\[ \pi_0 = (1 - p)\pi_0 + p \sum_{i=1}^{\infty} \pi_i, \]  
\[ \pi_1 = p\pi_0. \]

\[ \pi_{\Delta} = (1 - p)\pi_{\Delta-1}, \quad \Delta \geq 2. \]

\[ \sum_{i=0}^{\infty} \pi_i = 1. \]  

(11)

In the following, we will first solve the above system of linear equations. To this end, combining (10) and (11) yields

\[ \pi_0 = (1 - p)\pi_0 + p \sum_{i=1}^{\infty} \pi_i = (1 - p)\pi_0 + p(1 - \pi_0). \]

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, 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 transmissions, the cost for being at state \((\Delta, 0, -1)\) is nothing but \(\Delta\) itself. Hence, the expected AoII resulting from the adoption of the lazy policy can be calculated by
\[ \bar{\Delta}_\infty = \sum_{\Delta=1}^{\infty} \Delta \frac{p(1 - p)^{\Delta-1}}{2} = \frac{p}{2} \sum_{\Delta=1}^{\infty} \Delta (1 - p)^{\Delta-1} = \frac{1}{2p}. \]

**APPENDIX H**

**PROOF OF PROPOSITION 4**

We recall that, under the zero wait policy, the transmitter will transmit the new update whenever the channel is idle. Hence, combining with the probabilities detailed in Section IV-A, \(\pi_\Delta\)'s satisfy the following linear equations.

\[ \pi_\Delta = \sum_{i=0}^{\infty} P_{i,\Delta}(1) \pi_i, \quad 0 \leq \Delta \leq t_{max} - 1. \]  \hspace{1cm} (12)

\[ \pi_\Delta = \sum_{i=\Delta-t_{max}}^{\Delta-1} P_{i,\Delta}(1) \pi_i, \quad \Delta \geq t_{max}. \]  \hspace{1cm} (13)

\[ \sum_{i=0}^{\infty} \pi_i = \frac{1}{ET}. \]

The last equation comes from the fact that every transmission starts at state \((\Delta, 0, -1)\) follows a time period of expected length \(ET\) time slots, during which the channel is busy.

To solve the above system of linear equations, we first sum (13) from \(t_{max}\) to infinity, which yields
\[ \sum_{i=t_{max}}^{\infty} \pi_i = \sum_{i=t_{max}}^{\infty} \sum_{k=i-t_{max}}^{i-1} P_{k,i}(1) \pi_k. \]  \hspace{1cm} (14)

Then, we investigate the RHS of (14) by breaking down the summation.
\[ RHS = \sum_{k=0}^{t_{max}-1} P_{k,t_{max}}(1) \pi_k + \sum_{k=t_{max}}^{t_{max}} P_{k,t_{max}+1}(1) \pi_k + \cdots + \sum_{k=t_{max}-1}^{2t_{max}-2} P_{k,2t_{max}-1}(1) \pi_k + \cdots \]
Rearranging the summations yields
\[
RHS = P_{0,t_{\max}}(1)\pi_0 + 2 \sum_{k=1}^{t_{\max}} P_{1,t_{\max}+k-1}(1)\pi_1 + \cdots + \sum_{k=1}^{t_{\max}} P_{t_{\max}-1,t_{\max}+k-1}(1)\pi_{t_{\max}-1} + \sum_{k=1}^{t_{\max}} P_{t_{\max},t_{\max}+k}(1)\pi_{t_{\max}} + \sum_{k=1}^{t_{\max}} P_{t_{\max}+1,t_{\max}+k}(1)\pi_{t_{\max}+1} + \cdots
\]

According to property 2 Proposition 1, we have
\[
RHS = \sum_{i=0}^{t_{\max}-1} \left( \sum_{k=0}^{i} P_{i,t_{\max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{\max}} \left( P_{t_{\max},t_{\max}+i}(1) \right) \left( \sum_{i=t_{\max}}^{\infty} \pi_i \right).
\]

Then, we define \( \Pi \triangleq \sum_{i=\max}^{\infty} \pi_i \). Combining with the definition, equation (14) can be written as
\[
\Pi = \sum_{i=0}^{t_{\max}-1} \left( \sum_{k=0}^{i} P_{i,t_{\max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{\max}} \left( P_{t_{\max},t_{\max}+i}(1) \right) \Pi.
\]

Hence, we obtain
\[
\Pi = \frac{t_{\max}-1}{1 - \sum_{i=1}^{t_{\max}} \left( P_{t_{\max},t_{\max}+i}(1) \right)}.
\]

In the following, we calculate \( \pi_{\Delta} \) for \( 0 \leq \Delta \leq t_{\max} \). To start with, we consider the case of \( \Delta = 0 \). In this case, equation (12) reduces to
\[
\pi_0 = P_{0,0}(1)\pi_0 + \sum_{i=1}^{\infty} P_{i,0}(1)\pi_i = P_{0,0}(1)\pi_0 + P_{1,0}(1) \left( \frac{1}{ET} - \pi_0 \right).
\]

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

Moreover, by property 1 of Proposition 1, \( P_{i,\Delta}(1) = P_{\Delta,\Delta}(1) \) for any \( i \geq \Delta \) when \( 1 \leq \Delta \leq t_{\max} - 1 \). Hence, equation (12) can be written as
\[
\pi_{\Delta} = \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \sum_{i=\Delta}^{\infty} \pi_i
\]
\[
= \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \left( \frac{1}{ET} - \sum_{i=0}^{\Delta-1} \pi_i \right), \quad 1 \leq \Delta \leq t_{\max} - 1.
\]

Consequently, \( \pi_{\Delta} \) for \( 1 \leq \Delta \leq t_{\max} - 1 \) can be calculated iteratively. When \( \Delta = t_{\max} \), we have
\[
\pi_{t_{\max}} = \sum_{i=0}^{t_{\max}-1} P_{i,t_{\max}}(1)\pi_i.
\]
As we will see later, (15), (16), (17), and (18) are sufficient to obtain the resulting expected AoII \( \Delta_0 \). We notice that, under the zero wait policy, the expected cost for being at state \((\Delta, 0, -1)\) is always \( C(\Delta, 1) \). We define \( \Sigma = \sum_{i=t_{\text{max}}+1}^{\infty} C(i, 1) \pi_i \). Then, we have
\[
\Delta_0 = \sum_{i=0}^{t_{\text{max}}} C(i, 1) \pi_i + \Sigma.
\]

Then, it is sufficient to calculate the quantity \( \Sigma \). Hence, in the following, we focus on obtaining the closed form expression of \( \Sigma \). To this end, we rewrite (13) as
\[
\pi_{\Delta} = \sum_{i=\Delta-t_{\text{max}}}^{\Delta-1} P_{i, \Delta}(1) \pi_i = \sum_{i=1}^{t_{\text{max}}} P_{\Delta-t_{\text{max}}+i-1, \Delta}(1) \pi_{\Delta-t_{\text{max}}+i-1}, \quad \Delta \geq t_{\text{max}} + 1.
\]

We recall from Proposition 1, \( P_{\Delta, \Delta'}(1) = p_{\Delta} P_{\Delta, \Delta'}(1) + p_{\Delta'} P_{\Delta', \Delta}(1) \) where \( \Delta' = \Delta - \Delta \) when \( \Delta' \geq t_{\text{max}} + 1 \). Then,
\[
\pi_{\Delta} = \sum_{i=1}^{t_{\text{max}}} \left( p_{\Delta_{\text{max}}+1-i} P_{\Delta-t_{\text{max}}+i-1, \Delta}(1) + p_{\Delta} P_{\Delta_{\text{max}}+1-i, \Delta}(1) \right) \pi_{\Delta-t_{\text{max}}+i-1}, \quad \Delta \geq t_{\text{max}} + 1.
\]

Renaming the variables yields
\[
\pi_{\Delta} = \sum_{t=1}^{t_{\text{max}}} \left( p_{t} P_{\Delta-t, \Delta}(1) + p_{t+1} P_{\Delta-t, \Delta}(1) \right) \pi_{\Delta-t}
\]
\[
= \sum_{t=1}^{t_{\text{max}}} \Upsilon(\Delta, t) \pi_{\Delta-t}, \quad \Delta \geq t_{\text{max}} + 1,
\]
where \( \Upsilon(\Delta, t) \triangleq p_{t} P_{\Delta-t, \Delta}(1) + p_{t+1} P_{\Delta-t, \Delta}(1) \). To proceed, we define, for each \( 1 \leq t \leq t_{\text{max}} \),
\[
\pi_{\Delta, t} \triangleq \Upsilon(\Delta, t) \pi_{\Delta-t}, \quad \Delta \geq t_{\text{max}} + 1.
\]

Note that \( \sum_{t=1}^{t_{\text{max}}} \pi_{\Delta, t} = \pi_{\Delta} \). Then, for given \( t \), we have
\[
\sum_{i=t_{\text{max}}+1}^{\infty} C(i-t, 1) \pi_{i, t} = \sum_{i=t_{\text{max}}+1}^{\infty} C(i-t, 1) \Upsilon(i, t) \pi_{i-t}.
\] (19)

We define \( \Delta'_{\text{i}} \triangleq C(\Delta, 1) - C(\Delta - t, 1) \). Then, according to Proposition 2, we have
\[
\Delta'_{\text{i}} = \sum_{i=1}^{t_{\text{max}}} \left( C_{\text{i}}(\Delta, 1) - C_{\text{i}}(\Delta - t, 1) \right) + p_{t+1} \left( C_{\text{i}_{\text{max}}}(\Delta, 1) - C_{\text{i}_{\text{max}}}(\Delta - t, 1) \right).
\]

Moreover, we have
\[
C_{\text{i}}(\Delta - t, 1) = \Delta - t + \sum_{h=1}^{i-1} \sum_{k=1}^{h-1} k(1-p)^{h-k} p (1-p)^{k-1} + (\Delta - t + h)(1-p)^h.
\]
\[
C_{\text{i}}(\Delta, 1) = \Delta + \sum_{h=1}^{i-1} \sum_{k=1}^{h-1} k(1-p)^{h-k} p (1-p)^{k-1} + (\Delta + h)(1-p)^h.
\]
Subtracting the two equations yields
\[ C_i^\prime(\Delta, 1) - C_i^\prime(\Delta - t, 1) = t + \sum_{h=1}^{i-1} t(1 - p)^h = \frac{t - t(1 - p)^i}{p}. \]

Then, we have
\[ \Delta_t' = \sum_{i=1}^{t_{max}} p_i \left( \frac{t - t(1 - p)^i}{p} \right) + p_{t_+} \left( \frac{t - t(1 - p)^{t_{max}}}{p} \right), \quad 1 \leq t \leq t_{max}. \]

We notice that \( \Delta_t' = C(\Delta, 1) - C(\Delta - t, 1) \) is independent of \( \Delta \). Hence, equation (19) can be written as
\[ \sum_{i=t_{max}+1}^{\infty} \left( C(i, 1) - \Delta_t' \right) \pi_{i,t} = \sum_{i=t_{max}+1-t}^{\infty} C(i, 1) \Upsilon(i + t, t) \pi_i. \]

Then, we define \( \Pi_t \triangleq \sum_{i=t_{max}+1}^{\infty} \pi_{i,t} \) and \( \Sigma_t \triangleq \sum_{i=t_{max}+1}^{\infty} C(i, 1) \pi_{i,t} \). We recall that \( \Upsilon(\Delta, t) \) is independent of \( \Delta \) if \( \Delta \geq t_{max} + 1 \). Hence, plugging in the definitions yields
\[ \Sigma_t - \Delta_t' \Pi_t = \sum_{i=t_{max}+1-t}^{t_{max}} \Upsilon(i + t, t) C(i, 1) \pi_i + \Upsilon(t_{max} + 1 + t, t) \Sigma. \]

Summing the above equation over \( t \) from 1 to \( t_{max} \) yields
\[ \sum_{t=1}^{t_{max}} \left( \Sigma_t - \Delta_t' \Pi_t \right) = \sum_{t=1}^{t_{max}} \left( \sum_{i=t_{max}+1-t}^{t_{max}} \Upsilon(i + t, t) C(i, 1) \pi_i + \Upsilon(t_{max} + 1 + t, t) \Sigma \right). \]

Rearranging the above equation yields
\[ \Sigma - \sum_{i=1}^{t_{max}} \Delta_t' \Pi_t = \sum_{t=1}^{t_{max}} \left( \sum_{i=t_{max}+1-t}^{t_{max}} \Upsilon(i + t, t) C(i, 1) \pi_i \right) + \sum_{t=1}^{t_{max}} \Upsilon(t_{max} + 1 + t, t) \Sigma. \]

Then, the closed-form expression for \( \Sigma \) is given by
\[ \Sigma = \sum_{t=1}^{t_{max}} \left[ \left( \sum_{i=t_{max}+1-t}^{t_{max}} \Upsilon(i + t, t) C(i, 1) \pi_i \right) + \Pi_t \Delta_t' \right] \frac{1 - \sum_{t=1}^{t_{max}} \Upsilon(t_{max} + 1 + t, t)}{1 - \sum_{t=1}^{t_{max}} \Upsilon(t_{max} + 1 + t, t)}. \]

In the following, we calculate \( \Pi_t \). To this end, we have
\[ \Pi_t \triangleq \sum_{i=t_{max}+1}^{\infty} \pi_{i,t} = \sum_{i=t_{max}+1}^{\infty} \Upsilon(i, t) \pi_{i-t} = \sum_{i=t_{max}+1-t}^{\infty} \Upsilon(i + t, t) \pi_i. \]

Since \( \Upsilon(\Delta, t) \) is independent of \( \Delta \) if \( \Delta \geq t_{max} + 1 \), we have
\[ \Pi_t = \sum_{i=t_{max}+1-t}^{t_{max}-1} \Upsilon(i + t, t) \pi_i + \Upsilon(t_{max} + t, t) \Pi, \quad 1 \leq t \leq t_{max}. \]

Combining together, we recover the results presented in the proposition.
APPENDIX I

PROOF OF PROPOSITION 5

We define \( \omega \triangleq \tau + t_{\text{max}} \) and \( \Pi \triangleq \sum_{i=\omega}^{\infty} \pi_i \). Then, we recall that, under threshold policy \( \tau \), the transmitter will initiate a new transmission only when the current AoII is no less than threshold \( \tau \) and the channel is idle. Hence, \( \pi_{\Delta} \)'s satisfy the following linear equations.

\[
\pi_0 = (1 - p) \pi_0 + p \sum_{i=1}^{\tau-1} \pi_i + P_{\tau,0}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right). \tag{20}
\]

\[
\pi_1 = p \pi_0 + P_{\tau,1}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right). \tag{21}
\]

For each \( 2 \leq \Delta \leq t_{\text{max}} - 1 \),

\[
\pi_{\Delta} = \begin{cases} 
(1 - p) \pi_{\Delta - 1} + P_{\tau,\Delta}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right) & \text{if } \; \Delta - 1 < \tau, \\
\sum_{i=\tau}^{\Delta-1} P_{i,\Delta}(1) \pi_i + P_{\Delta,\Delta}(1) \left( \sum_{i=\Delta}^{\omega-1} \pi_i + \Pi \right) & \text{if } \; \Delta - 1 \geq \tau.
\end{cases} \tag{22}
\]

For each \( t_{\text{max}} \leq \Delta \leq \omega - 1 \),

\[
\pi_{\Delta} = \sum_{i=\Delta - t_{\text{max}}}^{\Delta-1} P_{i,\Delta}(1) \pi_i, \quad \Delta \geq \omega. \tag{23}
\]

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

Note that we can pull the probabilities in (20), (21), and (22) out of the summation since \( P_{\Delta,\Delta'}(1) \) is independent of \( \Delta \) when \( \Delta \geq \tau \) and \( \Delta' \leq t_{\text{max}} - 1 \) by property 1 of Proposition 1.

To solve the above system of linear equations, we first notice that there are infinity many equations as indicated by (23). To overcome the infinity, we sum (23) from \( \omega \) to infinity.

\[
\sum_{i=\omega}^{\infty} \pi_i = \sum_{i=\omega}^{\infty} \sum_{k=i-t_{\text{max}}}^{i-1} P_{k,i}(1) \pi_k. \tag{24}
\]
Then, we dive deep into the RHS of (24). To this end, we adopt a similar analysis as presented in the proof of Proposition 4. More precisely, we break down the first summation.

$$\text{RHS} = \sum_{k=\tau}^{\omega-1} P_{k,\omega}(1)\pi_k + \sum_{k=\tau}^{\omega} P_{k,\omega+1}(1)\pi_k + \cdots + \sum_{k=\omega-1}^{\omega+t_{\max}-2} P_{k,\omega+t_{\max}-1}(1)\pi_k + \sum_{k=\omega}^{\omega+t_{\max}} P_{k,\omega+t_{\max}}(1)\pi_k + \cdots$$

Then, we rearrange the summation.

$$\text{RHS} = P_{\tau,\omega}(1)\pi_{\tau} + \sum_{k=1}^{2} P_{\tau+1,\omega+k-1}(1)\pi_{\tau+1} + \cdots + \sum_{k=1}^{t_{\max}} P_{\omega-1,\omega+k-1}(1)\pi_{\omega-1} + \sum_{k=1}^{t_{\max}} P_{\omega,\omega+k}(1)\pi_{\omega} + \sum_{k=1}^{t_{\max}} P_{\omega+1,\omega+k+1}(1)\pi_{\omega+1} + \cdots$$

Then, according to property 2 Proposition 1 we have

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

Finally, equation (24) can be written as

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

(25)

Replacing (23) with (25) yields a system of linear equations with finite size. Hence, we can obtain $\pi_\Delta$ for $0 \leq \Delta \leq \omega - 1$ and $\Pi$ easily. Then, following along the same line as presented in the proof of Proposition 4, the expected AoII is given by the following equations.

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

where

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

$$\Pi_t = \sum_{i=\omega-t}^{\omega-1} \gamma(i + t, t)\pi_i + \gamma(\omega + t, t)\Pi, \quad 1 \leq t \leq t_{\max},$$

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

Remark 8. The derivation of $\Sigma$ overlaps considerably with the derivation detailed in the proof of Proposition 4. Therefore, details are omitted here.
We inherit the notations used in the proof of Proposition [5]. We notice that it is sufficient to obtain the expressions of $\pi_{\Delta}$ for $0 \leq \Delta \leq \omega - 1$. To this end, by letting $\tau = 1$, the system of linear equations that $\pi_{\Delta}$’s satisfy becomes the following.

$$\pi_0 = (1 - p)\pi_0 + P_{1,0}(1) \left( \sum_{i=1}^{\omega-1} \pi_i + \Pi \right).$$ \hspace{1cm} (26)

$$\pi_1 = p\pi_0 + P_{1,1}(1) \left( \sum_{i=1}^{\omega-1} \pi_i + \Pi \right).$$ \hspace{1cm} (27)

$$\pi_{\Delta} = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \left( \sum_{i=1}^{\omega-1} \pi_i + \Pi \right), \quad 2 \leq \Delta \leq t_{\text{max}} - 1.$$ \hspace{1cm} (28)

$$\pi_{t_{\text{max}}} = \sum_{i=1}^{t_{\text{max}}-1} P_{i,t_{\text{max}}}(1)\pi_i.$$ \hspace{1cm} (29)

To solve the above system of linear equations, we combine (26) and (29).

$$\pi_0 = (1 - p)\pi_0 + P_{1,0}(1) \left( \frac{1 - \pi_0}{ET} \right) \Rightarrow \pi_0 = \frac{P_{1,0}(1)}{pET + P_{1,0}(1)}.$$ 

Similarly, combining (27) and (29) yields

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

Then, we combine (28) and (29).

$$\pi_{\Delta} = \sum_{i=1}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \left( \frac{1 - \pi_0}{ET} - \sum_{i=1}^{\Delta-1} \pi_i \right), \quad 2 \leq \Delta \leq t_{\text{max}} - 1.$$ 

After $\pi_{\Delta}$ and $\Pi$ being obtained, the expected AoII $\bar{\Delta}_1$ can be obtained using the equations given in Proposition [5].
APPENDIX K

PROOF OF THEOREM 2

The proof is based on the results in [19]. Let $C(s, A)$ be the instant cost for being at state $s$ under policy $A$. We also define $P_{s,s'}^A$ as the probability that apply 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

$$C(s, B) + \sum_{s' \in S} P_{s,s'}^B V^A(s') \leq C(s, A) + \sum_{s' \in S} P_{s,s'}^A V^A(s'), \quad s \in S.$$ 

Then, we define

$$\gamma_s \triangleq C(s, B) + \sum_{s' \in S} P_{s,s'}^B V^A(s') - C(s, A) - \sum_{s' \in S} P_{s,s'}^A V^A(s') \leq 0.$$ 

Meanwhile, $V^A(s)$ and $V^B(s)$ satisfy their own Bellman equations. More precisely, we have

$$V^A(s) + \theta^A = C(s, A) + \sum_{s' \in S} P_{s,s'}^A V^A(s'), \quad s \in S.$$ 

$$V^B(s) + \theta^B = C(s, B) + \sum_{s' \in S} P_{s,s'}^B V^B(s'), \quad s \in S.$$ 

Then, subtracting the two expressions and bringing in the definition of $\gamma_s$ yield

$$V^B(s) - V^A(s) + \theta^B - \theta^A = \gamma_s + \sum_{s' \in S} P_{s,s'}^B (V^B(s') - V^A(s')) , \quad s \in S.$$ 

Let $V^\Delta(s) \triangleq V^B(s) - V^A(s)$ and $\theta^\Delta \triangleq \theta^B - \theta^A$. Then, we have

$$V^\Delta(s) + \theta^\Delta = \gamma_s + \sum_{s' \in S} P_{s,s'}^B V^\Delta(s'), \quad s \in S.$$ 

We know that

$$\theta^\Delta = \sum_{s \in S} \pi^B_s \gamma_s,$$

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 $\theta^\Delta \leq 0$. Equivalently, $\theta^B \leq \theta^A$.

Then, we prove that the resulting policy is optimal when the policy improvement step converges. To this end, we prove this by contradiction. Assume there exists a policy $B$ such that $\theta^B \leq \theta^A$ and the policy improvement step has converged to policy $A$. Then, we know $\gamma_s \geq 0$ for all $s \in S$. Hence, $\theta^\Delta \geq 0$. According to the definition of $\theta^\Delta$, we have $\theta^B \geq \theta^A$, which contradicts the assumption. Hence, the resulting policy is optimal when the policy improvement step converges.
APPENDIX L

PROOF OF THEOREM 3

We first recall from the definition of $\mathcal{M}$ that the feasible action when $i \neq -1$ is simply $a = 0$. Hence, for state with $i \neq -1$, the minimum operator in (4) is avoided. Consequently, the value function of state with $i \neq -1$ can be written as a linear combination of the value functions of other states. Now, let $V(\Delta)$ be short for $V(\Delta, 0, -1)$. Then, by repeatedly substituting the value functions of states with $i \neq -1$ with the corresponding expressions, we can show that $V(\Delta)$ satisfy the following equation.

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

(30)

where

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

Note that (30) provides a direct link between the value functions of states with $i = -1$. Then, the general procedures of the optimality proof can be summarized as follows.

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

2) **Policy Improvement:** we apply the value functions obtained in the previous step to Bellman equation and verify that the resulting policy remains the same.

In the following, we will elaborate the two steps.

a) **Policy Evaluation:** Here, we calculate the value function under threshold policy with $\tau = 1$. For simplicity of notation, we denote the policy as $\phi$ and the resulting value function as $V^{\phi}(\Delta)$. Then, the value functions satisfy the following system of linear equations.

$$V^{\phi}(0) = -\theta^{\phi} + pV^{\phi}(1) + (1 - p)V^{\phi}(0),$$

(31)

$$V^{\phi}(\Delta) = C(\Delta, 1) - ET\theta^{\phi} + \sum_{t=1}^{t_{\text{max}}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta, k}(1)V^{\phi}(k) + P_{\Delta, \Delta+t}(1)V^{\phi}(\Delta + t) \right) \right]$$

$$+ p_{t+} \left( \sum_{k=0}^{t_{\text{max}}-1} P_{\Delta, k}(1)V^{\phi}(k) + P_{\Delta, \Delta+t_{\text{max}}}(1)V^{\phi}(\Delta + t_{\text{max}}) \right), \quad \Delta \geq 1,$$

where $\theta^{\phi}$ is the expected AoII resulting from the adoption of $\phi$. Instead of solving the system of linear equations for the exact solution, some structural properties of the solution is sufficient for the analysis later. To this end, we provide the following lemma.
Lemma 4 (Properties of \(V^\phi(\cdot)\)). \(V^\phi(\cdot)\) satisfies the following equations.

\[
V^\phi(1) - V^\phi(0) = \frac{\theta^\phi}{p},
\]

\[
V^\phi(\Delta + 1) - V^\phi(\Delta) = \sigma^\phi, \quad \Delta \geq 1,
\]

where \(\sigma^\phi = \sigma\), whose closed-form expression is given by (9).

Proof. First of all, from (31), we have

\[
\theta^\phi = p(V^\phi(1) - V^\phi(0)) \Rightarrow V^\phi(1) - V^\phi(0) = \frac{\theta^\phi}{p}.
\]

Then, we will show that \(V^\phi(\Delta + 1) - V^\phi(\Delta)\) is constant for all \(\Delta \geq 1\). According to [20], the system of linear equations can be solved using iterative policy evaluation algorithm. Let \(V^\phi_\nu(\cdot)\) be the estimated value function at iteration \(\nu\) of the iterative policy evaluation algorithm. Then, the estimated value function is updated in the following way.

\[
V^\phi_{\nu+1}(0) = -\theta^\phi + pV^\phi_\nu(1) + (1 - p)V^\phi_\nu(0).
\]

\[
V^\phi_{\nu+1}(\Delta) = C(\Delta, 1) - ET\theta^\phi + \sum_{t=1}^{t_{\text{max}}} \left[ p_t \left( \sum_{k=0}^{t-1} P^k_{\Delta,k} V^\phi_\nu(k) + P^t_{\Delta,\Delta+t}(1) V^\phi_\nu(\Delta + t) \right) \right] + p_t+ \left( \sum_{k=0}^{t_{\text{max}}-1} P^k_{\Delta,k} V^\phi_\nu(k) + P^t_{\Delta,\Delta+t_{\text{max}}}(1) V^\phi_\nu(\Delta + t_{\text{max}}) \right), \quad \Delta \geq 1.
\]

Since the optimal policy exists according to Theorem 1, the iterative policy evaluation algorithm is guaranteed to converge to the value function \(V^\phi(\cdot)\) (i.e., \(\lim_{\nu \to \infty} V^\phi_\nu(\cdot) = V^\phi(\cdot)\)) [20]. Without loss of generality, we assume \(V^\phi_0(\cdot) = 0\). With this in mind, we can prove the desired results using mathematical induction. First of all, the base case \(\nu = 0\) is true by initialization. Then, we assume \(V^\phi_\nu(\Delta + 1) - V^\phi_\nu(\Delta) = \sigma^\phi_\nu\) where \(\sigma^\phi_\nu\) is independent of \(\Delta \geq 1\). Then, we will exam whether \(V^\phi_{\nu+1}(\Delta + 1) - V^\phi_{\nu+1}(\Delta)\) is also independent of \(\Delta \geq 1\). To this end, we have

\[
V^\phi_{\nu+1}(\Delta + 1) - V^\phi_{\nu+1}(\Delta) = C(\Delta + 1, 1) - ET\theta^\phi + \sum_{t=1}^{t_{\text{max}}} p_t P^t_{\Delta+1,\Delta+1+t}(1) V^\phi_\nu(\Delta + 1 + t) + p_t+ P^t_{\Delta+1,\Delta+1+t_{\text{max}}}(1) V^\phi(\Delta + 1 + t_{\text{max}}) - \left( C(\Delta, 1) - ET\theta^\phi + \sum_{t=1}^{t_{\text{max}}} p_t P^t_{\Delta+t}(1) V^\phi(\Delta + t) + p_t+ P^t_{\Delta+t_{\text{max}}}(1) V^\phi(\Delta + t_{\text{max}}) \right)
\]

\[
= C(\Delta + 1, 1) - C(\Delta, 1) + \sum_{t=1}^{t_{\text{max}}} p_t P^t_{\Delta+t}(1) \sigma^\phi_\nu + p_t+ P^t_{\Delta+t_{\text{max}}} \sigma^\phi_\nu.
\]
Moreover, we have
\[ \sum_{t=1}^{t_{\text{max}}} p_t P_{\Delta, \Delta+t}(1) + p_t+ P_{\Delta, \Delta+t_{\text{max}}}(1) = \sum_{t=1}^{t_{\text{max}}} p_t p(1 - p)^{t-1} + p_t+ (1 - p)^{t_{\text{max}}}. \]

Meanwhile,
\[ C(\Delta, 1) - C(\Delta - 1, 1) = \sum_{i=1}^{t_{\text{max}}} p_i \left( \frac{1 - (1 - p)^i}{p} \right) + p_t+ \left( \frac{1 - (1 - p)^{t_{\text{max}}}}{p} \right), \]
which is independent of \( \Delta \). Combining together, we can conclude that \( V^{\phi}_{\nu + 1}(\Delta + 1) - V^{\phi}_{\nu + 1}(\Delta) \) 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 by \( \sigma^\phi \) the constant. To obtain \( \sigma^\phi \), we have
\[
\sigma^\phi = \sum_{i=1}^{t_{\text{max}}} p_i \left( \frac{1 - (1 - p)^i}{p} \right) + p_t+ \left( \frac{1 - (1 - p)^{t_{\text{max}}}}{p} \right) + \left( \sum_{t=1}^{t_{\text{max}}} p_t p(1 - p)^{t-1} + p_t+ (1 - p)^{t_{\text{max}}} \right) \sigma^\phi.
\]
After some algebraic manipulations, we obtain
\[
\sigma^\phi = \sum_{i=1}^{t_{\text{max}}} p_i \left( \frac{1 - (1 - p)^i}{p} \right) + p_t+ \left( \frac{1 - (1 - p)^{t_{\text{max}}}}{p} \right)
\frac{1 - \left( \sum_{t=1}^{t_{\text{max}}} p_t p(1 - p)^{t-1} + p_t+ (1 - p)^{t_{\text{max}}} \right)}{1 - \left( \sum_{t=1}^{t_{\text{max}}} p_t p(1 - p)^{t-1} + p_t+ (1 - p)^{t_{\text{max}}} \right)}
\]  
\[ \square \]

\textit{b) Policy Improvement:} Here, we find policy resulting from the value function obtained in the previous step. 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 expected cost resulting from the operation of action \( a \) at state \( \Delta \). Then, for \( \Delta \geq 1 \), we have
\[
\delta V^{\phi}(\Delta) = \Delta - \theta^\phi + (1 - p)V^{\phi}(\Delta + 1) + pV^{\phi}(0) - V^{\phi,1}(\Delta)
\]
\[ = \Delta - \theta^\phi + (1 - p)V^{\phi}(\Delta + 1) + pV^{\phi}(0) - V^{\phi}(\Delta)
\]
\[ = \Delta - \theta^\phi + (1 - p)(V^{\phi}(\Delta + 1) - V^{\phi}(\Delta)) + p(V^{\phi}(0) - V^{\phi}(\Delta))
\]
\[ = \Delta - 2\theta^\phi + [(1 - p) - p(\Delta - 1)]\sigma^\phi.
\]
We notice that
\[ \delta V^{\phi}(\Delta + 1) - \delta V^{\phi}(\Delta) = 1 - p\sigma^\phi. \]
Plugging in the expression of $\sigma^\phi$ yields
\[
1 - po^\phi = 1 - \frac{\sum_{t=1}^{t_{max}} p_t(1 - (1 - p)^t) + p_{t+1}(1 - (1 - p)^{t_{max}})}{1 - \left(\sum_{t=1}^{t_{max}} p_t(1 - p)^{t-1} + p_{t+1}(1 - p)^{t_{max}}\right)} \\
\geq 1 - \frac{\sum_{t=1}^{t_{max}} p_t(1 - (1 - p)^t) + p_{t+1}(1 - (1 - p)^{t_{max}})}{1 - \left(\sum_{t=1}^{t_{max}} p_t(1 - p)^{t} + p_{t+1}(1 - p)^{t_{max}}\right)} \\
= 0.
\]
Consequently, $\delta V^\phi(\Delta + 1) \geq \delta V^\phi(\Delta)$ when $\Delta \geq 1$. Since $\theta^\phi = \bar{\Delta}_1$ and $\bar{\Delta}_1 \leq \frac{1 + (1 - p)\sigma^\phi}{2}$ by Condition 1, we know that
\[
\delta V^\phi(1) = 1 - 2\theta^\phi + (1 - p)\sigma^\phi = 1 - 2\bar{\Delta}_1 + (1 - p)\sigma^\phi \geq 0.
\]
Combining together, we have
\[
\delta V^\phi(\Delta) \geq \delta V^\phi(1) \geq 0, \quad \Delta \geq 1.
\]
Hence, the optimal action at state $(\Delta, 0, -1)$ when $\Delta \geq 1$ is to initiate the transmission. Now, the only missing part is the action at state $(0, 0, -1)$. To determine the action, we recall from Theorem 2 that the new policy will always be no worse than the old policies. Hence, it is sufficient to compare $\bar{\Delta}_1$ and $\bar{\Delta}_0$. Since $\bar{\Delta}_1 \leq \bar{\Delta}_0$ by Condition 1, we can conclude that the resulting policy is threshold policy with $\tau = 1$. Hence, the policy improvement step converges, meaning that the threshold policy with $\tau = 1$ is optimal.

**Appendix M**

**Algorithm 1** Policy Iteration Algorithm

**Require:**

Markov Decision Process: $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C})$

1. **procedure** POLICYITERATIONALGORITHM($\mathcal{M}$)
2. Initialize $\phi(s) \in \mathcal{A}$ and $V^\phi(s) \in \mathbb{R}$ for all $s \in \mathcal{S}$.
3. $(V^\phi(s), \theta^\phi) \leftarrow$ PolicyEvaluationStep($\mathcal{M}, \phi(s)$).
4. $\phi(s) \leftarrow$ PolicyImprovementStep($\mathcal{M}, V^\phi(s)$).
5. **if** $\phi(s)$ does not converges **then**
6. go to line 3
7. **return** $(\phi(s), \theta^\phi)$.