Task-oriented Age of Information for Wireless Monitoring Systems

***, ***, ***, ***, and ***

***Abstract*—The emergence of new intelligent applications has fostered the development of a task-oriented communication paradigm, where a comprehensive, universal, and practical metric is crucial for unleashing the potential of this paradigm. To this end, we introduce an innovative metric, the Task-oriented Age of Information (TAoI), to measure whether the content of information is relevant to the system task, thereby assisting the system in efficiently completing designated tasks. Also, we study the TAoI in a wireless monitoring system, whose monitoring task is to identify targets and transmit their images for subsequent analysis. We formulate the dynamic transmission problem as a Semi-Markov Decision Process (SMDP) and transform it into an equivalent Markov Decision Process (MDP) to minimize TAoI and find the optimal transmission policy. Furthermore, we demonstrate that the optimal strategy is a threshold-based policy regarding TAoI and propose a relative value iteration algorithm based on the threshold structure to obtain the optimal transmission policy. Finally, simulation results prove the superior performance of the optimal transmission policy compared to the two baseline policies.**

***Index Terms*—task-oriented communication, Age of information, semi-Markov decision process (SMDP).**

I. Introduction

Generally, conventional communications rely on Shannon's channel coding theory to achieve reliable transmission from sources to destinations [1]. The core idea is to abstract information into bits and design source coding/decoding, channel coding/decoding, modulation, and demodulation parts to minimize signal distortion measures (e.g., Mean Square Error) and achieve error-free replication of bits from source to destination [2]. This approach has been tremendously successful in systems where communicating, e.g., voice and data [3]. However, with the emergence of new intelligent systems such as real-time cyber-physical systems, interactive systems, and autonomous multi-agent systems, the communication goals of these physical systems are no longer to reconstruct the underlying message but to enable the destination to make the right inference or to take the right decision at the right time and within the right context [4]. This poses new technical challenges to conventional communication methods. Therefore, a new communication paradigm, task-oriented or goal-oriented communication, has been proposed as a promising solution [4]–[6].

The key to unlocking the potential of task-oriented communication lies in a comprehensive, universal, and practical metric to measure the importance and relevance of information to system tasks, thereby significantly reducing computational and transmission costs by only acquiring, transmitting, and reconstructing task-relevant information. Therefore, there have been several studies on metrics for task-oriented communication in recent years [7]–[9]. In [7], the Age of Information (AoI) is proposed as a pioneering metric to capture the freshness of data perceived by the destination, to measure the importance and relevance of information for physical systems. However, AoI has two significant limitations: first, AoI cannot measure the content of information and its dynamic changes; second, AoI considers only the process of information from its generation to its reception, not capturing the impact of received information on the system (i.e., the changes in the system tasks or decisions driven by the received information), thereby failing to provide a closed-loop metric.

To address these issues, the authors in [8] proposed the age of changed information (AoCI), which considers the impact of changes in information content on physical systems. Specifically, AoCI considers changes in semantics to be more beneficial to the system. However, AoCI does not directly measure whether the information content is relevant to the task. In [9], the authors proposed the Age of Incorrect Information (AoII), which combines a time penalty function and an estimation error penalty function to reflect the difference between the receiver's estimate and the actual state of the physical system. Although AoII measures the information content and dynamic changes of information based on AoI, the goal of adopting AoII is still to restore the information received by the source as perfectly as possible at the destination, which goes against the original intention of task-oriented communication. Additionally, calculating AoII at the destination requires the destination to know the current actual state of the physical system, which is too idealistic for most physical systems. The aforementioned works only focus on the process from information generation to reception, while [10], [11] consider the timeliness of closed-loop perception-driven aspects that they have overlooked. Among these, [10] introduced Age of Loop (AoL), which extends the Up/Down-Link AoI to a closed-loop AoI metric. [11] proposed Age of Actuation (AoA), which captures the elapsed time since the last performed actuation at a destination based on data received by a source. Although [10], [11] consider the system's closed loop, they still measure information from a temporal dimension, neglecting the content of the information.

Part of this work was presented at the IEEE/CIC ICCC, Aug. 2020 [**?**].

X. Wang, S. Gan, and X. Chen are with School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, 510006, China (e-mail: wangxijun@mail.sysu.edu.cn; chenxiang@mail.sysu.edu.cn).

Y. Huang is with Guangdong Communications and Networks Institute, Guangzhou, China. This work was done when he was with School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, China (e-mail: huangyz@gdcni.cn).

Y. Xu is with the Department of Mechanical and Automation Engineering, the Chinese University of Hong Kong, Hong Kong SAR, China, and also with the Chinese University of Hong Kong Shenzhen Research Institute,

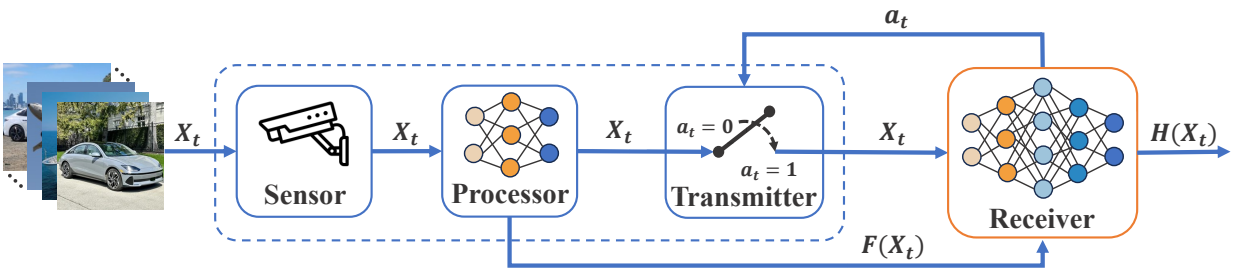


Fig. 1: An illustration of the task-oriented monitoring system.

In this paper, we consider a wireless monitoring system with a monitoring task consisting of a sensor, a processor, a transmitter, and a receiver. Specifically, the sensor captures real-time images, which are pre-classified by processors, and then the receiver decides whether to require the transmitter to transmit the images based on the pre-identification results and monitoring target. Then, a new task-oriented communication metric called Task-oriented Age of Information (TAoI) is introduced, which directly characterizes whether the information content is relevant to the system task. If the transmitted image matches the target, TAoI decreases; otherwise, it increases. We focus on finding the optimal transmission policy for the wireless monitoring system to minimize TAoI. By modeling the problem as an infinite time-horizon Semi-Markov Decision Process (SMDP) and transforming it into an equivalent MDP with uniform time steps, we prove that the optimal transmission policy is a threshold-type policy. Furthermore, we propose a relative value iteration algorithm based on the threshold structure to obtain the optimal transmission policy with low complexity. Finally, simulation results demonstrate that the optimal transmission policy outperforms the two baseline policies.

The rest of this paper is organized as follows. Section II presents the system model and introduces the proposed metric. In Section III, we provide the SMDP formulation of the problem, analyze the threshold structure of the optimal policy, and propose the relative value iteration algorithm based on the threshold structure. Simulation results are presented in Section IV, followed by the conclusion in Section V.

II. System Overview

A. System Model

As shown in Fig.1, we consider a task-oriented monitoring system consisting of a sensor, a processor, a transmitter, and a receiver. The monitoring task of this system is to capture the target image and send it to the receiver for subsequent analysis (e.g., highway traffic analysis and prediction). The sensor captures real-time images, which are then processed by the lightweight binary classifier in the processor. The pre-classification result is sent to the receiver to aid in recognizing the target image. The transmitter receives transmission decisions from the receiver and controls the image transmission through the channel. The receiver is equipped with a large binary classifier. It utilizes the pre-classification result from the processor along with the target to determine whether to instruct the transmitter to send the image. Upon receiving the image, the large binary classifier evaluates whether it matches the target.

A time-slotted system is considered as shown in Fig.2, where the duration of each time slot is τ (in seconds) and a decision epoch of the receiver as a time step. At the beginning of time step t, the sensor captures a fresh image $X_t \in \mathcal{X}$, whose label is defined by $Y_t \in \{0, 1\}$. To characterize the dynamics of the image generation, we assume that the image generation follows a Bernoulli process. Accordingly, we define the probability that the label Y_t of image X_t at time step t is 1 as $\Pr(Y_t = 1) = q$, and the probability that the label Y_t of image X_t at time step t is 0 as $\Pr(Y_t = 0) = 1 - q$. The processor performs binary classification on the image X_t and sends the obtained pre-classification result $F(X_t)$ to the receiver, where $F(X_t) \in \{0, 1\}$. Note that the processor may provide incorrect information. Let p_A and p_B respectively denote the misclassification probabilities of image X_t for labels $Y_t = 0$ and $Y_t = 1$, i.e.,

$$p_A \triangleq \Pr(F(X_t) = 1 | Y_t = 0), \forall t \tag{1}$$

$$p_B \triangleq \Pr(F(X_t) = 0 | Y_t = 1), \forall t. \tag{2}$$

We define the label of the target as G and assume that the label for the target in the monitoring task is 1, i.e, $G = 1$. Based on the pre-classification result $F(X_t)$ and the target G, the receiver must determine whether to request the transmitter to transmit the image X_t. Let $a_t \in \{0, 1\}$ denote the transmission decision of the receiver at time step t, where $a_t = 1$ indicates that the transmitter transmits the image X_t to the receiver, and $a_t = 0$, otherwise.

Assume that each image has the same size and the transmitter transmits an image over a reliable channel at a constant rate. As such, we specify that each image from the sensor to the receiver takes T_u time slots, where operations other than transmitting images only take 1 time slot. Note that the duration of a time step is not uniform. Specifically, let $L(a_t)$ denote the number of time slots in time step t with action a_t being taken, $L(a_t)$ can be expressed as

$$L(a_t) = \begin{cases} 1, & \text{if } a_t = 0 \\ T_u, & \text{if } a_t = 1 \end{cases} \tag{3}$$

When the image X_t arrives at the receiver, the large binary classifier in the receiver classifies it. It is assumed the classifier to be perfectly accurate, meaning the classification result $H(X_t)$ is identical to the image label Y_t, i.e., $H(X_t) = Y_t$.

We denote by $d_t \in \{0,1\}$ an indicator for whether the monitoring task is successful at time step t. If $d_t = 1$, then the classification result $H(X_t)$ matches the target G (i.e, $H(X_t) = G$) . Otherwise, it indicates they are different (i.e, $H(X_t) \neq G$). In particular, the probabilities for success and failure of the monitoring task are as follows:

$$\begin{aligned}\Pr(d_t = 1) =& \Pr(H(X_t) = G) = \Pr(Y_t = G) \\ =& (1-\hat{p}_A)\Pr(F(X_t) = 1) + \hat{p}_B \Pr(F(X_t) = 0).\end{aligned} \tag{4}$$

$$\begin{aligned}\Pr(d_t = 0) =& \Pr(H(X_t) \neq G) = \Pr(Y_t \neq G) \\ =& \hat{p}_A \Pr(F(X_t) = 1) + (1-\hat{p}_B)\Pr(F(X_t) = 0).\end{aligned} \tag{5}$$

where

$$\hat{p}_A \triangleq \Pr(Y_t = 0 | F(X_t) = 1) = \frac{(1-q)p_A}{(1-q)p_A + q(1-p_B)}, \tag{6}$$

$$1 - \hat{p}_A \triangleq \Pr(Y_t = 1 | F(X_t) = 1) = \frac{q(1-p_B)}{(1-q)p_A + q(1-p_B)}, \tag{7}$$

$$\hat{p}_B \triangleq \Pr(Y_t = 1 | F(X_t) = 0) = \frac{qp_B}{(1-q)(1-p_A) + qp_B}, \tag{8}$$

$$1 - \hat{p}_B \triangleq \Pr(Y_t = 0 | F(X_t) = 0) = \frac{(1-q)(1-p_A)}{(1-q)(1-p_A) + qp_B}. \tag{9}$$

B. Task-oriented AoI

In many real-time physical systems, AoI is extensively used to quantify the freshness of data perceived by the receiver, thereby enhancing the utility of decision-making processes [12]. These efforts are driven by the consensus that freshly received data typically contains more valuable information. However, AoI does not provide a direct measure of the data's content and ignores the dynamic impact of the content of the source data on the system. The metric we proposed, Task-oriented Age of Information (TAoI), differs from AoI in that TAoI not only captures the time lag of information received at the destination but also considers whether the content of this information is relevant to the goals of the system task.

For our system, the TAoI decreases only when the system captures the target image and sends it to the receiver (i.e., the monitoring task is successful); otherwise, it increases. Formally, let U_t denote the time step at which the most up-to-date the receiver receives the target image. Then, the TAoI at the ith time slot of time step t can be defined as

$$\Delta_{t,i} = \sum_{n=U_t}^{t-1} L(a_n) + i - 1 \tag{10}$$

where the first term is the number of time slots in the previous time steps since U_t and the second term is the number of time slots in the current time step. For ease of exposition, we represent the TAoI at the beginning of time step t as Δ_t. That is, $\Delta_t = \Delta_{t,1} = \sum_{n=U_t}^{t-1} L(a_n)$. When the transmitter sends an image to the receiver, and this image is the target image (i.e., $a_t = 1$ and $d_t = 1$), we specify that TAoI becomes the duration T_u taken to process this image. If the transmitter sends an image that is not the target image (i.e., $a_t = 1$ and $d_t = 0$), then the TAoI increases by T_u. When the receiver decides not to request the transmitter to send the image (i.e., $a_t = 0$), the TAoI increases by one. Thus, the dynamics of TAoI can be shown as follows:

$$\Delta_{t+1} = \begin{cases} T_u, & \text{if } a_t = 1 \text{ and } d_t = 1 \\ \Delta_t + T_u, & \text{if } a_t = 1 \text{ and } d_t = 0 \\ \Delta_t + 1, & \text{if } a_t = 0. \end{cases} \tag{11}$$

In addition, we illustrate an example of the AoI evolution with $T_u = 4$ in Fig.3.

C. Optimization Problem

In this paper, we aim at finding an transmission policy $\pi = (a_1, a_2, \cdots)$ that minimizes the long-term average TAoI, which means maximizing the number of successful monitoring tasks. Therefore, the dynamic transmission problem can be formulated as follows:

$$\min_{\pi} \limsup_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=1}^{T} \Delta_t\right]}{\mathbb{E}\left[\sum_{t=1}^{T} L(a_t)\right]}. \tag{12}$$

III. SMDP Formulation and Solution

A. SMDP Formulation

SMDP extends MDP to deal with situations in which the time interval between decision instants are not constant, as occurs with the dynamic transmission problem considered here. To this end, we formulate the dynamic transmission problem as an infinite time-horizon SMDP problem, which consists of a tuple $(\mathcal{S}, \mathcal{A}, t^+, \Pr(\cdot,\cdot), R(\cdot,\cdot))$ and depicted as follows:

1) State space $\mathcal{S}$: The state $\mathbf{s}_t$ of the SMDP at time step t is defined as $\mathbf{s}_t \triangleq (\Delta_t, F(X_t))$, where Δ_t denotes the TAoI at the beginning of time step t and $F(X_t)$ the pre-classification result for the image X_t at time step t. Denote the space of all possible state by $\mathcal{S}$ which countably infinite.

2) Action space $\mathcal{A}$: The action at time step t is the transmission decision a_t and the action space is $\mathcal{A} \triangleq \{0,1\}$.

3) Decision epoch t^+: As shown in Fig.2, a decision is made in the third step of the first time slot at each time step. The time interval $L(a_t)$ between two adjacent decisions depends on the action a_t taken at time step t, as specified in (3).

4) Transition probability $\Pr(\cdot,\cdot)$: Given current state $\mathbf{s}_t = (\Delta_t, F(X_t))$ and action a_t, the transition probability to next state $\mathbf{s}_{t+1} = (\Delta_{t+1}, F(X_{t+1}))$ is denoted by $\Pr(\mathbf{s}_{t+1}|\mathbf{s}_t, a_t)$. According to the TAoI evolution dynamic in (11), the transition probability can be given as in Table I. Particularly, we denote g as the probability that the pre-classification result is 1, i.e., $g \triangleq \Pr(F(X_t) = 1)$[1].

5) Reward function $R(\cdot,\cdot)$: Let Δ_t be the instantaneous reward under state $\mathbf{s}_t$ given action a_t, i.e.,

$$R(\mathbf{s}_t, a_t) = R((\Delta_t, F(X_t)), a_t)$$

[1] $g \triangleq \Pr(F(X_t) = 1) = \Pr(F(X_t) = 1 | Y_t = 0) + \Pr(F(X_t) = 1 | Y_t = 1) = p_A + (1 - p_B)$

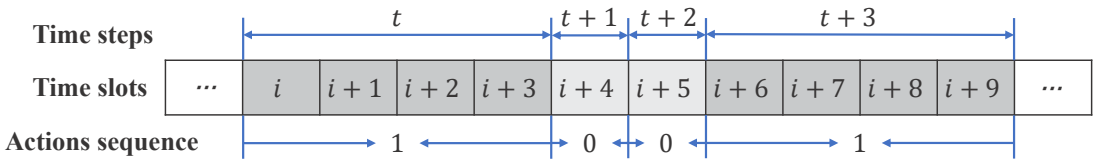


Fig. 2: An illustration of sequence of operations.

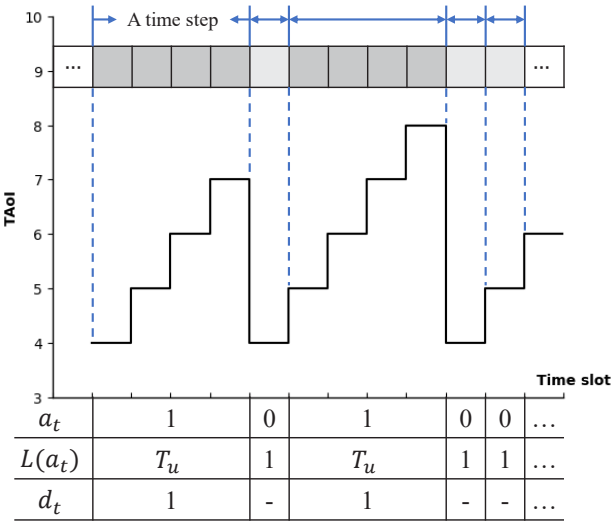


Fig. 3: An illustration of the evolution of the AoI, where $T_u = 4$.

TABLE I: Transition probability

$\Pr(\mathbf{s}_{t+1}\|\mathbf{s}_t, a_t)$	$\mathbf{s}_t$	a_t	$\mathbf{s}_{t+1}$
$(1-\hat{p}_A)g$	$(\Delta_t, F(X_t)=1)$	1	$(T_u, F(X_{t+1})=1)$
$\hat{p}_B g$	$(\Delta_t, F(X_t)=0)$	1	$(T_u, F(X_{t+1})=1)$
$(1-\hat{p}_A)(1-g)$	$(\Delta_t, F(X_t)=1)$	1	$(T_u, F(X_{t+1})=0)$
$\hat{p}_B(1-g)$	$(\Delta_t, F(X_t)=0)$	1	$(T_u, F(X_{t+1})=0)$
$\hat{p}_A g$	$(\Delta_t, F(X_t)=1)$	1	$(\Delta_t+T_u, F(X_{t+1})=1)$
$(1-\hat{p}_B)g$	$(\Delta_t, F(X_t)=0)$	1	$(\Delta_t+T_u, F(X_{t+1})=1)$
$\hat{p}_A(1-g)$	$(\Delta_t, F(X_t)=1)$	1	$(\Delta_t+T_u, F(X_{t+1})=0)$
$(1-\hat{p}_B)(1-g)$	$(\Delta_t, F(X_t)=0)$	1	$(\Delta_t+T_u, F(X_{t+1})=0)$
g	$(\Delta_t, F(X_t)=1)$	0	$(\Delta_t+1, F(X_{t+1})=1)$
g	$(\Delta_t, F(X_t)=0)$	0	$(\Delta_t+1, F(X_{t+1})=1)$
$1-g$	$(\Delta_t, F(X_t)=1)$	0	$(\Delta_t+1, F(X_{t+1})=0)$
$1-g$	$(\Delta_t, F(X_t)=0)$	0	$(\Delta_t+1, F(X_{t+1})=0)$

$$
\begin{aligned}
&= \sum_{i=1}^{L(a_t)} \Delta_{t,i} \\
&= \sum_{i=1}^{L(a_t)} \Delta_t + i - 1 \\
&= L(a_t)\left(\Delta_t + \frac{1}{2}\left(L(a_t)-1\right)\right) \qquad (13)
\end{aligned}
$$

Given an initial system state $\mathbf{s}_1$, the dynamic transmission problem (12) can be expressed as:

$$
\min_{\pi} \limsup_{T\to\infty} \frac{\mathbb{E}\left[\sum_{t=1}^{T} R(\mathbf{s}_t, a_t) \mid \mathbf{s}_1\right]}{\mathbb{E}\left[\sum_{t=1}^{T} L(a_t)\right]} \qquad (14)
$$

Since the non-uniform duration of time steps, the average reward in (14) is defined as the limit of the expected total reward over a finite number of time steps divided by the expected cumulative time of these time steps.

In this paper, our objective is to find a stationary deterministic optimal transmission policy that solves the long-term average TAoI minimization problem as (14). Since the infinite-time-average MDP/SMDP we are optimizing has a countably infinite state space and unbounded TAoI, there may not exist a stationary deterministic policy with an optimal average TAoI [13]. Therefore, before analyzing the stationary deterministic optimal policy for average TAoI, we need to prove the existence of such a policy.

To this end, we first use uniformization to transform the SMDP into an equivalent discrete-time MDP [14], [15]. Denoting the state and action spaces of the transformed MDP as $\hat{\mathcal{S}}$ and $\hat{\mathcal{A}}$ respectively, they remain identical to those in the original SMDP, i.e., $\hat{\mathcal{S}} = \mathcal{S}$ and $\hat{\mathcal{A}} = \mathcal{A}$. For any $\mathbf{s} = (\Delta, F(X)) \in \hat{\mathcal{S}}$ and $a \in \hat{\mathcal{A}}$, the reward in the MDP can be given by

$$
\bar{R}((\Delta, F(X)), a) = \Delta + \frac{1}{2}(L(a)-1) \qquad (15)
$$

and the transition probability is given by

$$
\bar{p}(\mathbf{s}_\prime|\mathbf{s}, a) = \begin{cases} \frac{\epsilon}{L(a)} p(\mathbf{s}'|\mathbf{s}, a), & \mathbf{s}' \neq \mathbf{s} \\ 1 - \frac{\epsilon}{L(a)}, & \mathbf{s}' = \mathbf{s} \end{cases} \qquad (16)
$$

where ϵ is chosen in $(0, \min_a L(a)]$. Next, the average TAoI under policy π is given by

$$
V_\pi(\mathbf{s}) = \frac{1}{T} \limsup_{T\to\infty} \mathbb{E}\left[\sum_{t=1}^{T} \bar{R}(\mathbf{s}_t, a_t) \mid \mathbf{s}_1\right] \qquad (17)
$$

For an infinite horizon, we focus on the set of deterministic stationary policies Π, where $\pi = \{a_1, a_2, \cdots\} \in \Pi$ such that $a_{t1} = a_{t2}$ when $\mathbf{s}_{t1} = \mathbf{s}_{t1}$ for any $t1, t2$. Thus, we omit the time index in the sequel. The objective is to find a policy $\pi \in \Pi$ that minimizes the average TAoI. In particular, if there is a policy that can minimize (17), i.e.,

$$
\min_{\pi\in\Pi} V_\pi(\mathbf{s}), \qquad (18)
$$

We refer to this policy as the average TAoI optimal policy and denote it as π^*. Based on [13, Theorem 4.2], we prove that there exists a deterministic optimal policy for (18). The proof

is provided in the Appendix D, under the proof of Theorem 1.

Then, we can obtain the optimal policy π^* for the original SMDP that minimizes the average TAoI by solving the Bellman equation in (19). According to [16], we have

$$V^* + V(\mathbf{s}) = \min_{a\in\mathcal{A}} \left\{ \bar{R}(\mathbf{s},a) + \sum_{\mathbf{s}'\in\mathcal{S}} \bar{p}(\mathbf{s}'|\mathbf{s},a) V_k(\mathbf{s}') \right\}, \ \forall \mathbf{s} \in \mathcal{S} \tag{19}$$

where V^* represents the optimal value to (14) for all initial states, and $V(\mathbf{s})$ is the value function for the discrete-time MDP. Moreover, the optimal policy π^* for any $\mathbf{s} \in \mathcal{S}$ can be given by

$$\pi^*(\mathbf{s}) = \arg\min_{a\in\mathcal{A}} \left\{ \bar{R}(\mathbf{s},a) + \sum_{\mathbf{s}'\in\mathcal{S}} \bar{p}(\mathbf{s}'|\mathbf{s},a) V_k(\mathbf{s}') \right\}, \ \forall \mathbf{s} \in \mathcal{S}. \tag{20}$$

B. Structural Analysis and Optimal Policy

In this section, we investigate the structure of the optimal policy is of threshold with respect to the TAoI. Then, we use the threshold structure to propose an optimal transmission policy for the dynamic transmission problem.

To solve (20), we apply the value iteration algorithm (VIA) to analyze the structure of the optimal policy. To begin with, we present some key properties of the value function $V(\mathbf{s})$ (i.e., $V(\Delta, F(X))$) in the following lemmas.

Lemma 1. *The value function $V(\Delta, F(X))$ is non-decreasing with Δ for any given $F(X)$.*

Proof: See Appendix A. □

Lemma 2. *Given $F(X)$, the value function $V(\Delta, F(X))$ is concave in Δ.*

Proof: See Appendix B. □

Since the value function $V(\Delta, F(X))$ is concave, its slope does not increase monotonically. The lower bound of the slope of $V(\Delta, F(X))$ is given by the following lemma.

Lemma 3. *For any $\mathbf{s}_1 = (\Delta_1, F(X)), \mathbf{s}_2 = (\Delta_2, F(X)) \in \mathcal{S}$, such that $\Delta_1 \leq \Delta_2$, $V_k(\Delta_2, F(X)) - V_k(\Delta_1, F(X)) \geq \frac{T_u}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1)$, where $p_1 = \hat{p}_A$ if $F(X) = 1$ and $p_1 = 1 - \hat{p}_B$ if $F(X) = 0$.*

Proof: See Appendix C. □

Based on Lemma 1-3, we can derive the structure of the optimal transmission policy as stated in the following theorem.

Theorem 4. *Given $F(X)$, there exists a stationary deterministic optimal policy that is of threshold-type in Δ. Specifically, if $\Delta \geq \Omega$, the $\pi^* = 1$, where Ω denotes the threshold given pair of Δ and $F(X)$.*

Proof: See Appendix D. □

According to Theorem 4, the optimal policy can be represented as a threshold policy on Δ, which is given by

$$\pi^*(\mathbf{s}) = \begin{cases} 1, & \text{if } \Delta \geq \Omega \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

Algorithm 1 Optimal Transmission Policy Based on the Threshold Structure

1: **Initialization:** Initialize $\mathbf{s}_0 = (0,0)$, $\pi_0^* = 0$ and T.
2: **for** $t = 1, T$ **do**
3: **Computing the Optimal Threshold:** According to (23) and $F(X_t)$, computing the optimal threshold Ω^*.
4: **Action Selection:** If $\Delta_t \geq \Omega^*$, $\pi_t^*(\mathbf{s}_t) = 1$. Otherwise, $\pi_t^*(\mathbf{s}_t) = 0$.
5: **Acting and Observing:** Execute the optimal policy $\pi_t^*(\mathbf{s}_t)$, receive reward $R(\mathbf{s}_t, a_t)$, and obtain the new state $\mathbf{s}_{t+1}$.
6: **end for**
7: **Output:** The optimal transmission policy π^*.

where Ω is the threshold that triggers the switch. With the threshold policy, we proceed to analyze the total average reward for any threshold Ω in the asymptotic regime, as detailed in the following lemma.

Lemma 5. *When $\hat{\Delta}$ goes to infinity, for any given $F(X)$ and threshold Ω, the total average reward $J(\Omega, F(X))$ of the threshold policy can be given by*

$$J(\cdot) = \frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}} \left(\frac{\Omega^2 - \Omega}{2} + \frac{2\Omega + T_u - 1}{2 - 2\hat{p}} + \frac{\hat{p}}{(1-\hat{p})^2} \right). \tag{22}$$

where $\hat{p} \triangleq \epsilon(1-p_1)/T_u$.

Proof: See Appendix E. □

Next, based on Theorem 1 and Lemma 4, we can derive the closed-form expression of the optimal threshold Ω^* of the optimal transmission policy. The optimal transmission policy based on the optimal threshold is detailed in Algorithm 1.

Theorem 6. *The optimal threshold Ω^* of the optimal transmission policy is given by*

$$\Omega^* = \arg\min(J(\lfloor \Omega' \rfloor), J(\lceil \Omega' \rceil)), \tag{23}$$

where $\Omega' = \dfrac{\sqrt{\hat{p} + (T_u - 1)(1-\hat{p})} - \hat{p}}{1-\hat{p}}$.

Proof: See Appendix F. □

Remark 7. The optimal threshold depends on two variables,$\hat{p}$ and T_u, as illustrated in Theorem 2. Firstly, when $T_u = 1$, indicating that the increase in TAoI from transmitting a non-target image is equivalent to not transmitting, the optimal threshold Ω^* is 0 for any $\hat{p}$ (where$0 \hat{\leq} p < 1$). This implies that once transmitting an image does not incur additional TAoI, the optimal policy is to continuously transmit regardless of the received image, aiming to reduce the average TAoI. Furthermore, the variable $\hat{p}$ depends on p_1, where p_1 is determined by $\hat{p_A}$ and $\hat{p_B}$. Specifically, if $F(X) = 1$, $p_1 = \hat{p_A}$; if $F(X) = 0$, $p_1 = 1 - \hat{p_B}$. When $\hat{p} = 0$, for any T_u, the optimal threshold is also 0, meaning the optimal policy is to transmit continuously. In fact, $\hat{p} = 0$ implies $\hat{p_A} = 0$ or $\hat{p_B} = 1$, indicating that all captured images are target images, thus validating the correctness of the optimal threshold Ω^*.

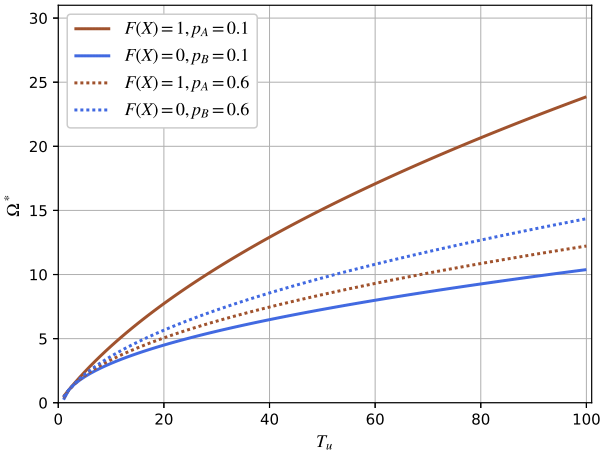


Fig. 4: The optimal threshold Ω^* versus T_u ($q = 0.5$).

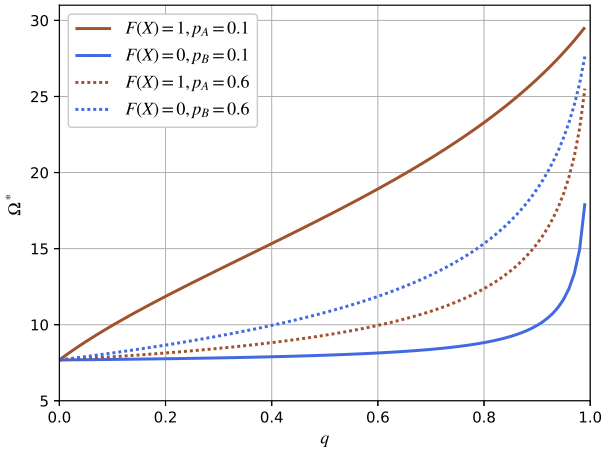


Fig. 5: The optimal threshold Ω^* versus q ($T_u = 60$).

Fig. 4 shows the optimal threshold Ω^* of the optimal transmission policy with respect to T_u under different $F(X)$, p_A and p_B. It can be observed that the optimal threshold increases with T_u. This is because a higher image transmission delay implies a larger TAoI, regardless of the success or failure of the monitoring task. Therefore, the optimal threshold naturally increases. We also observe that when $F(X) = 1$, the corresponding misclassification probability p_A decreases as the optimal threshold increases. That is, early transmission of a non-target image leads to a larger TAoI. Conversely, when $F(X) = 0$, the optimal threshold is positively correlated with its misclassification probability p_B. This means that the higher the misclassification probability , the later the target image is transmitted. This indicates that the more accurate the classifier in the processor, the more beneficial it is for the system to make better transmission decisions.

Fig. 5 illustrates the optimal threshold Ω^* of the optimal transmission policy with respect to q under different $F(X)$, p_A and p_B. We can see that the optimal threshold is positively correlated with q. This is due to the fact that a larger q represents a higher probability of target arrival, leading to a higher cost of failure for the monitoring task, which naturally increases the threshold. Additionally, it can be observed that when the misclassification probability is low (i.e., $p_A = 0.1$, and $p_B = 0.1$), the optimal threshold for preclassification as $F(X) = 0$, is lower than $F(X) = 1$, regardless of the value of q. This is because when the misclassification probability is low, for $F(X) = 0$, Y is likely to be 1. Thus, transmission will occur earlier. Similarly, for $F(X) = 1$, transmission will occur later.

IV. Simulation Results

In this section, numerical simulations are shown to evaluate the optimal policy for the dynamic transmission problem.

A. Simulation Setup

We evaluate the proposed policy on binary classification using the CIFAR-10 dataset. The CIFAR-10 dataset consists of $60,000$ 32×32 RGB images, including $50,000$ training images and $10,000$ test images. The original labels range from $\{0, \cdots, 9\}$, corresponding to the objects depicted in each image. We use binary labels $\{0, 1\}$ to indicate whether the image depicts an animal or a vehicle. We assume the target of the monitoring task in the system is vehicles, labeled as 1. During the training phase, we can choose a network from Table II as the processor to adjust the misclassification probabilities, p_A and p_B. Note that the equal classification accuracy for animals and vehicles in Table II is due to the relatively uniform distribution of data in the CIFAR-10 dataset, which may not hold true in practice. After obtaining the trained classifier, the system enters the inference phase, i.e., the transmission control phase. At the beginning of each decision epoch (i.e., time step) in the inference phase, the sensor selects an image from the CIFAR-10 dataset with a Bernoulli distribution probability of q.

TABLE II: Compare monitoring accuracy on CIFAR-10 dataset

Network	Test Accuracy of Vehicles $1 - \hat{p}_A$	Test Accuracy of Animals $1 - \hat{p}_B$
LeNet [17]	61.63%	59.76%
AlexNet [18]	78.76%	75.67%
VGG-16 [19]	84.53%	82.60%
ResNet-18 [20]	95.17%	93.52%

B. Simulation Results

The performance of the optimal policy is evaluated against two baseline policies, i.e., all-transmission policy and pre-classification based policy. In the all-transmission policy, regardless of the current state (i.e., TAoI and pre-classification result), the receiver requests transmission from the transmitter. In the pre-classification based policy, the receiver completely relies on the preclassifier and makes transmission decisions based on the preclassification result. Note that all simulation results are obtained by averaging 3 independent runs with

different seeds. For fair comparisons, the same seed is adopted for all policies in one run.

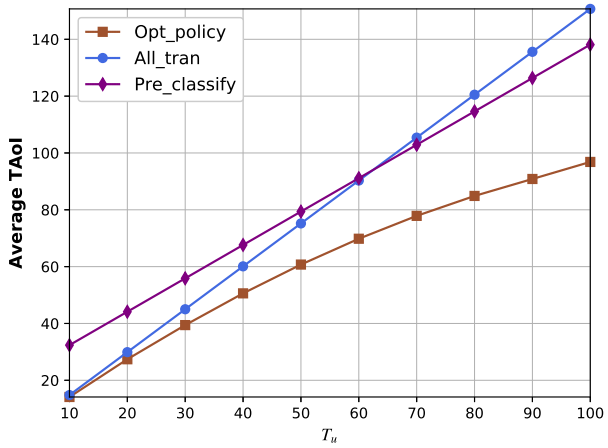
In Fig. 6, the average TAoI of the optimal policy and the two policies are compared with respect to T_u. Firstly, we can observe from both 6a and 6b that as T_u increases, the gap between the optimal policy and the other two baseline policies increases. This is because as T_u increases, the cost of task failure becomes larger, making the advantage of the optimal policy more significant. Secondly, as the misclassification probabilities p_A and p_B increase, the gap between the optimal policy and the baseline policies decreases. This indicates that the classifier has a significant impact on the system. Therefore, it is recommended that the accuracy of the system's classifier be as high as possible, given the physical conditions allow.

In Fig. 7, the average TAoI of the optimal policy and the two policies are compared with respect to the q. Note that in Fig. 7a and 7b, the average TAoI corresponding to $q = 1$ actually corresponds to the value when q is 0.99, due to (23). Firstly, we can observe that as q increases, the TAoI decreases. This is because the more frequently the target appears, the higher the probability of successful monitoring tasks. Additionally, regardless of the value of q and whether the misclassification probabilities are high or low, the average TAoI of our proposed optimal policy is lower than that of all baseline policies. This validates the effectiveness of the proposed policy.

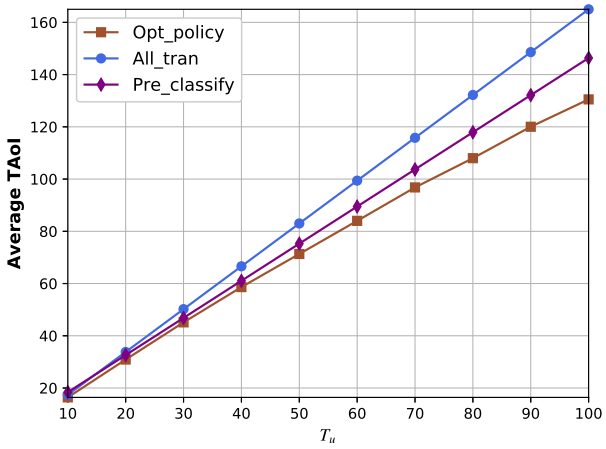
In Fig. 8, the average TAoI of the optimal policy and the two policies are compared with respect to the misclassification probabilities p_A and p_B. Firstly, it can be observed that the average TAoI of the all-transmission policy remains constant in both Fig. 8a and Fig. 8b. This is because the classifier does not affect the process of sensor perception of images, so the all-transmission policy is not affected by the misclassification rates. Secondly, in both Figure a and Figure b, we can see that the average TAoI of the optimal policy increases with the increase of the misclassification probabilities p_A and p_B. This is consistent with the conclusions of Fig. 6 and Fig. 7. Lastly, in Fig. 8a, the pre-classification-based policy shows a decrease in TAoI with the increase of p_A, but it does not go below the optimal policy.

V. Conclusions

We evaluate the proposed policy on binary classification using the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32×32 RGB images, including 50,000 training images and 10,000 test images. The original labels are $\{0, \cdots, 9\}$, corresponding to the objects depicted in each image. We use binary labels $\{0, 1\}$, indicating whether the image depicts an animal or a vehicle. We assume the target of the monitoring task in the system is vehicles, labeled as 1. During the training phase, we can choose a network from Table II as the processor to adjust the misclassification probabilities, p_A and p_B. After obtaining the trained classifier, the system enters the inference phase, i.e., the transmission control phase. At the beginning of each decision epoch (i.e., time step) in the inference phase, the sensor selects an image from the CIFAR-10 dataset with a Bernoulli distribution probability of q.



(a) $p_A = 0.1$, and $p_B = 0.1$.



(b) $p_A = 0.6$, and $p_B = 0.6$.

Fig. 6: Average TAoI versus T_u ($q = 0.9$).

Appendix

A. Proof of Lemma 1

Based on the value iteration algorithm (VIA) outlined in [16, Ch. 4.3], we utilize mathematical induction to establish the proof of Lemma 1. Initially, we introduce $Q_k(\mathbf{s}, a)$ and $V_k(\mathbf{s})$ to represent the state-action value function and the state value function at the k-th iteration, respectively. Particularly, $Q_k(\mathbf{s}, a)$ is defined as

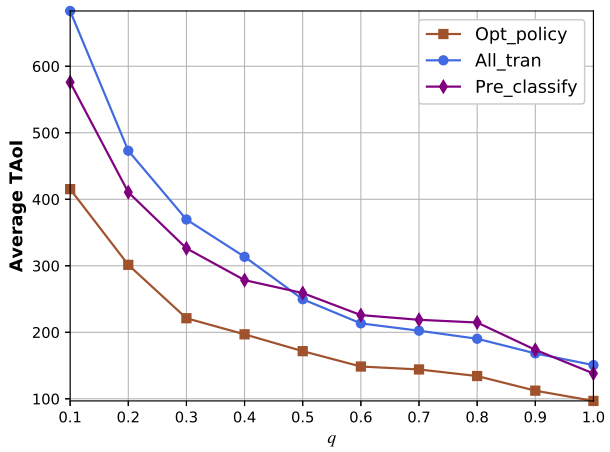
$$Q_k(\mathbf{s}, a) \triangleq \bar{R}(\mathbf{s}, a) + \sum_{\mathbf{s}' \in \mathcal{S}} \bar{p}(\mathbf{s}'|\mathbf{s}, a) V_k(\mathbf{s}'), \ \forall \mathbf{s} \in \mathcal{S}. \quad (24)$$

where $\mathbf{s}'$ is given by (11). For any given state $\mathbf{s}$, the update to the value function can be executed by

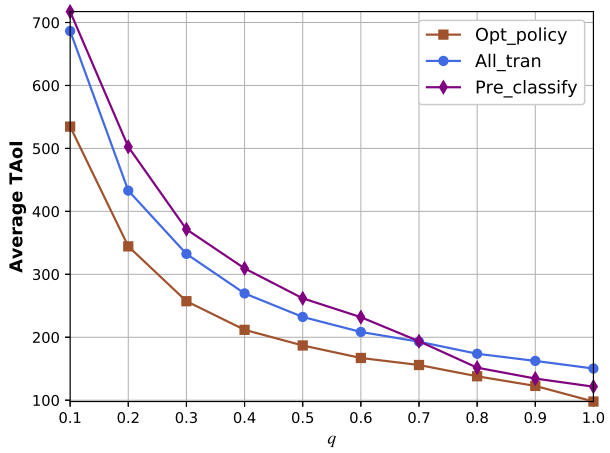
$$V_{k+1}(\mathbf{s}) = \min_{a \in \mathcal{A}} Q_k(\mathbf{s}, a), \ \forall \mathbf{s} \in \mathcal{S}. \quad (25)$$

Regardless of how $V_0(s)$ is initially set, the sequence $\{V_k(\mathbf{s})\}$ converges to $V(\mathbf{s})$ that satisfies the Bellman equation (19) , i.e.,

$$\lim_{k \to \infty} V_k(\mathbf{s}) = V(\mathbf{s}), \ \forall \mathbf{s} \in \mathcal{S}. \quad (26)$$

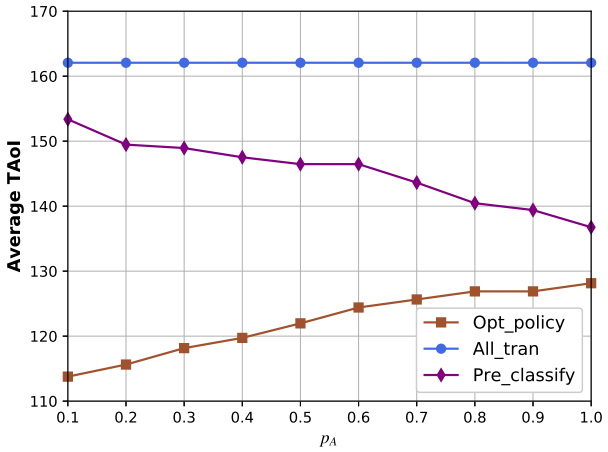


(a) $p_A = 0.1$, and $p_B = 0.1$.

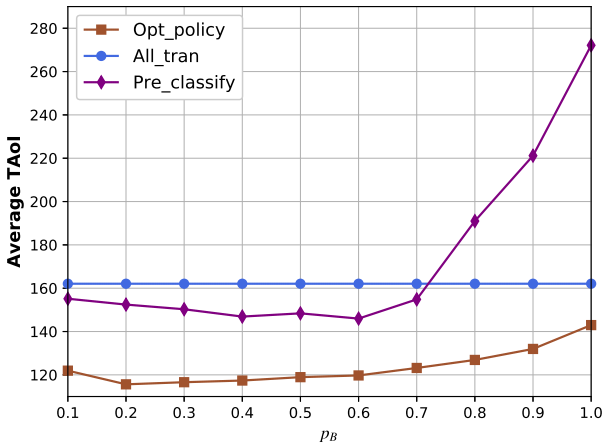


(b) $p_A = 0.6$, and $p_B = 0.6$.

Fig. 7: Average TAoI versus q ($T_u = 100$).



(a) Average TAoI versus p_A ($p_B = 0.2$).



(b) Average TAoI versus p_B ($p_A = 0.2$).

Fig. 8: Average TAoI versus misclassification probabilities ($q = 0.9$, $T_u = 100$).

Therefore, the monotonicity of $V(\mathbf{s})$ is validated by showing that, for any two states $\mathbf{s}_1 = (\Delta_1, F(X)), \mathbf{s}_2 = (\Delta_2, F(X)) \in \mathcal{S}$, whenever $\Delta_1 \leq \Delta_2$, it follows that

$$V_k(\mathbf{s}_1) \leq V_k(\mathbf{s}_2),\ k = 0, 1, \cdots. \tag{27}$$

Next, we prove (27) using mathematical induction. Without loss of generality, we set $V_0(\mathbf{s}) = 0$ for each $s \in \mathcal{S}$, ensuring that (27) is satisfied at $k = 0$. Then, assuming that (27) holds up to $k > 0$, we verify whether it holds for $k + 1$.

For $a = 0$, it follows that

$$Q_k(\mathbf{s}_1, 0) = \Delta_1 + (1 - \epsilon)V_k(\mathbf{s}_1) + \epsilon V_k(\mathbf{s}_1'), \tag{28}$$

and

$$Q_k(\mathbf{s}_2, 0) = \Delta_2 + (1 - \epsilon)V_k(\mathbf{s}_2) + \epsilon V_k(\mathbf{s}_2'), \tag{29}$$

where $\mathbf{s}_1' = (\Delta_1 + 1, F(X)_+)$ and $\mathbf{s}_2' = (\Delta_2 + 1, F(X)_+)$. Given that $\Delta_1 + 1 \leq \Delta_2 + 1$, $V_k(\Delta_1) \leq V_k(\Delta_2)$ and $V_k(\mathbf{s}_1') \leq V_k(\mathbf{s}_2')$, it can be easily deduced that $Q_k(\mathbf{s}_1, 0) \leq Q_k(\mathbf{s}_2, 0)$.

For $a = 1$, it follows that

$$Q_k(\mathbf{s}_1, 1) = \Delta_1 + \frac{1}{2}(T_u - 1) + \frac{\epsilon}{T_u} p_0 V_k(T_u, F(X)_+) + \tag{30}$$

$$\frac{\epsilon}{T_u} p_1 V_k(\Delta_1 + T_u, F(X)_+) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\mathbf{s}_1),$$

and

$$Q_k(\mathbf{s}_2, 1) = \Delta_2 + \frac{1}{2}(T_u - 1) + \frac{\epsilon}{T_u} p_0 V_k(T_u, F(X)_+) + \tag{31}$$

$$\frac{\epsilon}{T_u} p_1 V_k(\Delta_2 + T_u, F(X)_+) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\mathbf{s}_2),$$

where if $F(X) = 1$, then $p_0 = 1 - \hat{p}_A$ and $p_1 = \hat{p}_A$; if $F(X) = 0$, then $p_0 = \hat{p}_B$ and $p_1 = 1 - \hat{p}_B$. Similar to $a = 0$, we can obtain $Q_k(\mathbf{s}_1, 1) \leq Q_k(\mathbf{s}_2, 1)$ according to $\Delta_1 \leq \Delta_2$ and $V_k(\mathbf{s}_1) \leq V_k(\mathbf{s}_2)$.

By (25), we can get that $V_{k+1}(\mathbf{s}_1) \leq V_{k+1}(\mathbf{s}_2)$ for any k. This concludes the proof of Lemma 1.

B. Proof of Lemma 2

The concavity of $V(\mathbf{s})$ with respect to $\mathbf{s}$ for any given $F(X)$ can be demonstrated by showing that, for any $\mathbf{s}_1 =$

$(\Delta_1, F(X))$, $\mathbf{s}_2 = (\Delta_2, F(X)) \in \mathcal{S}$ and $w \in N$, whenever $\Delta_1 \leq \Delta_2$, it follows that

$$\begin{aligned} &V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X)) \geq \\ &V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X)), k = 0, 1, \ldots. \end{aligned} \tag{32}$$

Without sacrificing generality, we set $V_0(\mathbf{s}) = 0$ for all $\mathbf{s} \in \mathcal{S}$, ensuring that (32) is applicable at $k = 0$. Then, we assume that (32) holds up till $k > 0$ and inspect whether it holds for $k + 1$. Now, let $\mathbf{s} = (\Delta, F(X))$, $\mathbf{s}_1 = (\Delta_1, F(X))$, $\mathbf{s}_2 = (\Delta_2, F(X))$, $\mathbf{s}' = (\Delta + w, F(X))$, $\mathbf{s}'_1 = (\Delta_1 + w, F(X))$ and $\mathbf{s}'_2 = (\Delta_2 + w, F(X))$. For ease of explanation, we introduce $\Delta Q(\mathbf{s}', \mathbf{s}, a) = Q(\mathbf{s}', a) - Q(\mathbf{s}, a)$.

For $a = 0$, it follows that

$$\begin{aligned} &\Delta Q_k(\mathbf{s}'_1, \mathbf{s}_1, 0) - \Delta Q_k(\mathbf{s}'_2, \mathbf{s}_2, 0) \\ =&[\Delta_2 + (1-\epsilon)V_k(\Delta_2, F(X)) + \epsilon V_k(\Delta_2 + 1, F(X)_+)] \\ &- [\Delta_1 + (1-\epsilon)V_k(\Delta_1, F(X)) + \epsilon V_k(\Delta_1 + 1, F(X)_+)] \\ &+ [\Delta_1 + w + (1-\epsilon)V_k(\Delta_1 + w, F(X)) \\ &+ \epsilon V_k(\Delta_1 + w + 1, F(X)_+)] \\ &- [\Delta_2 + w + (1-\epsilon)V_k(\Delta_2 + w, F(X)) \\ &+ \epsilon V_k(\Delta_2 + w + 1, F(X)_+)] \\ =&(1-\epsilon)[(V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X))) \\ &- (V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X)))] \\ &+ \epsilon((V_k(\Delta_1 + w + 1, F(X)_+) - V_k(\Delta_1 + 1, F(X)_+)) \\ &- (V_k(\Delta_2 + w + 1, F(X)_+) - V_k(\Delta_2 + 1, F(X)_+))). \end{aligned} \tag{33}$$

Given that $V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X)) \geq V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X))$ and $V_k(\Delta_1 + w + 1, F(X)_+) - V_k(\Delta_1 + 1, F(X)_+) \geq V_k(\Delta_2 + w + 1, F(X)_+) - V_k(\Delta_2 + 1, F(X)_+)$, we can easily see that $\Delta Q_k(\mathbf{s}'_1, \mathbf{s}_1, 0) - \Delta Q_k(\mathbf{s}'_2, \mathbf{s}_2, 0) \geq 0$. Thus, $Q_k(\mathbf{s}, 0)$ is concave in Δ for any given $F(X)$.

For $a = 1$, it follows that

$$\begin{aligned} &\Delta Q(\mathbf{s}'_1, \mathbf{s}_1, 1) - \Delta Q(\mathbf{s}'_2, \mathbf{s}_2, 1) \\ =&\Delta_1 + w + \frac{1}{2}(T_u - 1) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\Delta_1 + w, F(X)) \\ &+ \frac{\epsilon}{T_u}\left(p_0 V_k(T_u, F(X)_+) + p_1 V_k(\Delta_1 + w + T_u, F(X)_+)\right) \\ &- [\Delta_1 + \frac{1}{2}(T_u - 1) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\Delta_1, F(X)) \\ &+ \frac{\epsilon}{T_u}\left(V_k(T_u, F(X)_+) + p_1 V_k(\Delta_1 + T_u, F(X)_+)\right)] \\ &- [\Delta_2 + w + \frac{1}{2}(T_u - 1) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\Delta_2 + w, F(X)) \\ &+ \frac{\epsilon}{T_u}\left(p_0 V_k(T_u, F(X)_+) + p_1 V_k(\Delta_2 + w + T_u, F(X)_+)\right)] \\ &+ [\Delta_2 + \frac{1}{2}(T_u - 1) + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\Delta_2, F(X)) \\ &+ \frac{\epsilon}{T_u}\left(p_0 V_k(T_u, F(X)_+) + p_1 V_k(\Delta_2 + T_u, F(X)_+)\right)] \\ =&(1 - \frac{\epsilon}{T_u})[(V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X))) \\ &- (V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X)))] \\ &+ \frac{\epsilon}{T_u} p_1 [(V_k(\Delta_1 + w + T_u, F(X)_+) \\ &- V_k(\Delta_1 + T_u, F(X)_+)) \\ &- (V_k(\Delta_2 + w + T_u, F(X)_+) \\ &- V_k(\Delta_2 + T_u, F(X)_+))]. \end{aligned} \tag{34}$$

Given that $V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X)) \geq V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X))$ and $V_k(\Delta_1 + w + T_u, F(X)_+) - V_k(\Delta_1 + T_u, F(X)_+) \geq V_k(\Delta_2 + w + T_u, F(X)_+) - V_k(\Delta_2 + T_u, F(X)_+)$, we can also get that $\Delta Q_k(\mathbf{s}'_1, \mathbf{s}_1, 1) - \Delta Q_k(\mathbf{s}'_2, \mathbf{s}_2, 1) \geq 0$. Thus, $Q_k(\mathbf{s}, 1)$ is concave in Δ for any given $F(X)$.

Since the value function $V_{k+1}(\mathbf{s})$ is the minimum of two concave functions, it is also concave in Δ for any given $F(X)$. Hence, we have $V_k(\Delta_1 + w, F(X)) - V_k(\Delta_1, F(X)) \geq V_k(\Delta_2 + w, F(X)) - V_k(\Delta_2, F(X))$, i.e., (32) holds for $k+1$. Therefore, we can show that (32) holds for any k by induction.

This concludes the proof of Lemma 2.

C. Proof of Lemma 3

The proof follows the same procedure of Lemma 1. The lower bound of $V(\mathbf{s}_2) - V(\mathbf{s}_1)$ can be proved by showing that for any $\mathbf{s}_1 = (\Delta_1, F(X))$, $\mathbf{s}_2 = (\Delta_2, F(X)) \in \mathcal{S}$, such that $\Delta_1 \leq \Delta_2$

$$\begin{aligned} V_k(\Delta_2, F(X)) - V_k(\Delta_1, F(X)) \geq \frac{L(a)}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1), \\ k = 0, 1, \cdots. \end{aligned} \tag{35}$$

Without sacrificing generality, we set $V_0(\mathbf{s}) = \frac{L(a)}{\epsilon(1-p_1)}\Delta$ for all $\mathbf{s} = (\Delta, F(X)) \in \mathcal{S}$, ensuring that (35) is satisfied at $k = 0$. Then, we assume that (35) holds up till $k > 0$ and hence we have $V_k(\Delta_2, F(X)) - V_k(\Delta_1, F(X)) \geq \frac{L(a)}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1)$ and $V_k(\Delta_2 + 1, F(X)) - V_k(\Delta_1 + 1, F(X)) \geq \frac{L(a)}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1)$.

Then, we inspect whether it holds for $k+1$. We first consider the case when $a = 1$ and we have $\frac{L(a)}{\epsilon(1-p_1)} = \frac{T_u}{\epsilon(1-p_1)}$. Since $V_{k+1}(\mathbf{s}) = \min_{a \in \mathcal{A}} Q_k(\mathbf{s}, a)$, we investigate the two state-action value functions, in the following, respectively.

When $F(X) = 1$ and $a = 0$, we have

$$\begin{aligned} &\Delta Q_k(\mathbf{s}_2, \mathbf{s}_1) \\ =&Q_k((\Delta_2, F(X)), 0) - Q_k((\Delta_1, F(X)), 0) \\ =&\Delta_2 - \Delta_1 + (1-\epsilon)\left(V_k(\Delta_2, F(X)) - V_k(\Delta_1, F(X))\right) \\ &+ \epsilon(V_k(\Delta_2 + 1, F(X)_+) - V_k(\Delta_1 + 1, F(X)_+)) \\ \geq&(\Delta_2 - \Delta_1) + \frac{L(a)}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1) \\ =&\left(1 + \frac{L(a)}{\epsilon(1-p_1)}\right)(\Delta_2 - \Delta_1) \\ \geq&\frac{L(a)}{\epsilon(1-p_1)}(\Delta_2 - \Delta_1). \end{aligned} \tag{36}$$

When $F(X) = 1$ and $a = 1$, we have

$$\begin{aligned} &\Delta Q_k(\mathbf{s}_2, \mathbf{s}_1) \\ =&Q_k((\Delta_2, F(X)), 1) - Q_k((\Delta_1, F(X)), 1) \\ =&\Delta_2 - \Delta_1 + \left(1 - \frac{\epsilon}{T_u}\right) V_k(\Delta_2, F(X)) - V_k(\Delta_1, F(X)) \end{aligned}$$

$$\begin{aligned}
&\frac{\epsilon}{T_u}p_1\left(V_k(\Delta_2+T_u,F(X)_+)-V_k(\Delta_1+T_u,F(X)_+)\right)\\
\geq&\Delta_2-\Delta_1+\left(1-\frac{\epsilon}{T_u}+\frac{\epsilon}{T_u}p_1\right)\frac{L(a)}{\epsilon(1-p_1)}(\Delta_2-\Delta_1)\\
=&\frac{L(a)}{\epsilon(1-p_1)}(\Delta_2-\Delta_1). \qquad (37)
\end{aligned}$$

When the optimal policy in $\mathbf{s}_1$ and $\mathbf{s}_2$ is two different actions, i.e., a_1 and a_2, we have

$$\begin{aligned}
&V_k(\Delta_2,F(X))-V_k(\Delta_1,F(X))\\
&=Q_k((\Delta_2,F(X)),a_2)-Q_k((\Delta_1,F(X)),a_1)\\
&\geq Q_k((\Delta_2,F(X)),a_2)-Q_k((\Delta_1,F(X)),a_2)\\
&\geq \frac{L(a)}{\epsilon(1-p_1)}(\Delta_2-\Delta_1). \qquad (38)
\end{aligned}$$

This concludes the proof of Lemma 3.

D. Proof of Theorem 1

For any $\mathbf{s}_1=(\Delta_1,F(X))$, $\mathbf{s}_2=(\Delta_2,F(X))\in\mathcal{S}$, such that $\Delta_1\leq\Delta_2$, we have

$$\begin{aligned}
&\Delta Q_k((\Delta_2,F(X)),a)-(V_k(\Delta_2,F(X))-V_k(\Delta_1,F(X)))\\
=&\Delta_2-\Delta_1-\frac{\epsilon}{L(a)}(V(\Delta_2,F(X))-V(\Delta_1),F(X))\\
&+\frac{\epsilon}{L(a)}p_1(V(\Delta_2+L(a),F(X))-V(\Delta_1+L(a),F(X))). \qquad (39)
\end{aligned}$$

Since the concavity of $V(\mathbf{s})$ have been proved in Lemma 2, we can easily see that $V(\Delta_2+L(a),F(X))-V(\Delta_1+L(a),F(X))\leq V(\Delta_2,F(X))-V(\Delta_1+L(a),F(X))$. Therefore, we have

$$\begin{aligned}
&\Delta Q_k((\Delta_2,F(X)),a)-(V_k(\Delta_2,F(X))-V_k(\Delta_1,F(X)))\\
\leq&\Delta_2-\Delta_1-\frac{\epsilon}{L(a)}(V(\Delta_2,F(X))-V(\Delta_1),F(X))\\
&+\frac{\epsilon}{L(a)}p_1(V(\Delta_2,F(X))-V(\Delta_1,F(X)))\\
=&\Delta_2-\Delta_1-\frac{\epsilon}{L(a)}(1-p_1)(V(\Delta_2,F(X))-V(\Delta_1,F(X))). \qquad (40)
\end{aligned}$$

As proved in Lemma 3 that $V_k(\Delta_2,F(X))-V_k(\Delta_1,F(X))\geq[L(a)/\epsilon(1-p_1)](\Delta_2-\Delta_1)$, it is easy to see that $\Delta Q_k(\mathbf{s}_2,\mathbf{s}_1)-(V(\mathbf{s}_2)-V(\mathbf{s}_1))\leq 0$.

Now, we can prove the threshold structure of the optimal policy. Suppose $\Delta_2\geq\Delta_1$ and $\pi^*(\Delta_1,F(X))=a$, it is easily to see that $V(\Delta_1,F(X))=Q((\Delta_1,F(X)),a)$, i.e., $V(\mathbf{s}_1)=Q(\mathbf{s}_1,a)$. According to Theorem 1, we know that $V(\mathbf{s}_2)-V(\mathbf{s}_1)\geq Q(\mathbf{s}_2,a)-Q(\mathbf{s}_1,a)$. Therefore, we have $V(\mathbf{s}_2)\geq Q(\mathbf{s}_2,a)$. Since the value function is a minimum of two state-action cost functions, we have $V(\mathbf{s}_2)\leq Q(\mathbf{s}_2,a)$. Altogether, we can assert that $V(\mathbf{s}_2)=Q(\mathbf{s}_2,a)$ and $\pi^*(\Delta_2,F(X))=a$.

This concludes the proof of Theorem 1.

E. Proof of Lemma 4

When $\hat{\Delta}$ goes to infinity, for any threshold policy with the threshold of Ω, the MDP can be modeled by a Discrete Time Markov Chain (DTMC) with the same states, as shown in Fig. 11. Let $\eta(\mathbf{s})$ denote the steady state probability of state $\mathbf{s}$, where $\mathbf{s}=(\Delta,F(X))$. According to Fig. 11, we have the balance equations of the DTMC as follows:

$$\begin{cases}\eta(\mathbf{s})=\eta(\mathbf{s}'), & 2\leq\Delta\leq\Omega\\ \eta(\mathbf{s})=\hat{p}\eta(\mathbf{s}'), & \Delta>\Omega,\end{cases} \qquad (41)$$

where $\mathbf{s}=(\Delta,F(X))$, $\mathbf{s}'=(\Delta-1,F(X))$, and $\hat{p}=\epsilon(1-p_1)/T_u$. Then, the steady-state probability of the DTMC can be calculated by $\eta(\hat{\mathbf{s}})$, where $\hat{\mathbf{s}}=(T_u,F(X))$. Specifically,

$$\eta(\mathbf{s})=\begin{cases}\eta(\hat{\mathbf{s}}), & 2\leq\Delta\leq\Omega\\ \eta(\hat{\mathbf{s}})\hat{p}^{\Delta-\Omega}, & \Delta>\Omega.\end{cases} \qquad (42)$$

Since $\sum_{\Delta=1}^{\infty}\eta(\Delta,F(X))=1$, we can obtain $\eta(\hat{\mathbf{s}})=\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}$. Hence, we have the closed-form of the steady-state probability as follows:

$$\eta(\mathbf{s})=\begin{cases}\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}, & 2\leq\Delta\leq\Omega\\ \frac{1-\hat{p}^{\Delta-\Omega}}{\Omega(1-\hat{p})+\hat{p}}, & \Delta>\Omega.\end{cases} \qquad (43)$$

According to Eq.(43), the average reward under the threshold policy can be derived by

$$\begin{aligned}
J(\Omega,F(X))=&J_1(\Omega,F(X))+J_2(\Omega,F(X))\\
=&\sum_{\Delta=1}^{\Omega-1}\eta(\hat{\mathbf{s}})\Delta+\sum_{\Delta=\Omega}^{\infty}\eta(\hat{\mathbf{s}})\left(\Delta+\frac{T_u-1}{2}\right)\\
=&\sum_{\Delta=1}^{\infty}\eta(\hat{\mathbf{s}})\Delta+\sum_{\Delta=\Omega}^{\infty}\eta(\hat{\mathbf{s}})\frac{T_u-1}{2}\\
=&\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}\left(\frac{\Omega^2-\Omega}{2}+\frac{\Omega}{1-\hat{p}}+\frac{\hat{p}}{(1-\hat{p})^2}\right)\\
&+\frac{T_u-1}{2\Omega(1-\hat{p})+2\hat{p}}\\
=&\left(\frac{\Omega^2-\Omega}{2}+\frac{2\Omega+T_u-1}{2-2\hat{p}}+\frac{\hat{p}}{(1-\hat{p})^2}\right)\\
&\times\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}. \qquad (44)
\end{aligned}$$

This concludes the proof of Lemma 4.

F. Proof of Theorem 2

To derive the optimal threshold Ω^*, we relax Ω to a continuous variable. First, we calculate the first derivative of $J(\Omega,F(X))$ as follows:

$$\begin{aligned}
\frac{\partial J(\Omega,F(X))}{\partial\Omega}=&\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}\left(\frac{2\Omega-1}{2}+\frac{1}{1-\hat{p}}\right)\\
&-\left(\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}\right)^2\\
&\times\left(\frac{\Omega^2-\Omega}{2}+\frac{2\Omega+T_u-1}{2-2\hat{p}}+\frac{\hat{p}}{(1-\hat{p})^2}\right). \qquad (45)
\end{aligned}$$

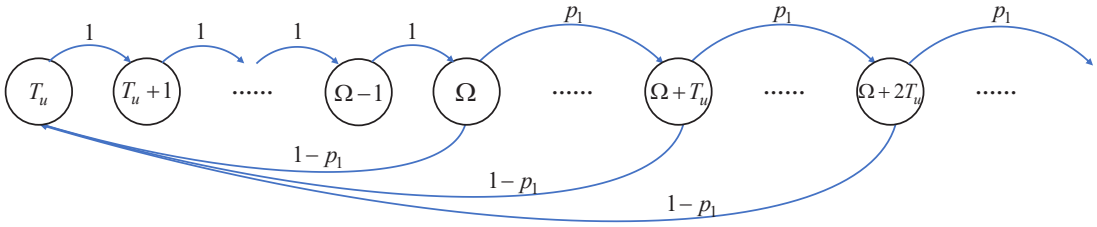


Fig. 9: DTMC

Then, we obtain the second derivative of $J(\Omega, F(X))$ as follows:

$$\begin{aligned}\frac{\partial^2 J(\Omega,F(X))}{\partial\Omega^2} =& \frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}} - 2\left(\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}\right)^2 \\ &\times\left(\frac{\Omega^2-\Omega}{2}+\frac{2\Omega+T_u-1}{2-2\hat{p}}+\frac{\hat{p}}{(1-\hat{p})^2}\right) \\ &+2\left(\frac{1-\hat{p}}{\Omega(1-\hat{p})+\hat{p}}\right)^3 \\ &\times\left(\frac{\Omega^2-\Omega}{2}+\frac{2\Omega+T_u-1}{2-2\hat{p}}+\frac{\hat{p}}{(1-\hat{p})^2}\right) \\ =& \frac{(1-\hat{p})\left(\hat{p}+(T_u-1)(1-\hat{p})\right)}{\left(\Omega(1-\hat{p})+\hat{p}\right)^3}, \end{aligned} \tag{46}$$

where $\hat{p} \leq 1$ and $T_u \geq 1$. Obviously, the second derivative $\partial^2 J(\Omega, F(X))/\partial\Omega^2 \geq 0$. Therefore, the function $J(\Omega, F(X))$ is concave with respect to Ω, and the optimal threshold can be obtained by setting the first derivative $\partial J(\Omega, F(X))/\partial\Omega$ to 0. The solution of $\partial J(\Omega, F(X))/\partial\Omega = 0$ is

$$\Omega' = \frac{\sqrt{\hat{p}+(T_u-1)(1-\hat{p})}-\hat{p}}{1-\hat{p}}. \tag{47}$$

Since Ω' may not be an integer, the optimal threshold can be given by

$$\Omega^* = \arg\min(J(\lfloor\Omega'\rfloor), J(\lceil\Omega'\rceil)). \tag{48}$$

This concludes the proof of Theorem 2.

REFERENCES

[1] X. Luo, H.-H. Chen, and Q. Guo, "Semantic Communications: Overview, Open Issues, and Future Research Directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, 2022.

[2] Y. E. Sagduyu, S. Ulukus, and A. Yener, "Task-Oriented Communications for NextG: End-to-end Deep Learning and AI Security Aspects," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 52–60, 2023.

[3] K. Niu, J. Dai, S. Yao, S. Wang, Z. Si, X. Qin, and P. Zhang, "A Paradigm Shift toward Semantic Communications," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 113–119, 2022.

[4] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, 2023.

[5] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 2023.

[6] Y. Shi, Y. Zhou, D. Wen, Y. Wu, C. Jiang, and K. B. Letaief, "Task-Oriented Communications for 6G: Vision, Principles, and Technologies," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 78–85, 2023.

[7] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE SECON*, 2011, pp. 350–358.

[8] X. Wang, W. Lin, C. Xu, X. Sun, and X. Chen, "Age of Changed Information: Content-Aware Status Updating in the Internet of Things," *IEEE Trans. Wireless Commun.*, vol. 70, no. 1, pp. 578–591, 2022.

[9] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The Age of Incorrect Information: A New Performance Metric for Status Updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, 2020.

[10] J. Cao, X. Zhu, S. Sun, P. Popovski, S. Feng, and Y. Jiang, "Age of Loop for Wireless Networked Control System in the Finite Blocklength Regime: Average, Variance and Outage Probability," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5306–5320, 2023.

[11] A. Nikkhah, A. Ephremides, and N. Pappas, "Age of Actuation in a Wireless Power Transfer System," in *Proc. IEEE INFOCOM WKSHPS*, Hoboken, NJ, USA, May 2023, pp. 1–6.

[12] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of Information: An Introduction and Survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, 2021.

[13] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus, "On the average cost optimality equation and the structure of optimal policies for partially observable markov decision processes," *Ann. Oper. Res.*, vol. 29, no. 1, pp. 439–469, 1991.

[14] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming (Wiley Series in Probability and Statistics)*. Hoboken, NJ, USA: Wiley, 2005.

[15] H. C. Tijms, *A First Course in Stochastic Models*. Chichester, U.K.: Wiley, Dec. 2004.

[16] P. Bertsekas, Dimitri, *Dynamic Programming and Optimal Control-II*, 3rd ed. Belmont, MA, USA: Athena Sci., 2007, vol. 2.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proc. IEEE*, vol. 86, no. 11, 1998.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Proc. NeurIPS*, Lake Tahoe, Nevada, USA, Dec. 2012.

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. ICLR*, San Diego, CA, USA, May 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE/CVF CVPR*, Las Vegas, Nevada, USA, June 2016.