Age of Incorrect Information under Delay

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

Abstract

This paper investigates the problem of minimizing the Age of Incorrect Information (AoII) when the communication channel has a random delay. We consider a slotted-time system where a transmitter observes a dynamic source and decides when to send updates to a remote receiver through a channel with random delay. The receiver maintains estimates of the state of the dynamic source based on the received updates. In this paper, we adopt 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 considered problem using Markov Decision Process (MDP). Then, leveraging the policy improvement theorem and under an easy-to-verify condition, we prove that the optimal decision for the transmitter is to initiate a transmission whenever the channel is idle and AoII is not zero. The results apply to generic delay distribution. Lastly, we verify the condition numerically and provide the numerical results that highlight the performance of the optimal policy.

I. INTRODUCTION

The development of wireless technology and cheap sensors have made the real-time monitoring system widely used. Typically in such systems, a monitor monitors one or more events simultaneously and transmits updates to allow one or more receivers at a distance to have a good knowledge of the events. Therefore, in real-time monitoring systems, the timeliness of information is often one of the most important performance indicators. The Age of Information (AoI), first introduced in [1], captures the freshness of information by tracking the time elapsed since the generation of the last received update. More precisely, let $V(t)$ be the generation time of the last update received up to time $t$. Then, AoI at time $t$ is defined by $\Delta_{AoI}(t) = t - V(t)$. Hence, AoI is small when the receiver has relatively timely information. After the introduction, it has attracted extensive attention and research [2]–[4]. Although AoI captures well the timeliness of information, the limitation is that AoI ignores the information content of the transmitted updates. Therefore, it falls short in the context of remote estimation. For example, we want to
estimate a rapidly changing event remotely. In this case, a small AoI does not necessarily mean
that the receiver has accurate information about the event. Likewise, when the event changes
slowly, the receiver does not need very timely information to make relatively accurate estimates
of the event.

Inspired by the above limitations of AoI, the Age of Incorrect Information (AoII) is introduced
in [5], which combines the timeliness and the accuracy of information. As presented in [5],
AoII is dominated by two penalty functions. The first one is the time penalty function which
can be any function that captures the time elapsed since the last time the receiver has perfect
information about the remote event. The second one is the information penalty function which
can be any function that captures the mismatch between the receiver’s estimate and the actual
state of the remote event. Therefore, AoII captures not only the information mismatch between
the event and the receiver but also the aging process of inconsistent information. Moreover, by
choosing different penalty functions, AoII is adaptable to various systems and communication
goals. Hence, AoII is regarded as a semantic metric [6].

Several works have been done since the introduction of AoII. In [5], the authors consider
a simple penalty function choice and study the minimization of AoII under power constraints.
Then, the authors extend the discussions to the case of the generic time penalty function in [7].
Three applications are also studied to highlight the performance advantages of AoII over AoI and
real-time estimation error. In [8], the authors investigate the AoII that considers the quantified
information mismatch between the event and the receiver. The optimization problem is studied
when the system is resource-constrained. AoII in scheduling problems where a base station
observes multiple events and needs to select a part of the receivers to update is also studied. [9]
investigates the problem of minimizing AoII when the channel state information is available and
the time penalty function is generic. [10] considers the case where the base station cannot know
the states of the events before receiving the updates. In this paper, we consider the problem of
minimizing AoII when the communication channel has a random delay. A similar system setup
is considered in [11], where the problem is studied based on the simulation results of a simplified
relative value iterations algorithm. In this paper, we provide a theoretical analysis of the problem
and the results apply to any delay distribution. The system with a random delay communication
channel has also been studied in the context of remote estimation and AoI [12]–[15]. 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 work are: 1) We investigate the system where the communication channel has a random delay. 2) We consider the problem of minimizing AoII when the delay distribution is generic. 3) 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 model and formulate the considered problem. Then, in Section III we characterize the problem using Markov Decision Process and evaluate the performance of several policies. The derivation of the optimal policy is presented in Section IV. Finally, in Section V we laid out 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 send updates to the receiver through a channel. The dynamic source is modeled by a two-state symmetric Markov chain with state transition probability $p$. An illustration of the Markov chain is shown 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 the new one. Then, the transmitter needs to decide whether to transmit the new update based on the current status of the system. When the channel is idle (i.e., no transmission in progress), the transmitter chooses between transmitting the new update and staying idle. When the channel is busy (i.e., there is a transmission in progress), the transmitter can’t do anything other than staying idle. We assume the transmitter’s decision is independent of the transmission history. Each transmission will take a random amount of time $T \in \mathbb{N}^*$. More precisely, the transmission time $T$ of an update is a random variable, whose probability distribution is denoted by $p_t \triangleq Pr(T = t)$. We assume $1 \leq T \leq t_{max}$ and $T$ is independent and identically distributed over time. In practice, we can

![Fig. 1: The two-state symmetric Markov chain with transition probability $p$.](image-url)
always choose a sufficiently large $t_{max}$ so that the probability that the transmission time is greater than $t_{max}$ is negligible. Moreover, we assume the channel is error-free, which means that the update will not be corrupted during transmission.

The receiver will estimate the state of the dynamic source using the received updates. We denote by $\hat{X}_k$ the receiver’s estimate at time slot $k$. According to [16], 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. In this paper, we consider only the case of $p \leq \frac{1}{2}$. Hence, the receiver uses the last received update as the estimate. For the case of $p > \frac{1}{2}$, the results can be extended using the corresponding best estimator. The receiver uses $ACK/NACK$ packets to inform the transmitter of its reception of the new update. $ACK/NACK$ packets are generally very small [17]. Hence, we assume that they are reliably and instantaneously received by the transmitter. Then, 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. In this way, the transmitter always knows the current estimate at the receiver side. An illustration of the system model is shown in Fig. 2.

B. Age of Incorrect Information

The system adopts Age of Incorrect Information (AoII) as the performance metric. We first define $U_k$ as the last time instant before time $k$ (including $k$) that the receiver’s estimate is correct. Mathematically, $U_k$ is defined as the following.

$$U_k \triangleq \max\{h : h \leq k, X_h = \hat{X}_h\}.$$  

Then, in a slotted-time system, AoII at time slot $k$ can be written as

$$\Delta_{AoII}(X_k, \hat{X}_k, k) = \sum_{h=U_k+1}^{k} \left( g(X_h, \hat{X}_h) \cdot F(h - U_k) \right), \quad (1)$$
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 $T_2$.

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, $F(k) = 1$ for all $k$. We recall that the dynamic source is modeled by a two-state Markov chain. Therefore, $g(X_k, \hat{X}_k) \in \{0, 1\}$. Then, equation (1) can be simplified as

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

The dynamic of $\Delta_k$ can be captured by 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

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

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

C. System Dynamic

In this section, we tackle down the system dynamic. We notice that the status of system at time $k$ can be fully captured by the triplet $(\Delta_k, t_k, i_k)$ where $t_k \in \{0, 1, \ldots, t_{max} - 1\}$ indicates the time
the current transmission has been in progress. We define \( t_k = 0 \) if there is no transmission in progress. The last element \( i_k \in \{-1, 0, 1\} \) indicates the state of the channel. We define \( i_k = -1 \) when the channel is idle. \( i_k = 0 \) if the channel is busy and the transmitting update is the same as receiver’s estimate, and \( i_k = 1 \) when the transmitting update is different from the receiver’s estimate.

**Remark 1.** Note that, according to the definition of \( t_k \) and \( i_k \), \( i_k = -1 \) if and only if \( t_k = 0 \). In this case, the channel is idle.

Then, it is sufficient to characterize the value of \( (\Delta_{k+1}, t_{k+1}, i_{k+1}) \) using \( (\Delta_k, t_k, i_k) \) and the transmitter’s action at time \( k \). We denote by \( a_k \in \{0, 1\} \) the transmitter’s decision at time \( k \). We define \( a_k = 0 \) when the transmitter decides not to initiate a new transmission and \( a_k = 1 \) otherwise. We also define \( Pr(T > k + 1 \mid t) \) as the probability that the current transmission will take more than \( t + 1 \) time slots conditioned on the fact that the current transmission has been in progress for \( t \) time slots. Hence, \( Pr(T > k + 1 \mid t) \) is given by

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

where \( 0 \leq t \leq t_{\text{max}} - 1 \) and \( P_t \triangleq \sum_{k=1}^{t} p_k \). We recall that \( 1 \leq T \leq t_{\text{max}} \). Hence, \( p_t \) satisfies the following conditions.

\[
\begin{align*}
p_t &\geq 0 & 1 \leq t \leq t_{\text{max}}, \\
p_t = 0 & t > t_{\text{max}} \text{ or } t < 1, \\
\sum_{t=1}^{t_{\text{max}}} p_t = 1.
\end{align*}
\]

With the above definitions in mind, we proceed with characterizing the system dynamic. We first notice that \( \Delta_k \) will evolve according to (2). Hence, in the following, we will omit the discussion on the evolution of \( \Delta_k \). Then, we distinguish between the following cases.

- \( (\Delta_k, t_k, i_k) = (0, 0, -1) \). In this case, the channel is idle. Hence, the feasible action is \( a_k \in \{0, 1\} \). When the transmitter decides not to initiate a new transmission (i.e., \( a_k = 0 \)), \( i_{k+1} = i_k \) and \( t_{k+1} = t_k \) by definition. Hence, we have

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

\[
Pr[(1, 0, -1) \mid (0, 0, -1), a_k = 0] = p.
\]
When the transmitter decides to initiate a new transmission (i.e., \( a_k = 1 \)), the update will be delivered after a random amount of time \( T \). When \( T > 1 \), which happens with probability \( Pr(T > 1 \mid 0) \), the channel will be busy at the next time slot and \( t_{k+1} = 1 \) due to the transmission. Since the transmission happens when \( \Delta_k = 0 \), we know \( i_{k+1} = 0 \). Moreover, the receiver’s estimate will not change since no new update will be delivered. Hence, we have

\[
Pr[(0, 1, 0) \mid (0, 0, -1), a_k = 1] = Pr(T > 1 \mid 0)(1 - p) = (1 - p_1)(1 - p).
\]

\[
Pr[(1, 1, 0) \mid (0, 0, -1), a_k = 1] = Pr(T > 1 \mid 0)p = (1 - p_1)p.
\]

When \( T = 1 \), which happens with probability \( 1 - Pr(T > 1 \mid 0) \), the update will be delivered in the next time slot. Hence, the channel will be idle in the next time slot, which means that \( t_{k+1} = 0 \) and \( i_{k+1} = -1 \). Since the transmission started when \( \Delta_k = 0 \), the newly arrived update will bring no new information to the receiver. Hence, the receiver’s estimate will not change. Then, we have

\[
Pr[(0, 0, -1) \mid (0, 0, -1), a_k = 1] = (1 - Pr(T > 1 \mid 0))(1 - p) = p_1(1 - p).
\]

\[
Pr[(1, 0, -1) \mid (0, 0, -1), a_k = 1] = (1 - Pr(T > 1 \mid 0))p = p_1p.
\]

- \((\Delta_k, t_k, i_k) = (0, t, 0) \) where \( 1 \leq t \leq t_{max} - 1 \). In this case, the channel is busy. Hence, the feasible action is simply \( a_k = 0 \). When the update will not arrive in the next time slot, which happens with probability \( Pr(T > t + 1 \mid t) \), \( i_{k+1} \) will not change since both the transmitting update and the receiver’s estimate remain the same. Apparently, \( t_{k+1} = t_k + 1 \) as the transmission continues. Moreover, the receiver’s estimate will not change. Hence, we have

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

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

When the update will arrive in the next time slot, which happens with probability \( 1 - Pr(T > t + 1 \mid t) \), \( t_{k+1} = 0 \) and \( i_{k+1} = -1 \) by definition. Since \( i_k = 0 \), the newly arrived update will bring no new information to the receiver. Hence, we have

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

\[
Pr[(1, 0, -1) \mid (0, t, 0)] = (1 - Pr(T > t + 1 \mid t))p.
\]
• \((\Delta_k, t_k, i_k) = (0, t, 1)\) where \(1 \leq t \leq t_{\text{max}} - 1\). The analysis is very similar to the case of \((\Delta_k, t_k, i_k) = (0, t, 0)\) except that, when the update arrives, the receiver’s estimate will flip. Hence, we have

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

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

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

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

• \((\Delta_k, t_k, i_k) = (\Delta, 0, -1)\) where \(\Delta > 0\). In this case, the analysis is very similar to case of \((\Delta_k, t_k, i_k) = (0, 0, -1)\). Therefore, the details are omitted here. Then, we have

\[
Pr[(\Delta + 1, 0, -1) | (\Delta, 0, -1), a_k = 0] = 1 - p.
\]

\[
Pr[(0, 0, -1) | (\Delta, 0, -1), a_k = 0] = p.
\]

\[
Pr[(\Delta + 1, 1, 1) | (\Delta, 0, -1), a_k = 1] = Pr(T > 1 | 0)(1 - p) = (1 - p_1)(1 - p).
\]

\[
Pr[(0, 1, 1) | (\Delta, 0, -1), a_k = 1] = Pr(T > 1 | 0)p = (1 - p_1)p.
\]

\[
Pr[(\Delta + 1, 0, -1) | (\Delta, 0, -1), a_k = 1] = (1 - Pr(T > 1 | 0))p = p_1p.
\]

\[
Pr[(0, 0, -1) | (\Delta, 0, -1), a_k = 1] = (1 - Pr(T > 1 | 0))(1 - p) = p_1(1 - p).
\]

• \((\Delta_k, t_k, i_k) = (\Delta, t, 0)\) where \(\Delta > 0\) and \(1 \leq t \leq t_{\text{max}} - 1\). Following along the same line, we notice that the analysis is very similar to the case of \((\Delta_k, t_k, i_k) = (0, t, 0)\). Hence, we have

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

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

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

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

• \((\Delta_k, t_k, i_k) = (\Delta, t, 1)\) where \(\Delta > 0\) and \(1 \leq t \leq t_{\text{max}} - 1\). Again, the analysis is very similar to the case of \((\Delta_k, t_k, i_k) = (0, t, 1)\). Hence, we have

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

Combing the above cases, we fully characterized the system dynamic.

D. Problem Formulation

A policy \(\phi\) specifies the action that the transmitter will choose in each time slot. The objective of this paper is to find the policy that minimizes the expected AoII of the system. Therefore, the problem can be formulated as the following.

\[
\arg\min_{\phi \in \Phi} \lim_{K \to \infty} \frac{1}{K} E_{\phi} \left( \sum_{k=0}^{K-1} \Delta_k \right),
\]

(4)

where \(E_{\phi}\) 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 minimum expected AoII.

In the next section, we will characterize the problem reported in (4) using an infinite horizon with average cost Markov Decision Process (MDP). Then, we will evaluate the performance of several policies. Finally, in Section [IV] we will find the optimal policy.

III. MARKOV DECISION PROCESS CHARACTERIZATION

The problem reported in (4) can be characterized by an infinite horizon with average cost Markov Decision Process (MDP) \(M^\Upsilon\), where \(\Upsilon\) indicates the probability distribution of \(T\). More precisely, \(M^\Upsilon\) consists of the following components.

- The state space \(S^\Upsilon\). The state can simply be the triplet \(s = (\Delta, t, i)\) where \(\Delta \in \mathbb{N}^0\) is the current AoII. \(t\) and \(i\) are the \(t_k\) and \(i_k\) defined in Section [II-C] respectively. In the following, we will use \(s\) and \((\Delta, t, i)\) to represent the state interchangeably.
- The feasible action \(A^\Upsilon\). When \(i = -1\), the feasible action is \(a \in \{0, 1\}\) where \(a = 0\) if the transmitter decides not to initiate a new transmission and \(a = 1\) otherwise. When \(i \neq -1\), the feasible action is simply \(a = 0\).
• The state transition probabilities $P^T$. The probability that action $a$ at state $s$ leads to state $s'$ is denoted by $P^T_{s,s'}(a)$. Therefore, the values of $P^T_{s,s'}(a)$ are those detailed in Section II-C.

• The instant cost $C^T$. The instant cost at state $s$ is $C^T(s) = \Delta$.

Let $V^T(s)$ be the value function of state $s \in S^T$. It is well known that the value function satisfies the Bellman equation, which is given by the following.

$$V^T(s) + \theta^T = \min_{a \in A^T} \left\{ C^T(s) + \sum_{s' \in S^T} P^T_{s,s'}(a)V^T(s') \right\},$$  

(5)

where $\theta^T$ is the expected AoII achieved by the optimal policy. When we plug in the values of $P^T_{s,s'}(a)$, the Bellman equation will be complicated and hard to analyze directly. To ease the analysis, we notice that the feasible action depends on the state. When the channel is busy (i.e., $i \neq -1$), the only feasible action is $a = 0$. Hence, the minimum operators in (5) are avoided for these states. Leveraging this observation, we will rewrite the Bellman equation corresponding to the state with $i = -1$ in the following subsections.

A. Special Transmission Time Distribution

In this section, we rewrite the Bellman equation corresponding to the state with $i = -1$. In order to establish good building blocks for the case of generic transmission time distribution, we start with considering a special case where the transmission time $T$ of an update is deterministic. Mathematically, we consider the case of $p_{t_{\text{max}}} = 1$ where $t_{\text{max}} > 1$. To highlight the special distribution, we write $\Upsilon$ as $t_{\text{max}}$ for the remainder of this subsection.

Remark 2. When $t_{\text{max}} = 1$, the problem reduces to the problem considered in [5], according to which the optimal policy is the one that transmits a new update every time slot.

To start with, we define and calculate some auxiliary quantities. We first define $C_{t_{\text{max}}}^t(\Delta, a)$ as the expected sum of instant cost during the busy period of the channel caused by action $a = 1$ at state $(\Delta, 0, -1)$. Note that $C_{t_{\text{max}}}^t(\Delta, a)$ includes the instant cost at state $(\Delta, 0, -1)$.

Remark 3. In order to have a more intuitive understanding of the definition of $C_{t_{\text{max}}}^t(\Delta, a)$, we use $\eta$ to denote a sample path and let $H$ be the set of all such paths. Moreover, we denote by $C_\eta$ and $P_\eta$ the cost and probability associated with path $\eta$, respectively. Then, $C_{t_{\text{max}}}^t(\Delta, a)$ can be calculated as

$$C_{t_{\text{max}}}^t(\Delta, a) = \sum_{\eta \in H} P_\eta C_\eta.$$
As an example, we consider the case where $t_{\text{max}} = 2$ and action $a = 1$ is taken at state $(2, 0, -1)$. Then, a sample path of the state before the channel becomes idle again can be the following.

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

For the above sample path, $C_{\eta} = 2 + 3 = 5$ and $P_{\eta} = Pr[(3, 1, 1) \mid (2, 0, -1)] \cdot Pr[(4, 0, -1) \mid (3, 1, 1)].$

To obtain the closed-form expression for $C_{t_{\text{max}}}^t(\Delta, a)$, we denote by $p(t)$ the probability that the dynamic source will be in the same state after $t$ time slots. Since the Markov chain is symmetric, $p(t)$ is independent of the state and can be calculated by

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

where the subscript indicates the row number and the column number of the needed probability. For the consistency of notation, we define $p^{(0)} \triangleq 1$. Then, $C_{t_{\text{max}}}^t(\Delta, a)$ is given by the following lemma.

**Lemma 1.** $C_{t_{\text{max}}}^t(\Delta, a)$ is given by the following equation.

$$C_{t_{\text{max}}}^t(\Delta, a) = \begin{cases} \Delta & a = 0, \\
\sum_{t=0}^{t_{\text{max}}-1} C(\Delta, t) & a = 1,
\end{cases}$$

where

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

**Proof.** When $a = 0$, $C_{t_{\text{max}}}^t(\Delta, a)$ can be obtained easily from the definition and the system dynamic detailed in Section II-C. For the case of $a = 1$, we notice that $C(\Delta, t)$ is nothing but the expected AoII $t$ time slots after the start of transmission. Hence, $C(\Delta, t)$ can be obtained using the results in Section II-C. The complete derivation can be found in Appendix A.

We also define $P_{\Delta_{t_{\text{max}}}, \Delta}^t(a)$ as the probability that taking action $a$ at state $(\Delta, 0, -1)$ causes the system to reach state $(\Delta', 0, -1)$ for the first time.
Remark 4. The probability is the same as $P_\eta$ defined in Remark 3 in which case $P_\eta = P_2^2(1)$. This probability provides us with a direct transition between any two states with $i = -1$.

We first note that $P_{0,\Delta'}(0)$ can be obtained easily from the results in Section II-C. Then, we only need to consider the case of $a = 1$. Hence, we have the following lemma.

Lemma 2. $P_{0,\Delta'}(1)$ is given by the following equation.

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

$P_{\Delta,\Delta'}(1)$, where $\Delta > 0$, is given by the following equation.

$$P_{\Delta,\Delta'}(1) = \begin{cases} 
    p(t_{\text{max}}) & \Delta' = 0, \\
    (1 - p(t_{\text{max}}-1))(1 - p) & \Delta' = 1, \\
    (1 - p(t_{\text{max}}-k))p^2(1 - p)^{k-2} & \Delta' = k \in \{2, 3, \ldots, t_{\text{max}} - 1\}, \\
    p(1 - p)^{t_{\text{max}}-1} & \Delta' = \Delta + t_{\text{max}}, \\
    0 & \text{otherwise}.
\end{cases}$$

Proof. We recall that each update will take exactly $t_{\text{max}}$ time slots to be delivered. Then, we can derive the expressions leveraging the results in Section II-C. The complete derivation can be found in Appendix B.

By examining the expressions presented in Lemma 2, $P_{\Delta,\Delta'}(1)$ possesses the following properties. We will see later that these properties are essential for the following analysis.

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

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

Proof. The proof is obvious as we can verify easily using the closed-form expressions detailed in Lemma 2.
Leveraging Lemmas 1 and 2, we can rewrite the Bellman equation. We recall that the feasible action at the state with \( i \neq -1 \) is \( a = 0 \). Hence, for such state, the minimum operator in (5) is avoided. Then, by induction, the value function of state \( s = (\Delta, 0, -1) \) satisfies the following.

\[
V_{t_{\text{max}}}^t(\Delta) + \theta_{t_{\text{max}}}^t = \min_{a \in \{0, 1\}} \left\{ C_{t_{\text{max}}}^t(\Delta, a) - \theta_{t_{\text{max}}}^t(a) + \sum_{\Delta' \geq 0} P_{\Delta,\Delta'}(a)V_{t_{\text{max}}}^t(\Delta') \right\},
\]

where \( V_{t_{\text{max}}}^t(\Delta) \) is short for \( V_{t_{\text{max}}}^t(\Delta, 0, -1) \) and

\[
\theta_{t_{\text{max}}}^t(a) = \begin{cases} 
0 & a = 0, \\
(t_{\text{max}} - 1)\theta_{t_{\text{max}}}^t & a = 1.
\end{cases}
\]

Equation (6) provides a direct connection between states with \( i = -1 \). Hence, we can ignore the states with \( i \neq -1 \) in the analysis. In the following, we will extend the results to the case of generic transmission time distribution.

**B. Generic Transmission Time Distribution**

Using the quantities calculated in the previous subsection as building blocks, we rewrite the Bellman equation for the state with \( i = -1 \) when the probability distribution of \( T \) is generic. Similar to \( C_{t_{\text{max}}}^t(\Delta, a) \), we define \( C^T(\Delta, a) \) as the sum of expected instant costs during the busy period of the channel caused by the operation of \( a \) at state \( (\Delta, 0, -1) \) when the transmission time \( T \) is distributed according to \( T \). Leveraging the results in Lemma 1, we can easily conclude that

\[
C^T(\Delta, a) = \begin{cases} 
\Delta & a = 0, \\
t_{\text{max}} \sum_{t=1}^{t_{\text{max}}} p_t C^t(\Delta, a) & a = 1.
\end{cases}
\]

Similarly, we also define \( P_{\Delta,\Delta'}^T(a) \) as the probability that action \( a \) at state \( (\Delta, 0, -1) \) will result in state \( (\Delta', 0, -1) \) for the first time after the action. We first notice that, when \( a = 0 \), \( P_{\Delta,\Delta'}^T(0) \) can be obtained easily from the discussion in Section II-C. Then, we tackle down the case of \( a = 1 \). Leveraging the results in Lemma 2, we obtain

\[
P_{\Delta,\Delta'}^T(1) = \sum_{t=1}^{t_{\text{max}}} p_t P_{\Delta,\Delta'}^t(1).
\]

To get more insights into the transition probabilities, we have the following lemma.
Lemma 3. Equation (8) can be expressed equivalently as the following.

\[
P_{\Delta', \Delta}(1) = \begin{cases} 
\sum_{t=\Delta'}^{t_{\text{max}}} p_t P_{\Delta', \Delta'}^t(1), & 0 \leq \Delta' \leq t_{\text{max}} - 1, \Delta \geq \Delta', \\
\sum_{t=\Delta'}^{t_{\text{max}}} p_t P_{\Delta', \Delta'}^t(1) + p'_{t'} P_{\Delta', \Delta'}^{t'}(1), & 0 \leq \Delta' \leq t_{\text{max}} - 1, \Delta < \Delta', \\
p'_{t'} P_{\Delta', \Delta'}^{t'}(1), & \Delta' \geq t_{\text{max}},
\end{cases}
\]

where \( t' = \Delta' - \Delta \). Moreover, we define \( p'_{t'} P_{\Delta', \Delta'}^{t'}(1) \equiv 0 \) when \( t' \leq 0 \) or when \( t' > t_{\text{max}} \).

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 \( \Delta \geq \Delta' \).
2) \( P_{\Delta', \Delta}(1) = P_{\Delta + 1, \Delta' + 1}(1) \) when \( \Delta' \geq t_{\text{max}} \) and \( \Delta \geq 0 \).
3) \( P_{\Delta', \Delta}(1) = 0 \) when \( \Delta' > \Delta + t_{\text{max}} \) or when \( t_{\text{max}} - 1 < \Delta' < \Delta + 1 \).

Proof. The equivalent expression can be obtained easily using Lemma 2 and Proposition 1. Hence, the details are omitted here. In the following, we will focus on showing the presented properties. To this end, we prove the properties one by one.

1) When \( \Delta' = 0 \), \( P_{\Delta, \Delta}^0(1) = \sum_{t=1}^{t_{\text{max}}} p_t P_{\Delta, \Delta}^t(1) \). Since \( P_{\Delta, \Delta}^t(1) \) is independent of \( \Delta \), property 1 is true in this case. Then, we consider the case of \( 1 \leq \Delta' \leq t_{\text{max}} - 1 \) and \( \Delta \geq \Delta' \). We notice that

\[
P_{\Delta, \Delta'}^t(1) = \sum_{t=\Delta'}^{t_{\text{max}}} p_t P_{\Delta, \Delta'}^t(1).
\]

According to Property 2 of Proposition 1, \( P_{\Delta, \Delta'}^t(1) \) is independent of \( \Delta \). Hence, \( P_{\Delta, \Delta'}^t(1) \) is independent of \( \Delta \) in this case. Combining together, property 1 is true.

2) We notice that, when \( \Delta' \geq t_{\text{max}} \),

\[
P_{\Delta, \Delta'}^t(1) = p'_{t'} P_{\Delta, \Delta'}^{t'}(1) = p'_{t'} P_{\Delta, \Delta'+t'}(1).
\]

By Property 3 of Proposition 1, \( P_{\Delta, \Delta'+t'}(1) \) is independent of \( \Delta \). Then, we can conclude that \( P_{\Delta, \Delta'}^t(1) \) depends only on \( t' \). Thus, property 2 is true.

3) When \( \Delta' > \Delta + t_{\text{max}} \), the property holds due to Proposition 1. When \( t_{\text{max}} - 1 < \Delta' < \Delta + 1 \),

\[
P_{\Delta, \Delta'}^t(1) = p'_{t'} P_{\Delta, \Delta'}^{t'}(1),
\]

where \( t' \leq 0 \). Hence, \( P_{\Delta, \Delta'}^t(1) = 0 \) by definition. Then, property 3 is true.

\[\square\]
Let $V^\Upsilon(\Delta)$ be short for $V^\Upsilon(\Delta, 0, -1)$. Then, like we did in the previous subsection, the value function of state with $i = -1$ satisfies the following equation.

$$V^\Upsilon(\Delta) + \theta^\Upsilon = \min_{a \in \{0, 1\}} \left\{ C^\Upsilon(\Delta, a) - \theta^\Upsilon(a) + \sum_{\Delta' \geq 0} P^\Upsilon_{\Delta}\Delta'(a)V^\Upsilon(\Delta') \right\}, \quad (9)$$

where

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

and $ET$ is the expected transmission time of an update, which can be calculated by $ET \triangleq \sum_{t=1}^{t_{max}} t_p$. In the same spirit, we will, in the next subsection, evaluate the performance of several policies. As we will see later, the performance plays a crucial role in finding the optimal policy.

**C. Policy Evaluation**

In this subsection, we evaluate the performance of policy by calculating the achieved expected AoII. In particular, we consider threshold policy with different thresholds.

**Definition 2** (Threshold policy). Under 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.

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

**Remark 5.** When $\tau = 0$, the transmitter will initiate new transmission every time the channel is idle. We define $\tau \triangleq \infty$ as the policy under which the transmitter never initiates any transmissions.

We notice that threshold policy can be fully characterized by the threshold $\tau$. In the following, we will use $\tau$ to represent threshold policy. The system dynamic under threshold policy $\tau$ can be characterized by a discrete-time Markov chain. Let the state of the Markov chain be $s$. Without loss of generality, we assume the Markov chain starts at state $s = (0, 0, -1)$. Then, we can easily conclude that the states $s = (\Delta, 0, -1)$ with $\Delta \geq 0$ are positive recurrent under any policy. Hence, the steady-state probabilities of these states exist. We define $\pi_\Delta$ as the steady-state probability of state $s = (\Delta, 0, -1)$. Since the channel will remain idle if no transmission
is initiated and the expected transmission time of an update is \( ET \), \( \pi_\Delta \)'s satisfy the following equation:

\[
\sum_{i=0}^{\tau-1} \pi_i + ET \sum_{i=\tau}^{\infty} \pi_i = 1.
\]

Let \( \bar{\Delta}_\tau \) be the expected AoII achieved by threshold policy \( \tau \). Then, leveraging the quantity \( C^\tau(\Delta, a) \) derived in Section [III-B] we know that

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

(10)

Hence, it is sufficient to obtain the values of \( \pi_\Delta \)'s. Leveraging the results in section [III-B] we know that the transition probabilities between states \( s \) with \( i = -1 \) are given by \( P^\tau_{\Delta, \Delta'}(a) \). The problem arises since there are infinity many \( \pi_\Delta \)'s to calculate. To overcome the infinity, we define \( \Pi \triangleq \sum_{i=\omega}^{\infty} \pi_i \) where \( \omega = t_{\text{max}} + \tau \). As we will see later, \( \Pi \) and \( \pi_\Delta \) for \( 0 \leq \Delta < \omega - 1 \) are sufficient to obtain the expected AoII. With this in mind, \( \pi_\Delta \)'s and \( \Pi \) can be calculated using the following proposition.

**Proposition 2** (Steady-state probability). For \( 0 \leq \tau < \infty \), \( \Pi \) and \( \pi_\Delta \) for \( 0 \leq \Delta < \omega - 1 \) are the solution to the following system of linear equations.

\[
\begin{align*}
\pi_0 &= (1 - p)\pi_0 + p \sum_{i=1}^{\tau-1} \pi_i + P^\tau_{0,0}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right), \\
\pi_1 &= p\pi_0 + P^\tau_{1,1}(1) \left( \sum_{i=\tau}^{\omega-1} \pi_i + \Pi \right).
\end{align*}
\]

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

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

\[
\pi_{t_{\text{max}}} = \begin{cases} 
(1 - p)\pi_{t_{\text{max}}-1} & t_{\text{max}} - 1 < \tau, \\
\sum_{i=\tau}^{t_{\text{max}}-1} P^\tau_{i,t_{\text{max}}}(1)\pi_i & t_{\text{max}} - 1 \geq \tau.
\end{cases}
\]

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

\[
\pi_\Delta = \begin{cases} 
(1 - p)\pi_{\Delta-1} & \Delta - 1 < \tau, \\
\sum_{i=\tau}^{\Delta-1} P^\tau_{i,\Delta}(1)\pi_i & \Delta - 1 \geq \tau.
\end{cases}
\]
\[ \Pi = \sum_{i=\tau}^{\omega-1} \left( \sum_{k=\tau}^{i} P^Y_{i,t_{max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{max}} \left( P^X_{\omega,\omega+1}(1) \right) \Pi. \]

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

When \( \tau = 0, \omega = t_{max} \). Then, \( \Pi \) and \( \pi_\Delta \) for \( 0 \leq \Delta < \omega - 1 \) are given by

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

\[ \Pi = \frac{\sum_{i=0}^{t_{max}-1} \left( \sum_{k=0}^{i} P^Y_{i,t_{max}+k}(1) \right) \pi_i}{1 - \sum_{i=1}^{t_{max}} P^Y_{t_{max},t_{max}+i}(1)}. \]

When \( \tau = 1, \omega = t_{max} + 1 \). Then, \( \Pi \) and \( \pi_\Delta \) for \( 0 \leq \Delta < \omega - 1 \) are given by

\[ \pi_0 = \frac{P^Y_{1,0}(1)}{pET + P^Y_{1,0}(1)}. \]

\[ \pi_1 = \frac{pP^Y_{1,0}(1) + pP^Y_{1,1}(1)}{pET + P^Y_{1,0}(1)}. \]

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

\[ \pi_{t_{max}} = \sum_{i=1}^{t_{max}-1} P^Y_{i,t_{max}+1}(1) \pi_i. \]

\[ \Pi = \frac{\sum_{i=1}^{t_{max}} \left( \sum_{k=1}^{i} P^Y_{i,t_{max}+k}(1) \right) \pi_i}{1 - \sum_{i=1}^{t_{max}} P^Y_{t_{max},t_{max}+i}(1)}. \]

**Proof.** We notice that \( \pi_\Delta \)'s satisfy the balance equations, whose state transition probabilities are given by \( P^Y_{\Delta,\Delta'}(a) \). Hence, we can solve the balance equations to obtain \( \pi_\Delta \). The problem arises since there are infinitely many \( \pi_\Delta \)'s. To overcome this, we leverage the quantity \( \Pi \). By such grouping, we only need to solve a system of linear equations with finite size. The complete calculation can be found in Appendix [C].

As the steady-state probabilities are given, we proceed with evaluating \( \bar{\Delta}_\tau \) for different \( \tau \). Leveraging [10], the expected AoII can be calculated using the following proposition.
Proposition 3 (Threshold policy performance). For $0 \leq \tau < \infty$, the expected AoII resulting from the adoption of threshold policy $\tau$ is given by

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

where $\omega = t_{\text{max}} + \tau$ and

$$\Sigma = \sum_{t=1}^{t_{\text{max}}} \left[ p_t P_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i \right) + \Pi_t \left( \sum_{t=1}^{t_{\text{max}}} p_t [t - t((1 - p))^{i}] \right) \right] \left( 1 - \sum_{t=1}^{t_{\text{max}}} (p_t P_{1,1+t}(1)) \right),$$

$$\Pi_t = p_t P_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} \pi_i + \Pi \right), \quad 1 \leq t \leq t_{\text{max}}.$$ 

$\Pi$ and $\pi_\Delta$ for $0 \leq \Delta < \omega - 1$ are those detailed in Proposition 2.

Proof. We recall that the expected AoII can be calculated using (10). The problem arises because of the infinite sum. To overcome this, we adopt a similar approach as proposed in proof of Proposition 2. More precisely, we define $\Sigma = \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i$. The complete calculation can be found in Appendix D.

Remark 6. Note that the closed-form expressions for both $\Delta_0$ and $\Delta_1$ are given. For $\tau \in \mathbb{N}^* \setminus \{0, 1\}$, $\bar{\Delta}_\tau$ can be obtained by solving a system of linear equations with size $\omega + 1$.

IV. OPTIMAL POLICY

As the performance of the threshold policy is evaluated, we will, in the section, find the optimal policy for $\mathcal{M}^T$. First of all, we need to show that the optimal policy exists.

A. Existence of Optimal Policy

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

Proof. The proof is based on the results presented in [18]. More precisely, we show that $\mathcal{M}^T$ verifies the two conditions given in [18]. Then, the existence of optimal policy is guaranteed. The complete proof can be found in Appendix E. 

March 8, 2022 DRAFT
As the optimal policy exists, we proceed with finding the optimal policy. To this end, we leverage **Policy Improvement Theorem**. To start with, we first introduce **Policy Iteration Algorithm**, which is an iterative algorithm consisting of two steps.

1) The first step is *policy evaluate step*. In this step, we calculate the value function $V^\psi(\cdot)$ and the expected AoII $\theta^\psi$ resulting from the adoption of policy $\psi$. More precisely, we solve the system of linear equations defined by the following equation.

$$
V^\psi(s) + \theta^\psi = C^T(s) + \sum_{s' \in \mathcal{S}^T} P^T_{s,s'}(\psi(s))V^\psi(s'), \quad s \in \mathcal{S}^T,
$$

where $\psi(s)$ is the action suggested by policy $\psi$ at state $s$. Note that this is a underdetermined system. Hence, we can select a reference state $s^{ref}$ arbitrarily and set $V^\psi(s^{ref}) = 0$. In this way, we can obtain a unique solution.

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

$$
\psi'(s) = \arg \min_a \left\{ C^T(s) + \sum_{s' \in \mathcal{S}^T} P^T_{s,s'}(a)V^\psi(s') \right\}.
$$

The pseudocode of policy iteration algorithm can be found in Appendix I. Then, the policy improvement theorem tells us the following.

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

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

**Proof.** The proof is based on the results presented in [19]. The complete proof can be found in Appendix F. \square

B. Optimality Proof

In this subsection, we find the optimal policy for $\mathcal{M}^T$. To start with, we present the inequality that is important for the analysis later on.
**Condition 1.** The condition is the following.

\[
\Delta_1 \leq \min\left\{ \Delta_0, \frac{1 + (1-p)\sigma^\tau}{2} \right\},
\]

where

\[
\sigma^\tau = \frac{\sum_{t=1}^{t_{max}} (p_t - p_t(1-p)^t)}{p - p^2 \sum_{t=1}^{t_{max}} p_t(1-p)^{t-1}}.
\]

We recall that the closed-form expressions for both \(\Delta_1\) and \(\Delta_0\) are given by Proposition 3. Hence, the inequality is easy to verify. At the end of this section, we will verify the condition analytically when \(p_t = \frac{1}{2}\) for \(t \in \{1, 2\}\). For other transmission time distributions, we can verify numerically as did in Section V-A. Now, we assume that Condition 1 holds. Then, the optimal policy is given by the following theorem.

**Theorem 3** (Optimal policy). When Condition 1 is verified, the optimal policy is threshold policy with \(\tau = 1\).

**Proof.** According to Theorem 2, it is sufficient to show that the policy iteration algorithm converges to threshold policy with \(\tau = 1\). The complete proof can be found in Appendix G.

**Proposition 4** (Verification of a special case). Condition 1 holds for \(p \leq \frac{1}{2}\) when \(p_t = \frac{1}{2}\), where \(t \in \{1, 2\}\). Consequently, the optimal policy in this case is threshold policy with \(\tau = 1\).

**Proof.** The verification is simple because we have the closed-form expressions for all the quantities involved. The complete verification can be found in Appendix H.

For other transmission time distributions, we can verify Condition 1 numerically. The numerical results when \(T\) is geometrically distributed are given in Section V-A.

**V. NUMERICAL RESULTS**

In this section, we present the numerical results for the verification of Condition 1. Meanwhile, we provide the performance of the optimal policy under various system settings to highlight the characteristics of the considered problem.
A. Verification of Condition

We first verify Condition I numerically. We notice that it is equivalent to verify the following two inequalities.

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

The numerically results are shown in Figs. 4 and 5. More precisely, Fig. 4 shows the results when \( T \) is geometrically distributed with success probability \( p_s = 0.6 \). Meanwhile, we force \( p_{t_{\text{max}}} = 1 - \sum_{t=1}^{t_{\text{max}}-1} p_t \) to meet the assumptions given in (3). We vary the value of \( p \) and plot the results. From the figure, we can see that both inequalities are verified for \( p \in [0.025, 0.5] \) under all the cases considered. Fig. 5 shows the results when \( p = 0.35 \) and \( T \) is geometrically distributed. We vary the value of success probability \( p_s \) and plot the corresponding results. Same as before, we force \( p_{t_{\text{max}}} = 1 - \sum_{t=1}^{t_{\text{max}}-1} p_t \). We can conclude from the plot that both inequalities are verified for \( p_s \in [0, 1] \) under all the cases considered. For other transmission time distributions, the inequalities can be verified in a very similar way.

Remark 7. We notice that, in Fig. 5b, \( \bar{\Delta}_1 = \bar{\Delta}_0 \) when \( p_s = 1 \). We first recall that when \( p_s = 1 \), the update is bound to arrive after one time slot. Hence, the problem is equivalent to the problem considered in [5]. As proved in [5], threshold policies with \( \tau = 1 \) and \( \tau = 0 \) are indeed equivalent and are both optimal.

B. Optimal Policy Performance

Here, we show the performance of the optimal policy when \( T \) is geometrically distributed. We force \( t_{\text{max}} = 1 - \sum_{t=1}^{t_{\text{max}}-1} p_t \) to meet the assumptions given in (3). 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 the following. When the transmission time distribution is fixed and as \( p \) increases, the transmitted updates will more likely be incorrect by the time they arrive. This will result in a faster aging process during the transmission. Mathematically, we can show that \( C_T(\Delta, 1) \) 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 as \( t_{\text{max}} \) increases, the transmitted updates will more likely be incorrect by the time they arrive. Again, we can verify that \( C_T(\Delta, 1) \) is increasing in
(a) Verification for $\bar{\Delta}_1 - \frac{1+(1-p)\sigma^2}{2} \leq 0$.

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

Fig. 4: Condition 1 verification. Different lines represent different values of $t_{max}$. In this case, we fix $\Upsilon$ to a geometric distribution with success parameter $p_s = 0.6$ and plot the results under various $p$.

(a) Verification for $\bar{\Delta}_1 - \frac{1+(1-p)\sigma^2}{2} \leq 0$.

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

Fig. 5: Condition 1 verification. Different lines represent different values of $t_{max}$. In this case, we fix $p = 0.35$ and $\Upsilon$ to the geometric distribution. We plot the results under various success probability $p_s$. 
Fig. 6: The performance of the optimal policy when the transmission time $T$ is geometrically distributed. Different lines represent different values of $t_{\text{max}}$. We force $t_{\text{max}} = 1 - \sum_{t=1}^{t_{\text{max}}-1} p_t$ to meet the assumptions in (3).

Hence, the expected AoII will increase. We also notice that $p^{(t)}$ will converge to $\frac{1}{2}$ as $t$ increases, in which case the correctness of the transmitting update is independent of the past. Hence, $t_{\text{max}}$ will have mild impact on the expected AoII when $t_{\text{max}}$ is large. When $p = 0.5$, the state of the dynamic source is independent of the past. Hence, the transmission time has no impact on the performance. In Fig. 6b, we fix $p = 0.6$ and vary the value of $p_s$. As shown in the figure, 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 more likely be correct when they arrive. We also notice that the expected AoII increases as $t_{\text{max}}$ increases, and the impact of $t_{\text{max}}$ will be weak as $t_{\text{max}}$ becomes large. The reason is the same as the one stated for Fig. 6a.

VI. Conclusion

In this paper, we study the problem of minimizing AoII in a system where the communication channel has a random delay. More precisely, we investigate the transmitter-receiver pair in a slotted-time system where the transmitter observes a dynamic source and needs to decide when to send updates through a communication channel to a remote receiver. The receiver estimates the actual state of the dynamic source based on the received updates. The dynamic source is modeled
by a two-state symmetric Markov chain and the channel has a random delay. We study the problem of minimizing AoII by controlling the transmitter’s action in each time slot. Leveraging the notation of MDP, we characterize the considered problem and evaluate the performance of threshold policy with various thresholds. When the inequality in Condition I holds, using the policy improvement theorem, we prove that the optimal policy is the one that initiates a transmission whenever the channel is idle and AoII is not zero. The theoretical results apply to generic delay distribution. Then, in Section V we verify Condition I numerically and provide the numerical results that showcase the performance of the optimal policy.

REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

[6] 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.

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

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

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

[10] 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.

[11] 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.

[12] 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), 2017, pp. 321–325.

[13] Y. Sun, Y. Polyanskiy, and E. Uysal, “Sampling of the wiener process for remote estimation over a channel with random delay,” IEEE Transactions on Information Theory, vol. 66, no. 2, pp. 1118–1135, 2019.
We first consider the case of $a = 0$. In this case, the transmitter decides not to initiate a new transmission. Hence, the channel will remain idle in the next time slot. Then, we can easily obtain that $C_{t_{\text{max}}}^t(\Delta, 0) = \Delta$.

Then, we tackle down the case of $a = 1$. In this case, the update will take $t_{\text{max}}$ time slots to be delivered. Hence, $C_{t_{\text{max}}}^t(\Delta, 1)$ can be calculated as

$$C_{t_{\text{max}}}^t(\Delta, 1) = \sum_{t=0}^{t_{\text{max}}-1} C(\Delta, t),$$

where $C(\Delta, t)$ is the expected instant cost $t$ time slots after the start of transmission. To derive the expression for $C(\Delta, t)$, we first note that the receiver’s estimate remains unchanged during the transmission. With this in mind, we distinguish between the following cases.

- The transmission starts at state with $\Delta = 0$. In this case, the transmitted update is the same as the receiver’s estimate.
  - When $t = 0$, $C(0, 0) = 0$.
  - For each $1 \leq t \leq t_{\text{max}} - 1$, $C(0, t) = 0$ when the dynamic source remains the same state after $t$ time slots, which happens with probability $p^{(t)}$. $C(0, t) = k$ for $k \in \{1, 2, \ldots, t\}$ when the estimate is correct at the $(t - k)$th time slot after the transmission and then, the receiver’s estimate remains incorrect for the following $k$ time slots. Since $\Delta = 0$, this happens with probability $p^{(t-k)}p(1-p)^{k-1}$.

---

**APPENDIX A**

**PROOF OF LEMMA**

We first consider the case of $a = 0$. In this case, the transmitter decides not to initiate a new transmission. Hence, the channel will remain idle in the next time slot. Then, we can easily obtain that $C_{t_{\text{max}}}^t(\Delta, 0) = \Delta$.

Then, we tackle down the case of $a = 1$. In this case, the update will take $t_{\text{max}}$ time slots to be delivered. Hence, $C_{t_{\text{max}}}^t(\Delta, 1)$ can be calculated as

$$C_{t_{\text{max}}}^t(\Delta, 1) = \sum_{t=0}^{t_{\text{max}}-1} C(\Delta, t),$$

where $C(\Delta, t)$ is the expected instant cost $t$ time slots after the start of transmission. To derive the expression for $C(\Delta, t)$, we first note that the receiver’s estimate remains unchanged during the transmission. With this in mind, we distinguish between the following cases.

- The transmission starts at state with $\Delta = 0$. In this case, the transmitted update is the same as the receiver’s estimate.
  - When $t = 0$, $C(0, 0) = 0$.
  - For each $1 \leq t \leq t_{\text{max}} - 1$, $C(0, t) = 0$ when the dynamic source remains the same state after $t$ time slots, which happens with probability $p^{(t)}$. $C(0, t) = k$ for $k \in \{1, 2, \ldots, t\}$ when the estimate is correct at the $(t - k)$th time slot after the transmission and then, the receiver’s estimate remains incorrect for the following $k$ time slots. Since $\Delta = 0$, this happens with probability $p^{(t-k)}p(1-p)^{k-1}$.

---

[14] T. Z. Ornee and Y. Sun, “Sampling for remote estimation through queues: Age of information and beyond,” in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2019, pp. 1–8.

[15] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, “Towards an effective age of information: Remote estimation of a markov source,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 367–372.

[16] ——, “Towards an effective age of information: Remote estimation of a markov source,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 367–372.

[17] 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.

[18] L. I. Sennott, “Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs,” *Operations Research*, vol. 37, no. 4, pp. 626–633, 1989.

[19] R. A. Howard, “Dynamic programming and markov processes.” 1960.

[20] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
Combining together, we obtain the following.

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

- The transmission starts at state with \( \Delta > 0 \). In this case, the transmitted update is different from the receiver’s estimate.
  - When \( t = 0 \), \( C(\Delta, 0) = \Delta \).
  - For each \( 1 \leq t \leq t_{\text{max}} - 1 \), \( C(\Delta, t) = 0 \) when the dynamic source remains the same state after \( t \) time slots, which happens with probability \( p(t) \). \( C(\Delta, t) = k \) for \( k \in \{1, 2, \ldots, t - 1\} \) when the estimate is correct at the \((t - k)\)th time slot after the transmission and then, the receiver’s estimate remains incorrect for the following \( k \) time slots. Since \( \Delta > 0 \), this happens with probability \( (1 - p)^t \).

Combining together, we obtain the following.

\[ C(\Delta, t) = \sum_{k=1}^{t-1} k(1 - p^{(t-k)})p(1 - p)^{k-1} + (t + \Delta)(1 - p)^t, \quad \Delta > 0. \]

Hence, we have

\[ C_{t_{\text{max}}}^{\Delta}(\Delta, a) = \begin{cases} 
\Delta & a = 0, \\
\sum_{t=0}^{t_{\text{max}}-1} C(\Delta, t) & a = 1,
\end{cases} \]

where

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

**APPENDIX B**

**PROOF OF LEMMA 2**

We recall that each transmission will take exactly \( t_{\text{max}} \) time slots to be delivered. Then, we can derive the expressions leveraging the results in Section II-C. To this end, we distinguish between the following cases.
When $\Delta = 0$ and $a = 1$, the transmitted update is the same as the receiver’s estimate. Hence, the receiver’s estimate will not change as a result of receiving the transmitted update. Moreover, according to the dynamic of AoII reported in (2), we know that $\Delta' \in \{0, 1, 2, \ldots, t_{\text{max}}\}$.

With these in mind, we further distinguish our discussion into the following cases.

- $\Delta' = 0$ happens when the receiver’s estimate is correct as a result of receiving the update. Hence, the probability of this happening is $p(t_{\text{max}})$.

- $\Delta' = k \in \{1, 2, \ldots, t_{\text{max}}\}$ happens when the receiver’s estimate is correct at $(t_{\text{max}} - k)$th time slot after the transmission, which happens with probability $p(t_{\text{max}} - k)$. Then, the estimate remains incorrect for the remainder of the transmission time. This happens when the source first changes state, then remains in the same state throughout the rest of the transmission. Hence, the probability of this happening is $p(1 - p)^k$. Combining together, $\Delta' = k$ with probability $p(t_{\text{max}} - k) p(1 - p)^k - 1$.

Combining together, we have

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

- When $\Delta > 0$ and $a = 1$, the transmitted update is different from the receiver’s estimate. Hence, the receiver’s estimate will flip as a result of receiving the transmitted update. As indicated by (2), $\Delta$ will either increases by one or decreases to zero. Hence, $\Delta' \in \{0, 1, 2, \ldots, t_{\text{max}} - 1, \Delta + t_{\text{max}}\}$. With these in mind, we further distinguish between the following cases.

  - $\Delta' = 0$ happens in the same case as discussed in the previous case. Since the receiver’s estimate will flip upon the arrival of the transmitted update, the flipped estimate is correct with probability $p(t_{\text{max}})$.

  - $\Delta' = 1$ happens when the estimate is correct at $(t_{\text{max}} - 1)$th time slot after the transmission, which happens with probability $1 - p(t_{\text{max}} - 1)$. Then, the estimate becomes incorrect as a result of receiving the update. Since the estimate flips upon the arrival of the transmitted update, it happens when the source remains the same state. Hence, the probability of this happening is $1 - p$. Combining together, $\Delta' = 1$ with probability $(1 - p(t_{\text{max}} - 1))(1 - p)$. 

\(- \Delta' = k \in \{2, 3, \ldots, t_{\text{max}} - 1\}\) happens when the estimate is correct at \((t_{\text{max}} - k)\)th time slot after the transmission, which happens with probability \(1 - p^{(t_{\text{max}} - k)}\). Then, the estimate remains incorrect for the remainder of the transmission time. This happens when the dynamic source behaves in the following way during the remaining transmission time. The dynamic source should first change state, then remain the same state, and finally change state again when the update arrives. This happens with probability \(p^2 (1 - p)^{k-2}\). Hence, \(\Delta' = k\) with probability \((1 - p^{(t_{\text{max}} - k)}) p^2 (1 - p)^{k-2}\).

\(- \Delta' = \Delta + t_{\text{max}}\) happens when the estimate is incorrect throughout the transmission. Since the estimate will flip when the update is received, this happens when the source doesn’t change state until the update arrives. Hence, \(\Delta' = \Delta + t_{\text{max}}\) with probability \(p (1 - p)^{t_{\text{max}} - 1}\).

Combining together, we have

\[
P^{t_{\text{max}}}_{\Delta, \Delta'}(1) = \begin{cases} 
 p^{(t_{\text{max}})} & \Delta' = 0, \\
 (1 - p^{(t_{\text{max}} - 1)})(1 - p) & \Delta' = 1, \\
 (1 - p^{(t_{\text{max}} - k)}) p^2 (1 - p)^{k-2} & \Delta' = k \in \{2, 3, \ldots, t_{\text{max}} - 1\}, \\
 p(1 - p)^{t_{\text{max}} - 1} & \Delta' = \Delta + t_{\text{max}}, \\
 0 & \text{otherwise}.
\end{cases}
\]

**APPENDIX C**

**PROOF OF PROPOSITION 2**

We recall that \(\pi_\Delta\)'s satisfy the balance equations, whose state transition probabilities are given by \(P^T_{\Delta, \Delta'}(a)\). Then, by plugging in the values detailed in Section III-B, \(\pi_\Delta\)'s satisfy the following balance equations.

\[
\pi_0 = (1 - p)\pi_0 + p \sum_{i=1}^{\tau-1} \pi_i + \sum_{i=\tau}^{\infty} P^T_{i,0}(1)\pi_i = (1 - p)\pi_0 + p \sum_{i=1}^{\tau-1} \pi_i + P^T_{0,0}(1) \sum_{i=\tau}^{\infty} \pi_i. \quad (11)
\]

\[
\pi_1 = p\pi_0 + \sum_{i=\tau}^{\infty} P^T_{i,1}(1)\pi_i = p\pi_0 + P^T_{1,1}(1) \sum_{i=\tau}^{\infty} \pi_i. \quad (12)
\]

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

\[
\pi_\Delta = \begin{cases} 
 (1 - p)\pi_{\Delta-1} + P^T_{\tau, \Delta}(1) \sum_{i=\tau}^{\infty} \pi_i & \Delta - 1 < \tau, \\
 \sum_{i=\tau}^{\Delta-1} P^T_{i,\Delta}(1)\pi_i + P^T_{\Delta, \Delta}(1) \sum_{i=\Delta}^{\infty} \pi_i & \Delta - 1 \geq \tau.
\end{cases} \quad (13)
\]
\[ \pi_{t_{\text{max}}} = \begin{cases} (1 - p)\pi_{t_{\text{max}} - 1} & \text{if } t_{\text{max}} - 1 < \tau, \\ \sum_{i=\tau}^{t_{\text{max}} - 1} P_{i,t_{\text{max}}}^\tau (1)\pi_i & \text{if } t_{\text{max}} - 1 \geq \tau. \end{cases} \]

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

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

\[ \pi_{\Delta} = \sum_{i=\Delta-t_{\text{max}}}^{\Delta-1} P_{i,\Delta}^\tau (1)\pi_i, \quad \Delta \geq \tau + t_{\text{max}}. \quad (14) \]

Note that we can pull the state transition probabilities in (11), (12), and (13) out of the summation due to Property 1 of Lemma 3. The problem arises since there are infinitely many balance equations to solve. To overcome the infinity, we define \( \omega \equiv \tau + t_{\text{max}} \) and sum (14) over \( \Delta \) from \( \omega \) to \( \infty \). Then, we have

\[ \sum_{i=\omega}^{\infty} \pi_i = \sum_{i=\omega}^{\infty} \sum_{k=1-t_{\text{max}}}^{\omega-1} P_{k,i}^\tau (1)\pi_k. \quad (15) \]

We dive deep into the RHS of (15). To this end, we expand the first summation, which yields

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

Then, we rearrange the summation.

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

Leveraging Property 2 of Lemma 3, we have

\[ RHS = \sum_{i=1}^{\omega-1} \left( \sum_{k=1}^{t_{\text{max}}} P_{i,\omega+k}^\tau (1) \right) \pi_i + \sum_{i=1}^{t_{\text{max}}} \left( P_{i+1,\omega}^\tau (1) \right) \left( \sum_{k=\omega}^{\infty} \pi_k \right). \]
We define \( \Pi \triangleq \sum_{i=\omega}^{\infty} \pi_i \). Then,

\[
\Pi = \sum_{i=\tau}^{\omega-1} \left( \sum_{k=\tau}^{i} P_{i,t_{\max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{\max}} P_{\omega,\omega+t}(1) \Pi. \tag{16}
\]

Finally, replacing (14) with (16) and applying the definition of \( \Pi \) to the balance equations yield a system of linear equations with finite size as presented in the Proposition. Hence, we can obtain \( \pi_{\Delta} \) for \( 0 \leq \Delta \leq \omega - 1 \) and \( \Pi \) easily.

In the following, we consider two special cases where \( \tau = 0 \) and \( \tau = 1 \). We start with the case of \( \tau = 0 \). In this case, \( \omega = t_{\max} \) and the balance equations reduce to the following.

\[
\pi_{\Delta} = \sum_{i=0}^{\infty} P_{i,\Delta}^{Y}(1) \pi_i = \sum_{i=0}^{\Delta-1} P_{i,\Delta}^{Y}(1) \pi_i + P_{\Delta,\Delta}(1) \sum_{i=\Delta}^{\infty} \pi_i, \quad 0 \leq \Delta \leq t_{\max} - 1. \tag{17}
\]

\[
\Pi = \sum_{i=0}^{t_{\max}-1} \left( \sum_{k=0}^{i} P_{i,t_{\max}+k}(1) \right) \pi_i + \sum_{i=1}^{t_{\max}} P_{t_{\max},t_{\max}+1}(1) \Pi. \tag{18}
\]

\[
ET \sum_{i=0}^{\infty} \pi_i = 1. \tag{19}
\]

We first combine (17) and (19), which yields

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

Moreover, according to (18), we obtain

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

Then, we consider the case of \( \tau = 1 \). In this case, \( \omega = t_{\max} + 1 \) and the balance equations reduce to the following.

\[
\pi_0 = (1 - p) \pi_0 + P_{1,0}^{Y}(1) \sum_{i=1}^{\infty} \pi_i. \tag{20}
\]

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

\[
\pi_{\Delta} = \sum_{i=1}^{\infty} P_{i,\Delta}^{Y}(1) \pi_i = \sum_{i=1}^{\Delta-1} P_{i,\Delta}^{Y}(1) \pi_i + P_{\Delta,\Delta}(1) \sum_{i=\Delta}^{\infty} \pi_i, \quad 2 \leq \Delta \leq t_{\max} - 1.
\]

\[
\pi_{t_{\max}} = \sum_{i=1}^{t_{\max}-1} P_{i,t_{\max}}^{Y}(1) \pi_i.
\]
\[ \Pi = \sum_{i=1}^{t_{\text{max}}} \left( \sum_{k=1}^{i} P_{i,t_{\text{max}}+k}^T \right) \pi_i + \sum_{i=1}^{t_{\text{max}}} P_{t_{\text{max}}+1,t_{\text{max}}+1+i}^T \pi_i. \] (21)

\[ \pi_0 + ET \sum_{i=1}^{\infty} \pi_i = 1. \] (22)

To get the stationary distribution, we first combine (20) and (22). Then, we obtain

\[ \pi_0 = (1 - p)\pi_0 + P_{1,0}^T(1) \left( \frac{1 - \pi_0}{ET} \right). \]

Hence, we have

\[ \pi_0 = \frac{P_{1,0}^T(1)}{pE[T] + P_{1,0}^T(1)}. \]

Similarly,

\[ \pi_1 = \frac{pP_{1,0}^T(1) + pP_{1,1}^T(1)}{pET + P_{1,0}^T(1)}. \]

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

\[ \pi_\Delta = \sum_{i=0}^{\Delta-1} P_{i,\Delta}(1)\pi_i + P_{\Delta,\Delta}(1) \left( \frac{1 - \pi_0}{ET} - \sum_{i=1}^{\Delta-1} \pi_i \right). \]

Moreover,

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

Finally, according to (21), we obtain

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

**APPENDIX D**

**PROOF OF PROPOSITION 3**

According to (10), the expected AoII \(\tilde{\Delta}_r\) is given by

\[ \tilde{\Delta}_r = \sum_{i=0}^{\tau-1} C^T(i,0)\pi_i + \sum_{i=\tau}^{\infty} C^T(i,1)\pi_i. \]

In the same spirit as the proof of Proposition 2, we define \(\Sigma \triangleq \sum_{i=\omega}^{\infty} C^T(i,1)\pi_i\) where \(\omega = \tau + t_{\text{max}}\). Then, we have

\[ \tilde{\Delta}_r = \sum_{i=1}^{\tau-1} C^T(i,0)\pi_i + \sum_{i=\tau}^{\omega-1} C^T(i,1)\pi_i + \Sigma. \]
Hence, it is sufficient to obtain the closed-form expression for $\Sigma$. To this end, we recall that

$$
\pi_\Delta = \sum_{i=\Delta-t_{\max}}^{\Delta-1} P_i^{\Delta} (1) \pi_i = \sum_{i=1}^{t_{\max}} P_{i-t_{\max}+\Delta-1,\Delta} (1) \pi_{i-t_{\max}+\Delta-1}, \quad \Delta \geq \omega.
$$

According to Lemma 3, when $\Delta \geq \omega$, $P_{\Delta,\Delta'}^{\Delta} (1) = p_t P_{\Delta,\Delta'}^{\Delta'} (1)$ where $t' = \Delta' - \Delta$. Hence,

$$
\pi_\Delta = \sum_{i=1}^{t_{\max}} p_{t_{\max}+1-i} P_{i-t_{\max}+\Delta-1,\Delta} (1) \pi_{i-t_{\max}+\Delta-1}, \quad \Delta \geq \omega.
$$

Renaming the variables yields

$$
\pi_\Delta = \sum_{t=1}^{t_{\max}} p_t P_{\Delta-t,\Delta} (1) \pi_{\Delta-t}, \quad \Delta \geq \omega.
$$

To proceed, we define, for each $1 \leq t \leq t_{\max}$,

$$
\pi_{\Delta,t} \triangleq p_t P_{\Delta-t,\Delta} (1) \pi_{\Delta-t}, \quad \Delta \geq \omega. \quad (23)
$$

Note that $\sum_{t=1}^{t_{\max}} \pi_{\Delta,t} = \pi_\Delta$. Then, we multiple both side of (23) by $C_{\Delta-t,1}^{\Delta}$ and sum over $\Delta$ from $\omega$ to $\infty$. Hence, we have

$$
\sum_{t=1}^{t_{\max}} C_{\Delta-t,1}^{\Delta} \pi_{t,t} = \sum_{t=1}^{t_{\max}} C_{\Delta-t,1}^{\Delta} \pi_{t-t,1}. \quad (24)
$$

We first investigate the LHS of (24). We notice that $C_{\Delta-t,1}^{\Delta} - C_{\Delta-t,1}^{\Delta-t,1}$ is independent of $\Delta$. Hence, we define $\Delta_t' \triangleq C_{\Delta-t,1}^{\Delta} - C_{\Delta-t,1}^{\Delta-t,1}$. Then, according to (7), we have

$$
\Delta_t' = \sum_{i=1}^{t_{\max}} \left( C_i^{\Delta-t,1} - C_i^{\Delta-t,1} \right).
$$

According to the results in Section III-A we have

$$
C_i^{\Delta-t,1} = \Delta - t + \sum_{h=1}^{i-1} \left( \sum_{k=1}^{h-1} k(1-p^{(h-k)})p(1-p)^{k-1} + (\Delta-t+h)(1-p)^h \right).
$$

Subtracting the two equations yields

$$
C_i^{\Delta-t,1} - C_i^{\Delta-t,1} = t + \sum_{h=1}^{i-1} \left( t(1-p)^h \right) = \frac{t-t(1-p)^i}{p}.
$$

Then, we have

$$
\Delta_t' = \sum_{i=1}^{t_{\max}} \left( \frac{t-t(1-p)^i}{p} \right).
$$
Hence, (24) can be rewritten as
\[
\sum_{i=\omega}^{\infty} (C^T(i, 1) - \Delta') \pi_{i,t} = \sum_{i=\omega-t}^{\infty} C^T(i, 1) p_t P^t_{i,i+t}(1) \pi_i.
\]

Then, we define \( \Pi_t \triangleq \sum_{i=\omega}^{\infty} \pi_{i,t} \) and \( \Sigma_t \triangleq \sum_{i=\omega}^{\infty} C^T(i, 1) \pi_{i,t} \). We notice that \( p_t P^t_{\Delta,\Delta+t}(1) \) is independent of \( \Delta \) if \( \Delta > 0 \). Hence, we obtain
\[
\sum_{i=\omega}^{\infty} C^T(i, 1) \pi_{i,t} - \Delta' \sum_{i=\omega}^{\infty} \pi_{i,t} = p_t P^t_{1,1+t}(1) \sum_{i=\omega-t}^{\infty} C^T(i, 1) \pi_i.
\]

Plugging in the definitions yields
\[
\Sigma_t - \Delta'_t \Pi_t = p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i + \Sigma \right).
\]

Summing the above equation over \( t \) from 1 to \( t_{\text{max}} \) yields
\[
\sum_{t=1}^{t_{\text{max}}} \left( \Sigma_t - \Delta'_t \Pi_t \right) = \sum_{t=1}^{t_{\text{max}}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i + \Sigma \right) \right].
\]

Rearranging the above equation yields
\[
\Sigma - \sum_{t=1}^{t_{\text{max}}} \Delta'_t \Pi_t = \sum_{t=1}^{t_{\text{max}}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i \right) \right] + \sum_{t=1}^{t_{\text{max}}} \left( p_t P^t_{1,1+t}(1) \right) \Sigma. \tag{25}
\]

Then, we calculate \( \Pi_t \). Combining the definition with (23), we have
\[
\Pi_t \triangleq \sum_{i=\omega}^{\infty} \pi_{i,t} = \sum_{i=\omega}^{\infty} \left( p_t P^t_{t-i,1+t}(1) \pi_{t-i} \right) = \sum_{i=\omega-t}^{\infty} \left( p_t P^t_{t-i,t+1}(1) \pi_i \right).
\]

Since \( P^t_{\Delta,\Delta+t}(1) \) is independent of \( \Delta \) when \( \Delta > 0 \), we have
\[
\Pi_t = p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} \pi_i + \Pi \right).
\]

Finally, according to (25), the closed-form expression for \( \Sigma \) is given by the following.
\[
\Sigma = \sum_{t=1}^{t_{\text{max}}} \left[ p_t P^t_{1,1+t}(1) \left( \sum_{i=\omega-t}^{\omega-1} C^T(i, 1) \pi_i \right) + \Pi_t \left( \sum_{i=1}^{t_{\text{max}}} p_i [t - t(1 - p)] \right) \right] + 1 - \sum_{t=1}^{t_{\text{max}}} \left( p_t P^t_{1,1+t}(1) \right).
\]
The proof is based on the results presented in [18]. More precisely, we show that \( M^T \) verifies the two conditions given in [18]. Then, the existence of the optimal policy is guaranteed. Before verifying the conditions, we first introduce the infinite horizon \( \gamma \)-discounted cost of \( M^T \), where \( 0 < \gamma < 1 \) is a discount factor. The expected \( \gamma \)-discounted cost under policy \( \phi \) starting from state \( s \) can be calculated as

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

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

\[
V_{\gamma}^T(s) = \min_{a \in A^T} \left\{ C^T(s) + \gamma \sum_{s' \in S^T} P_{ss'}^T(a)V_{\gamma}^T(s') \right\}.
\]

We also define \( h_{\gamma}^T(s) \triangleq V_{\gamma}^T(s) - V_{\gamma}^T(0) \) as the relative value function and chose the reference state \( 0 = (0, 0, -1) \). According to [18], we only need to verify that the following two conditions for \( M^T \).

- **There exists a non-negative \( N \) such that \( -N \leq h_{\gamma}^T(s) \) for all \( s \) and \( \gamma \):** We first prove one structural property of the value function \( V_{\gamma}^T(s) \) for \( s = (\Delta, t, i) \) where \( \Delta > 0 \).

**Lemma 4 (Monotonicity).** \( V_{\gamma}^T(s) \) is increasing in \( \Delta \) when \( \Delta > 0 \).

**Proof.** Since \( V_{\gamma}^T(s) \) is the value function of state \( s \), value iteration is a canonical procedure to calculate \( V_{\gamma}^T(\cdot) \). Let \( V_{\gamma, \nu}^T(s) \) be the estimated value function at iteration \( \nu \) of value iteration. Then, the estimated value function is updated in the following way.

\[
V_{\gamma, \nu+1}^T(s) = \min_{a \in A^T} \left\{ C^T(s) + \gamma \sum_{s' \in S^T} P_{ss'}^T(a)V_{\gamma, \nu}^T(s') \right\}.
\]

**Lemma 5 (Convergence of value iteration).** The value iteration reported in (27) will converge to value function as iteration goes to infinity. More precisely, \( \lim_{\nu \to \infty} V_{\gamma, \nu}^T(s) = V_{\gamma}^T(s) \) for all \( \gamma \) and \( s \in S^T \).
Proof. According to Propositions 1 and 3 of [18], it is sufficient to show that $V_\gamma^\tau(s)$ is finite for all $s$ and $\gamma$. To this end, we consider the policy $\phi$ being threshold policy with $\tau = \infty$. According to (26), we have

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

Then, by definition, we have $V_\gamma^\tau(s) \leq V^\tau_{\gamma,\phi}(s) < \infty$ for all $s$ and $\gamma$. Then, we can conclude that the value iteration will converge to the value function.

Leveraging Lemma 5, Lemma 4 can be proved by mathematical induction. To start with, we initialize $V^\tau_{\gamma,0}(s) = 0$ for all $s$. Hence, the base case (i.e., $\nu = 0$) is true. We assume the lemma holds at iteration $\nu$. Then, we want to examine whether the lemma holds at iteration $\nu + 1$. We notice that the state transitions corresponding to the state $s = (\Delta, t, i)$ where $\Delta > 0$ share the same structure. Hence, we can easily show that the lemma holds at iteration $\nu + 1$ of value iteration. The details are omitted here. Then, by mathematical induction, we can conclude that the lemma holds.

With Lemma 4 in mind, we can easily conclude that $h_\gamma^\tau(s)$ is also increasing in $\Delta$ when $\Delta > 0$. To proceed, let $c^\tau_{s,s'}(\phi)$ be the expected cost of a first passage from $s \in S^\tau$ to $s' \in S^\tau$ when policy $\phi$ is adopted. We know from Proposition 4 of [18] that $c^\tau_{s,0}(\phi)$ is finite. In the following, we consider the policy $\phi$ being threshold policy with $\tau = 0$, under which the transmitter initiate transmission whenever the channel is idle. Since the policy $\phi$ induces an irreducible ergodic Markov chain and the expected cost is finite, $h_\gamma^\tau(s) \leq c^\tau_{s,0}(\phi)$ by Proposition 5 of [18]. We notice that

$$-h_\gamma^\tau(s) = V_\gamma^\tau(0) - V_\gamma^\tau(s) \leq c^\tau_{s,0}(\phi).$$

Hence, we have $h_\gamma^\tau(s) \geq -c^\tau_{s,0}(\phi)$. Combining with the monotonicity proved in Lemma 4, we can choose $N = \max_{s \in G} \{ c^\tau_{s,0}(\phi) \}$, where $G = \{ s = (\Delta, t, i) : \Delta \in \{0, 1\} \}$.

- $M^\tau$ has a stationary policy $\phi$ inducing an irreducible, ergodic Markov chain. Moreover, the resulting expected AoII is finite: We consider the policy $\phi$ being threshold policy with $\tau = 0$. Clearly, it induces a irreducible, ergodic Markov chain. Moreover, the resulting expected AoII is finite as we can conclude easily from the closed-form expression reported in Proposition 3.
As the two conditions are verified, the existence of the optimal policy is guaranteed by the results presented in [18]. Moreover, the minimum expected AoII is independent of the initial state as also presented in [18].

APPENDIX F
PROOF OF THEOREM 2

The proof is based on the results presented in [19]. To this end, we consider a generic Markov Decision Process \( \mathcal{M} = (S, A, \mathcal{P}, \mathcal{C}) \). Let \( C(s, A) \) be the instant cost for being at state \( s \in S \) under policy \( A \). We also define \( P^A_{s,s'} \) as the probability that applying policy \( A \) at state \( s \) will lead to state \( s' \). Finally, \( V^A(s) \) is defined as the value function resulting from the operation of policy \( A \). Since \( B \) is chosen over \( A \), we have

\[
C(s, B) + \sum_{s' \in S} P^B_{s,s'} V^A(s') \leq C(s, A) + \sum_{s' \in S} P^A_{s,s'} V^A(s'), \quad s \in S.
\]

Then, we define

\[
\gamma_s \triangleq C(s, B) + \sum_{s' \in S} P^B_{s,s'} V^A(s') - C(s, A) - \sum_{s' \in S} P^A_{s,s'} V^A(s') \leq 0, \quad s \in S.
\]

Meanwhile, both policies satisfy their own Bellman equation. More precisely, we have

\[
V^A(s) + \theta^A = C(s, A) + \sum_{s' \in S} P^A_{s,s'} V^A(s'), \quad s \in S.
\]

\[
V^B(s) + \theta^B = C(s, B) + \sum_{s' \in S} P^B_{s,s'} V^B(s'), \quad s \in S.
\]

where \( \theta^A \) and \( \theta^B \) are the expected costs resulting from the operation of policy \( A \) and policy \( B \), respectively. Then, subtracting the two expressions and bringing in the expression for \( \gamma_s \) yield

\[
V^B(s) - V^A(s) + \theta^B - \theta^A = \gamma_s + \sum_{s' \in S} P^B_{s,s'} (V^B(s') - V^A(s')), \quad s \in S.
\]

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

\[
V^\Delta(s) + \theta^\Delta = \gamma_s + \sum_{s' \in S} P^B_{s,s'} V^\Delta(s'), \quad s \in S.
\]

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 \( \theta^\Delta \leq 0 \). Consequently, \( \theta^B \leq \theta^A \).
Then, we prove that the resulting policy is optimal when policy improvement step converges. To this end, we prove by contradiction. We assume there exists two policies $A$ and $B$ such that $\theta^B < \theta^A$. Meanwhile, the policy improvement step has converged to policy $A$. Since the policy has converged, we know $\gamma_s \geq 0$ for all $s \in S$. Hence, $\theta^A \geq 0$. Then, according to the definition of $\theta^\Delta$, we have $\theta^B \geq \theta^A$ which contradicts the assumption. Hence, superior policies cannot go undiscovered. Then, we can conclude that the resulting policy is optimal when policy improvement step converges.

**Appendix G**

**Proof of Theorem 3**

According to Theorem 2, it is sufficient to show that *policy iteration algorithm* converges to threshold policy with $\tau = 1$. To this end, we first present the general procedures we adopted in the optimality proof.

1) **Policy Evaluation**: In this step, we calculate the value function and the expected AoII resulting from the operation of threshold policy with $\tau = 1$.

2) **Policy Improvement**: In this step, we verify that the policy converges under the calculated value function and expected AoII using Bellman equation.

Following the above-mentioned steps, we first calculate the value function and the expected AoII under threshold policy with $\tau = 1$. The value function can be calculated using the Bellman equation with actions specified by threshold policy with $\tau = 1$. Hence, the Bellman equation reduces to a system of linear equations. We denote by $\psi$ the threshold policy with $\tau = 1$. We recall that we only need to determine the action when the channel is idle (i.e., $i = -1$). Hence, we only calculate the value function for the states with $i = -1$. Let $V^\psi(\Delta, 0, -1)$ be the value function of state $(\Delta, 0, -1)$ resulting from the operation of policy $\psi$ and $V^\psi(\Delta)$ being short for $V^\psi(\Delta, 0, -1)$. Then, according to (9), $V^\psi(\Delta)$ satisfies the following system of linear equations.

\[
\begin{align*}
V^\psi(0) &= -\theta^\psi + pV^\psi(1) + (1 - p)V^\psi(0), \\
V^\psi(\Delta) &= C^\psi(\Delta, 1) - ET\theta^\psi + \sum_{t=1}^{t_{\max}} p_t \left( \sum_{k=0}^{t-1} P_{\Delta,k}^t V^\psi(k) + P_{\Delta,\Delta+t}^t V^\psi(\Delta + t) \right), \quad \Delta \geq 1,
\end{align*}
\]  

where $V^\psi(\cdot)$ and $\theta^\psi$ are the value function and the expected AoII resulting from the adoption of $\psi$, respectively. It is very difficult to solve the above system of linear equations directly. However, as we will see later, some structural properties of the value functions are sufficient. The properties are summarized in the following lemma.
Lemma 6 (Structural properties). $V^\psi(\cdot)$ satisfies the following equations.

\[
V^\psi(1) - V^\psi(0) = \frac{\theta^\psi}{p},
\]
\[
V^\psi(\Delta + 1) - V^\psi(\Delta) = \sigma^\psi, \quad \Delta \geq 1,
\]

where

\[
\sigma^\psi = \frac{\sum_{t=1}^{t_{\text{max}}} p_t p (1 - p^r)^t}{1 - \sum_{t=1}^{t_{\text{max}}} p_t (1 - p)^{t-1}}.
\]

Proof. First of all, from (28), we have

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

Then, we will show that $V^\psi(\Delta + 1) - V^\psi(\Delta)$ is constant for all $\Delta \geq 1$. According to [20], the system of linear equations can be solved using iterative policy evaluation algorithm. More precisely, the value function is updated in the following way.

\[
V_{\nu+1}^\psi(0) = -\theta^\psi + p V_{\nu}(1) + (1 - p) V_{\nu}^\psi(0),
\]
\[
V_{\nu+1}^\psi(\Delta) = C^\psi(\Delta, 1) - ET\theta^\psi + \sum_{t=1}^{t_{\text{max}}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta, k}^t V_{\nu}^\psi(k) + P_{\Delta+1, \Delta+t}^t V_{\nu}^\psi(\Delta + t) \right) \right], \quad \Delta \geq 1.
\]

Without loss of generality, we can assume $V_0^\psi(\cdot) = 0$. Since the optimal policy exists according to Theorem [3], iterative policy evaluation algorithm is guaranteed to converge the value function [20]. More precisely, $\lim_{\nu \to \infty} V_\nu^\psi(\cdot) = V^\psi(\cdot)$. Hence, we can prove the desired results by mathematical induction. The base case $\nu = 0$ is true by initialization. Then, we assume $V_\nu^\psi(\Delta + 1) - V_\nu^\psi(\Delta) = \sigma_\nu$ where $\sigma_\nu$ is independent of $\Delta \geq 1$. Then, we will exam whether $V_{\nu+1}^\psi(\Delta + 1) - V_{\nu+1}^\psi(\Delta)$ is independent of $\Delta \geq 1$. We first recall that $P_{\Delta, \Delta+t}^t(1)$ is independent of $\Delta \geq 1$ for any $t$. With this in mind, we have

\[
V_{\nu+1}^\psi(\Delta + 1) - V_{\nu+1}^\psi(\Delta)
= C^\psi(\Delta + 1, 1) - ET\theta^\psi + \sum_{t=1}^{t_{\text{max}}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta+1, k}^t V_{\nu}^\psi(k) + P_{\Delta+1, \Delta+t}^t V_{\nu}^\psi(\Delta + t + 1) \right) \right] -
C^\psi(\Delta, 1) + ET\theta^\psi - \sum_{t=1}^{t_{\text{max}}} \left[ p_t \left( \sum_{k=0}^{t-1} P_{\Delta, k}^t V_{\nu}^\psi(k) + P_{\Delta, \Delta+t}^t V_{\nu}^\psi(\Delta + t) \right) \right]
= C^\psi(\Delta + 1, 1) - C^\psi(\Delta, 1) + \sum_{t=1}^{t_{\text{max}}} \left( p_t P_{\Delta, \Delta+t}^t \sigma_\nu \right).
\]
According to (7), we have
\[
C^T(\Delta + 1, 1) - C^T(\Delta, 1) = \sum_{t=1}^{t_{\text{max}}} (C^t(\Delta + 1, 1) - C^t(\Delta, 1)).
\]
In the case of $\Delta \geq 1$, leveraging Lemma 1, we have
\[
C^t(\Delta + 1, 1) - C^t(\Delta, 1) = 1 + \sum_{k=1}^{t-1} \left( (k + \Delta + 1)(1 - p)^k - (k + \Delta)(1 - p)^k \right)
= 1 + \sum_{k=1}^{t-1} (1 - p)^k = 1 + \frac{(1 - p) - (1 - p)^t}{p},
\]
where $1 \leq t \leq t_{\text{max}}$.
Combining together, we obtain
\[
C^T(\Delta + 1, 1) - C^T(\Delta, 1) = \sum_{t=1}^{t_{\text{max}}} \frac{p_t - p_t(1 - p)^t}{p}.
\]
Hence, we can conclude that $V^\psi_{\nu+1}(\Delta + 1) - V^\psi_{\nu+1}(\Delta)$ is independent of $\Delta$ when $\Delta \geq 1$. Then, by mathematical induction, $V^\psi(\Delta) - V^\psi(\Delta + 1)$ is independent of $\Delta$ when $\Delta \geq 1$. We denote by $\sigma^T$ the constant. Then, $\sigma^T$ satisfies the following.
\[
\sigma^T = V^\psi(\Delta) - V^\psi(\Delta + 1) = \sum_{t=1}^{t_{\text{max}}} \left( \frac{p_t - p_t(1 - p)^t}{p} + p_t(1 - p)^{t-1}\sigma^T \right).
\]
After some algebraic manipulations, we obtain
\[
\sigma^T = \frac{\sum_{t=1}^{t_{\text{max}}} \left(\frac{p_t - p_t(1 - p)^t}{p}\right)}{p - p^2 \sum_{t=1}^{t_{\text{max}}} p_t(1 - p)^{t-1}}.
\]
In the next step, we need to verify that the optimal policy resulting from $V^\psi(\cdot)$ and $\theta^\psi$ is threshold policy with $\tau = 1$. As we discussed before, it is sufficient to determine the action at state $(\Delta, 0, -1)$. To this end, we define $\delta V^\psi(\Delta) \triangleq V^\psi,0(\Delta) - V^\psi,1(\Delta)$ where $V^\psi,a(\Delta)$ is the value function resulting from taking action $a$ at state $(\Delta, 0, -1)$. Then, for $\Delta \geq 1$, we have
\[
\delta V^\psi(\Delta) = \Delta - \theta^\psi + (1 - p)V^\psi(\Delta + 1) + pV(0) - V^\psi,1(\Delta)
= \Delta - \theta^\psi + (1 - p)V^\psi(\Delta + 1) + pV(0) - V^\psi(\Delta)
= \Delta - \theta^\psi + (1 - p)(V^\psi(\Delta + 1) - V^\psi(\Delta)) + p(V^\psi(0) - V^\psi(\Delta))
= \Delta - 2\theta^\psi + [(1 - p) - p(\Delta - 1)]\sigma^T.
\]
We notice that
\[ \delta V^\psi(\Delta + 1) - \delta V^\psi(\Delta) = 1 - p^\sigma. \]

Plugging in the expression for \(\sigma\) yields
\[
1 - p^\sigma = 1 - \frac{\sum_{t=1}^{t_{\text{max}}} (p_t - p_t(1-p)^t)}{1 - \sum_{t=1}^{t_{\text{max}}} p_t(1-p)^{t-1}}
\]
\[
= 1 - \frac{\sum_{t=1}^{t_{\text{max}}} p_t(1-p)^{t-1} - \sum_{t=1}^{t_{\text{max}}} (p_t - p_t(1-p)^t)}{1 - \sum_{t=1}^{t_{\text{max}}} p_t(1-p)^{t-1}}
\]
\[
= \frac{(1-2p) \sum_{t=1}^{t_{\text{max}}} p_t(1-p)^{t-1}}{1 - \sum_{t=1}^{t_{\text{max}}} p_t(1-p)^{t-1}}.
\]

Since \(p \leq \frac{1}{2}\), we know \(1 - p^\sigma \geq 0\). Hence, we have \(\delta V^\psi(\Delta + 1) \geq \delta V^\psi(\Delta)\). Since \(\theta^\psi = \bar{\Delta}_1\) and \(\bar{\Delta}_1 \leq \frac{1+(1-p)p^\tau}{2}\) by Condition \([1]\), we know that
\[ \delta V^\psi(1) = 1 - 2\theta^\psi + (1-p)\sigma^\tau = 1 - 2\bar{\Delta}_1 + (1-p)\sigma^\tau \geq 0. \]

Combining together, we have
\[ \delta V^\psi(\Delta) \geq \delta V^\psi(1) \geq 0, \quad \Delta \geq 1. \]

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 polices. 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 \(\tau = 1\). Combining together, we can conclude that the policy converges, which means that threshold policy with \(\tau = 1\) is optimal.

**APPENDIX H**

**PROOF OF PROPOSITION 4**

We first provide the specific expressions of some quantities that are needed in the case when \(p_t = \frac{1}{2}\) for \(t \in \{1, 2\}\). According to \([7]\) and Lemma \([3]\) we have
\[
C(0, 1) = \sum_{t=1}^{2} p_t C^t(0, 1) = \frac{p}{2},
\]
\[
C(1, 1) = \sum_{t=1}^{2} p_t C^t(1, 1) = 2 - p.
\]
\[
C(2, 1) = \sum_{t=1}^{2} p_t C^t(2, 1) = \frac{7 - 3p}{2}.
\]
\[ P_{\Delta, \Delta'}(1) = \begin{cases} 
 2 \sum_{t=1}^{2} p_t P_{\Delta, 0}^t(1) = \frac{2 - 3p + 2p^2}{2}, & \Delta' = 0, \\
 2 \sum_{t=1}^{2} p_t P_{\Delta, 1}^t(1) = p_2 P_{\Delta, 2}^2(1) = \frac{p - p^2}{2}, & \Delta' = 1, \Delta \geq 1, \\
 2 \sum_{t=1}^{2} p_t P_{\Delta, \Delta + 1}^t(1) = p_1 P_{\Delta, \Delta + 1}^1(1) = \frac{2p - p^2}{2}, & \Delta' = 1, \Delta = 0, \\
 p_1 P_{\Delta, \Delta + 1}^1(1) = \frac{p}{2}, & \Delta' \geq 2, t' = 1, \\
 p_2 P_{\Delta, \Delta + 2}^2(1) = \frac{p - p^2}{2}, & \Delta' \geq 2, t' = 2. 
\]

Moreover, we have
\[ \sigma' = \frac{3 - p}{2 - 2p + p^2}. \]
\[ ET = \sum_{t=1}^{2} t p_t = \frac{3}{2}. \]

Then, we calculate \( \bar{\Delta}_0 \) using Proposition 3. To this end, we first calculate some auxiliary quantities.
\[ \pi_0 = \frac{P_{0,0}(1)}{ET} = \frac{2 - 3p + 2p^2}{3}. \]
\[ \pi_1 = P_{0,1}(1) \pi_0 + P_{1,1}(1) \left( \frac{1}{ET} - \pi_0 \right) = \frac{2p^3 - 5p^2 + 4p}{6}. \]
\[ \pi_2 = \sum_{s=0}^{1} P_{s, \Delta}(1) \pi_s + P_{2,2}(1) \left( \frac{1}{ET} - \sum_{s=0}^{1} \pi_s \right) = \frac{-2p^4 + 5p^3 - 6p^2 + 4p}{12}. \]
\[ \Pi = \frac{\sum_{s=1}^{2} \left( \sum_{k=1}^{s} P_{s, 2+k}(1) \right) \pi_s}{1 - \sum_{s=1}^{2} \left( P_{2,2+s}(1) \right)} = \frac{2p^6 - 13p^5 + 30p^4 - 34p^3 + 16p^2}{12(2 - 2p + p^2)}. \]

Moreover,
\[ \Delta'_1 = \frac{3 - p}{2}. \]
\[ \Delta'_2 = 3 - p. \]
\[ \Pi_1 = \frac{-4p^6 + 10p^5 - 8p^4 - 4p^3 + 8p^2}{24(2 - 2p + p^2)}. \]
\[ \Pi_2 = \frac{8p^6 - 36p^5 + 68p^4 - 64p^3 + 24p^2}{24(2 - 2p + p^2)}. \]
Hence,
\[
\sum_{t=1}^{2} \Delta'_t \Pi_t = \frac{-12p^7 + 98p^6 - 314p^5 + 516p^4 - 452p^3 + 168p^2}{48(2 - 2p + p^2)}.
\]

Then,
\[
\Sigma = \sum_{t=1}^{2} \left[ p_t P^t_{s,s+t}(1) \left( \sum_{s=3-t}^{2} C(s, 1) \pi_s \right) \right] + \sum_{t=1}^{2} s'_t \Pi_t \frac{1 - \sum_{t=1}^{2} \left( p_t P^t_{s,s+t}(1) \right)}{24(2 - 2p + p^2)^2}
\]
\[
= -6p^9 + 61p^8 - 277p^7 + 758p^6 - 1352p^5 + 1556p^4 - 1076p^3 + 344p^2.
\]

Finally, we have
\[
\bar{\Delta}_0 = \sum_{s=0}^{2} C(s, 1) \pi_s + \Sigma = \frac{16p^7 - 96p^6 + 300p^5 - 616p^4 + 832p^3 - 688p^2 + 272p}{24(2 - 2p + p^2)^2}.
\]

In the following, we calculate \(\bar{\Delta}_1\) using Proposition 3. Same as before, we calculate some auxiliary quantities.
\[
\pi_1 = \frac{pP_{1,0}(1) + pP_{1,1}(1)}{pET + P_{1,0}(1)} = \frac{2p - 2p^2 + p^3}{2 + 2p^2}.
\]
\[
\pi_2 = P_{1,2}(1) \pi_1 = \frac{2p^2 - 2p^3 + p^4}{4 + 4p^2}.
\]
\[
\Pi = \frac{4p^2 - 4p^3 + 2p^5 - p^6}{8 - 8p + 12p^2 - 8p^3 + 4p^4}.
\]

Moreover,
\[
\Delta'_1 = \frac{3 - p}{2}.
\]
\[
\Delta'_2 = 3 - p.
\]
\[
\Pi_1 = \frac{8p^3 - 12p^4 + 8p^5 - 2p^6}{2(4 + 4p^2)(2 - 2p + p^2)}.
\]
\[
\Pi_2 = \frac{8p^2 - 16p^3 + 12p^4 - 4p^5}{2(4 + 4p^2)(2 - 2p + p^2)} = \frac{2p^2 - 4p^3 + 3p^4 - p^5}{(2 + 2p^2)(2 - 2p + p^2)}.
\]

Then,
\[
\sum_{t=1}^{2} \Delta'_t \Pi_t = \frac{24p^2 - 44p^3 + 30p^4 - 6p^5 - 3p^6 + p^7}{2(4 + 4p^2)(2 - 2p + p^2)}.
\]
Hence,
\[
\sum = \sum_{t=1}^{2} \left[ p_t P_{s,s+t}^t(1) \left( \sum_{s=3-t}^{2} C(s,1)\pi_s \right) \right] + \sum_{t=1}^{2} s'_t \Pi_t \\
1 - \sum_{t=1}^{2} \left( p_t P_{s,s+t}^t(1) \right)
\]
\[
= \frac{3p^9 - 21p^8 + 64p^7 - 102p^6 + 56p^5 + 72p^4 - 144p^3 + 80p^2}{2(4 + 4p^2)(2 - 2p + p^2)^2}
\]

Finally, we have
\[
\bar{\Delta}_1 = \sum_{s=1}^{2} C(s,1)\pi_s + \Sigma = \frac{p(-p^4 - p^3 + 12p^2 - 22p + 16)}{2(1 + p^2)(2 - 2p + p^2)^2}
\]

Using the above expressions, we can verify Condition 1. It is equivalent to verify the following two inequalities.
\[
\begin{cases}
\bar{\Delta}_1 \leq \frac{1 + (1 - p)\sigma^\gamma}{2}, \\
\bar{\Delta}_1 \leq \bar{\Delta}_0.
\end{cases}
\]

In the following, we will verify the inequalities one by one.

- Here, we verify the inequality \( \bar{\Delta}_1 \leq \frac{1 + (1 - p)\sigma^\gamma}{2} \). To this end, we have
  \[
  \frac{1 + (1 - p)\sigma^\gamma}{2} = \frac{5 - 6p + 2p^2}{2(2 - 2p + p^2)}.
  \]
  Then,
  \[
  \bar{\Delta}_1 - \frac{1 + (1 - p)\sigma^\gamma}{2} = \frac{(1 - 2p)(p^5 - 4p^4 + 10p^3 - 17p^2 + 18p - 10)}{2(1 + p^2)(2 - 2p + p^2)^2}.
  \]

- Here, we verify the condition \( \bar{\Delta}_1 \leq \bar{\Delta}_0 \). To this end, we have
  \[
  \bar{\Delta}_1 - \bar{\Delta}_0 = \frac{4p(1 - 2p)(2p^7 - 11p^6 + 34p^5 - 72p^4 + 107p^3 - 108p^2 + 66p - 20)}{24(1 + p^2)(2 - 2p + p^2)^2}.
  \]

We can conclude that both inequalities hold when \( p \in (0, \frac{1}{2}] \). Hence, by Theorem 3, the optimal policy in this case is threshold policy with \( \tau = 1 \).
Algorithm 1 Policy Iteration Algorithm

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

1: procedure POLICY ITERATION ALGORITHM($\mathcal{M}$)
2: Initialize $\psi(s) \in \mathcal{A}$ for all $s \in \mathcal{S}$.
3: $(V^\psi(s), \theta^\psi) \leftarrow \text{PolicyEvaluateStep}(\mathcal{M}, \psi(s))$.
4: $\psi(s) \leftarrow \text{PolicyImproveStep}(\mathcal{M}, V^\psi(s))$.
5: if $\psi(s)$ does not converge then
6: go to line 3
return $(\psi(s), \theta^\psi)$.