Age of Information Minimization in Backscatter-Assisted Wireless Powered Communication Post-print

Authors: Song Zhaoxi, Tang Dong, Huang Gaofei, Zhao Sai, Liu Guiyun

Date: 2022-04-07T00:00:00+00:00

Abstract

Age of Information (AoI) is a performance metric that quantifies the freshness of captured data from the destination' s perspective. In energy-constrained real-time sensing IoT scenarios, a joint sampling and hybrid backscatter communication update policy is proposed to enhance the system' s AoI performance. This policy minimizes the system' s long-term average AoI by enabling the source to select status sampling actions and transmission modes for the update process. Specifically, the optimization problem is first formulated as an average-cost Markov Decision Process (MDP). Subsequently, when environment dynamics are known, the optimal policy is derived through a relative value iteration algorithm; when environment dynamics are unknown, a Q-learning algorithm with exploration-exploitation methods is adopted to learn the optimal policy via trial-and-error interactions with the environment. Simulation results indicate that the proposed policy significantly improves the system' s AoI performance compared with two reference policies, and also reveal that the system' s AoI performance improves with decreasing update packet size or increasing battery capacity.

Full Text

Preamble

Age of Information Minimization for Backscatter Assisted Wireless Powered Communications

Song Zhaoxi, Tang Dong†, Huang Gaofei, Zhao Sai, Liu Guiyun

School of Electronics & Communication Engineering, Guangzhou University, Guangzhou 510006, China

Abstract: Age of Information (AoI) is a performance metric that captures the freshness of data from the destination' s perspective. In energy-constrained real-time sensing Internet of Things scenarios, this paper proposes a joint sampling and hybrid backscatter communication updating policy to improve the system' s AoI performance. The policy minimizes the long-term average AoI of the system by allowing the source to select state sampling actions and transmission modes for the updating process. Specifically, we first model the optimization problem as an average-cost Markov decision process (MDP). When the dynamic environment information is known, we obtain the optimal strategy through a relative value iteration algorithm. When the system lacks dynamic environment information, we employ a Q-learning algorithm with exploration-exploitation techniques to learn the optimal strategy through trial-and-error interactions with the environment. Simulation results demonstrate that compared with two reference policies, the proposed policy significantly improves the system' s AoI performance. Additionally, we find that the system' s AoI performance improves as the update packet size decreases or the battery capacity increases.

Key words: age of information; wireless powered communication; backscatter communication; Markov decision process; reinforcement learning; Q-learning

0 Introduction

With the development of Internet of Things (IoT) technology, an increasing number of wireless sensor nodes have been deployed in various real-time monitoring systems in recent years, such as environmental monitoring, intelligent transportation, and smart agriculture systems. These IoT applications output decisions based on real-time status updates of physical processes, where the accuracy of decisions depends on the freshness of received information [1]. To measure and quantify the freshness of received information, Kaul et al. [2] proposed the concept of Age of Information (AoI), which quantifies the freshness of received information from the destination' s perspective. AoI is defined as the time elapsed since the latest status update generated at the source successfully arrived at the destination. A shorter time (smaller AoI value) indicates better freshness (better AoI performance). However, the energy-constrained nature of IoT devices prevents them from sending updates in a timely manner, thereby increasing the likelihood that IoT applications receive outdated status updates. Energy harvesting (EH) technology is considered one of the most promising solutions to this problem, as its development has greatly alleviated the energy constraints of IoT devices. It can capture surrounding kinetic, thermal, solar, or radio frequency (RF) energy and convert it into electrical energy to sustain perpetual device operation [3, 4]. In particular, due to the ubiquity of radio waves, RF-based wireless power transfer (WPT) is regarded as a potential energy harvesting technology. On the other hand, backscatter communication (BC) technology, with its ultra-low power consumption characteristics, can be widely applied in energy-constrained IoT and wireless sensor network scenar-

ios to reduce device communication energy consumption and operational costs. Therefore, in time-sensitive IoT networks, combining WPT and BC technologies can reduce the overall system energy consumption while maintaining the freshness of information received by IoT applications and sustaining continuous monitoring services.

Early work on AoI primarily focused on minimizing AoI from a queueing theory perspective, i.e., modeling the update system as a queueing system consisting of a source, service facility, and monitor, and utilizing optimization theory tools to minimize AoI [2, 5]. Recently, references [6–8] investigated the analysis and optimization of AoI in energy harvesting communication systems, where the source uses energy harvested from nature for update transmission. Due to the unpredictability of energy generation, the energy harvesting process is typically modeled as an independent random process. However, when the source harvests energy from surrounding RF signals [9–11], the amount of harvested energy depends on the transmission power of the RF source and the current channel state information (CSI). Reference [12] further considered the generation time of updates and proposed a joint sampling and updating policy. In this policy, the source needs to determine when to generate and transmit update packets, and then implements status update packet transmission through wireless powered communication (WPC) when needed. However, since WPC requires substantial energy for active information transmission, this leads to high power consumption issues, further exacerbating the battery energy constraints at the source.

Unlike WPC, BC is an emerging green low-power communication technology [13] and a promising option for achieving sustainable communication. Specifically, BC can transmit information by reflecting incident signals from external RF sources without generating active RF signals, thus consuming several orders of magnitude less power than WPC. However, BC has limited transmission range and relatively low data rates. To overcome these limitations, references [14–17] studied a hybrid backscatter communication (HBC) scheme combining BC and WPC to maximize system throughput performance, where transmitters can adaptively select BC or WPC for data transmission. In particular, reference [17] proposed a novel hybrid communication protocol that allows hybrid transmitters to adaptively switch among EH, BC, or IT modes within a time block in a fine-grained manner to further improve system throughput performance. However, references [14–17] did not consider how to minimize the system' s AoI value in backscatter-assisted wireless powered communication.

Although literature on AoI in backscatter communication is limited, AoI remains a critical factor. Therefore, developing a sampling and updating policy that minimizes the system' s average AoI is the focus of this paper in time-sensitive IoT applications. While the joint sampling and WPC updating policy proposed in reference [12] improves the system' s AoI performance to some extent, the high power consumption characteristics of WPC indirectly limit further improvement in system AoI performance. In this context, this paper considers combining WPT and BC technologies for status update transmission. By employing both

model-based relative value iteration algorithms and model-free Q-learning algorithms [18] to solve the optimization problem, we propose a joint sampling and HBC updating policy that minimizes the system' s long-term average AoI. This policy allows the source to adaptively select status sampling actions and update transmission modes based on current channel state, battery energy status, and AoI information at both source and destination to further improve system AoI performance.

1 System Model

The system model is shown in [Figure 1: see original paper]. We consider a wireless backscatter sensor network consisting of an energy transmitter (ET), a source node S, and a destination node D. The ET is connected to the power grid and provides RF energy to the source. The source includes a sensor that can perform real-time status sampling of physical processes and a hybrid transmitter that can send status update information to the destination. The hybrid transmitter is equipped with RF energy harvesting circuits, backscatter circuits, and active RF circuits to enable RF energy collection and status information transmission through hybrid backscatter and wireless powered communication.

We assume that system time is divided into time slots indexed by $n = 0, 1, 2, ..., N$. Without loss of generality, each time slot has a duration of 1 second. The source S decides on sampling actions and update modes at the beginning of each time slot, and both status sampling and update transmission can be completed within one time slot. Furthermore, we consider that the source can perform complex tasks, thus the time and energy costs of status sampling cannot be ignored [19]. Let h_n and g_n denote the channel link gains from ET to S and from S to D in time slot n, respectively. We assume they are subject to quasi-static channel fading, meaning the channel state remains constant within one time slot and changes independently between different time slots.

1.1 Monitoring Model

We consider a joint sampling and hybrid backscatter communication updating policy where, at the beginning of each time slot, the source must decide not only the sensor' s status sampling action but also the hybrid transmitter' s status update mode. The schematic diagram of status update modes is shown in [Figure 2: see original paper]. In time slot n, the source can control its hybrid transmitter to perform EH mode for energy collection, or execute single modes such as BC or IT, or perform combination modes such as EH-BC, EH-IT, BC-IT, or EH-BC-IT for status update transmission. For ease of handling, we denote EH mode as a mode, single modes BC and IT for status update transmission as b mode and c mode, respectively, and combination modes EH-BC, EH-IT, BC-IT, and EH-BC-IT as d mode, e mode, f mode, and g mode, respectively.

Let $\mathbf{m}_n = (w_n, z_n)$ denote the action vector in time slot n, where $w_n \in \{0, 1\}$

represents the source' s status sampling action, and $z_n \in \{a, b, c, d, e, f, g\}$ represents the source' s status update mode. If the source performs status sampling in time slot n, then $w_n = 1$; otherwise, $w_n = 0$. If $z_n = a$, it means the source performs energy harvesting in time slot n; otherwise, the source transmits status updates through mode $k' \in \{b, c, d, e, f, g\}$ in time slot n.

1.2 Energy Harvesting Model

We assume the energy transmitter ET continuously sends RF energy to source S with constant power P_{max}. The source stores harvested energy in a battery with capacity B for future status information sampling and update packet transmission. Let $\mathbf{t}_n = (t_{EH}(n), t_{BC}(n), t_{IT}(n))$ denote the mode operation time vector, where $t_{EH}(n)$, $t_{BC}(n)$, and $t_{IT}(n)$ represent the operation times of EH, BC, and IT modes in time slot n, respectively.

Therefore, the time allocation for different modes at the source must satisfy the following constraints: For mode a, $t_{EH}(n) = 1$; for mode b, $t_{BC}(n) = 1$; for mode c, $t_{IT}(n) = 1$; similarly, for mode d, $t_{EH}(n) + t_{BC}(n) = 1$; for mode e, $t_{EH}(n) + t_{IT}(n) = 1$; for mode f, $t_{BC}(n) + t_{IT}(n) = 1$; finally, for mode g, $t_{EH}(n) + t_{BC}(n) + t_{IT}(n) = 1$. For ease of handling, these equalities can be expressed as:

$$at_{EH}(n)+bt_{BC}(n)+ct_{IT}(n)+d(t_{EH}(n)+t_{BC}(n))+e(t_{EH}(n)+t_{IT}(n))+f(t_{BC}(n)+t_{IT}(n))+g(t_{EH}(n)+t_{BC}(n)+t_{I}$$

Let $E_H(n)$ and $E_T(n)$ denote the energy harvested and consumed by the source' s hybrid transmitter when operating in mode m during time slot n, respectively. The consumed energy includes circuit energy consumption in BC and IT modes and energy consumption for transmitting status update packets. Therefore, the harvested and consumed energy at the source can be expressed as:

$$E_H(m,n) = \begin{cases} \eta P_{max} h_n t_{EH}(n), & \text{if } m \in \{a, d, e, g\} \\ 0, & \text{otherwise} \end{cases}$$

$$E_T(m,n) = \begin{cases} 0, & \text{if } m = a \\ P_c^{BC} t_{BC}(n), & \text{if } m \in \{b, d, f, g\} \\ P_c^{IT} t_{IT}(n) + p_n t_{IT}(n), & \text{if } m \in \{c, e, f, g\} \end{cases}$$

where $\eta \in (0, 1)$ is the RF-to-DC energy conversion efficiency, $\alpha_n \in [0, 1]$ is the backscatter coefficient in time slot n, P_c^{BC} and P_c^{IT} are the circuit power consumption in BC and IT modes, respectively, and p_n is the active information transmission power of the source in time slot n.

According to the Shannon formula, the packet size transmitted in BC mode during time slot n is:

$$R_{BC}(n) = \log_2\left(1 + \frac{\alpha_n P_{max} h_n g_n}{\delta^2}\right)$$

where δ^2 is the noise power at the destination. Similarly, the packet size transmitted in IT mode during time slot n is:

$$R_{IT}(n) = \log_2\left(1 + \frac{p_n g_n}{\delta^2}\right)$$

If the source decides to transmit a status update packet of size M bits in time slot n, the backscatter coefficient α_n and active information transmission power p_n must satisfy the following constraints:

$$M \leq R_{BC}(n), \quad M \leq R_{IT}(n)$$

Let B_n denote the battery energy state of the source in time slot n, where $B_n \in \{0, \frac{B_{max}}{q}, 2\frac{B_{max}}{q}, \ldots, B_{max}\}$ represents the quantized battery energy level, with q being the maximum quantization level and B_{max} the battery capacity. The battery energy must satisfy the following energy causality constraint:

$$B_{n+1} = \max\left\{\min\{B_n - w_n E_s + E_H(m,n) - E_T(m,n), B_{max}\}, 0\right\}$$

where E_s is the energy cost of status sampling. Therefore, the battery energy evolution at the source can be expressed as:

$$B_{n+1} = \max\left\{\min\left\{B_n - w_n E_s + E_H(m,n) - E_T(m,n), B_{max}\right\}, 0\right\}$$

1.3 Age of Information Model

AoI is defined as the time elapsed since the latest update generated at the source arrived at the destination. Let C_n and A_n denote the AoI at the source and destination in time slot n, respectively, where $C_n \in \{1, 2, \ldots, C_{max}\}$ and $A_n \in \{1, 2, \ldots, A_{max}\}$ represent the upper bounds of AoI at the source and destination. We assume that status sampling at the source requires a time cost of 1 time slot and an energy cost of E_s. If the source decides to perform status sampling, then due to the 1 time slot sampling cost, C_n is set to 1; otherwise, C_n increases linearly by 1. Therefore, the AoI dynamics at the source can be expressed as:

$$C_{n+1} = \begin{cases} 1, & \text{if } w_n = 1 \\ \min\{C_n + 1, C_{max}\}, & \text{if } w_n = 0 \end{cases}$$

For simplicity, this can be rewritten as:

$$C_{n+1} = \min\{w_n + C_n + 1, C_{max}\}$$

We also assume that transmitting status updates at the source requires 1 time slot of transmission time. If the source decides to perform a status update, then A_n is set to C_n; otherwise, A_n increases linearly by 1. Therefore, the AoI dynamics at the destination can be expressed as:

$$A_{n+1} = \begin{cases} \min\{C_n + 1, A_{max}\}, & \text{if } k' \in \{b, c, d, e, f, g\} \\ \min\{A_n + 1, A_{max}\}, & \text{if } z_n = a \end{cases}$$

For simplicity, this can be represented by the following constraint:

$$A_{n+1} = \min\{k' A_n + C_n + 1, A_{max}\}$$

1.4 Optimization Problem

Let $\pi = (x(0), x(1), \ldots, x(N))$ denote a deterministic decision policy adopted by the source, which determines the status sampling and update mode decisions for each time slot. Here, $x(n)$ represents a specific status sampling action and update mode taken by the source in time slot n, and Π is the set of all possible policies. If the source adopts policy π, the long-term average AoI at the destination can be expressed as:

$$\bar{A}^{\pi} = \limsup_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} A_n^{\pi}$$

Our objective is to find an age-optimal policy π^* that minimizes the long-term average AoI at the destination. Therefore, finding the age-optimal policy π^* corresponds to solving the following problem (P1):

$$(\text{P1}): \quad \min_{\pi \in \Pi} \limsup_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} A_n^{\pi}$$

subject to constraints (1), (6), (8), (10), (12), and (14)-(16).

2 Optimal Decision Policy

The independence of channel states over time leads to uncertainty in the source' s energy state and its energy state transitions. Therefore, the problem of minimizing long-term average AoI is a stochastic optimization problem. To solve this problem, we first convert it into an MDP problem. Then, for the case where environmental dynamic information is known, we solve the problem using a model-based relative value iteration algorithm in Section 2.3. For the case

where environmental dynamic information is unknown, we propose a model-free Q-learning algorithm to solve the problem in Section 2.4.

2.1 Markov Decision Process

Due to the independence of channel gains h_n and g_n over time and the Markovian nature of the source's decision process, we can model the long-term average AoI minimization problem as an infinite-horizon MDP. Following [20], we detail the main components of the MDP below.

a) State Space: Since actual channel gains are continuous random variables, we adopt the FSMC model [21] to partition the channel gains into K discrete levels with equal probability. In this case, we define the system state in time slot n as $s_n = (B_n, C_n, A_n, h_n, g_n) \in \mathcal{S}$, where $\mathcal{S}$ is the state space containing all possible system states, which is a finite set.

b) Action Space: In time slot n, the source needs to decide the sensor' s sampling action $w_n \in \{0, 1\}$ and determine the operation parameters of the update mode (including backscatter coefficient α_n, active information transmission power p_n, and mode time allocation vector $\mathbf{t}_n$). Therefore, the action taken by the source in state s_n can be expressed as $x_n = (w_n, z_n, \alpha_n, p_n, \mathbf{t}_n) \in \chi(s_n)$, where $\chi(s_n)$ represents the action space under system state s_n.

c) Transition Probability: For simplicity, we use $s = (B, C, A, h, g)$ to denote the current system state and $s' = (B', C', A', h', g')$ to denote the next system state. Since state variables are independent of each other, the probability of transitioning from s to s' given the current system state s and action x is:

$$P(s'|s, x) = P(B'|B, x) \cdot P(C'|C, x) \cdot P(A'|A, C, x) \cdot P(h'|h) \cdot P(g'|g)$$

d) Reward Function: Let $G(s, x)$ denote the immediate cost of taking action x in system state s at time slot n. Then $G(s, x)$ can be defined as:

$$G(s, x) = A$$

2.2 Problem Transformation

Based on the MDP components described in Section 2.1, the system state space and action space of optimization problem (P1) are finite, allowing it to be converted into a finite-state finite-action average-cost MDP problem. In particular, the per-stage average cost of the optimization problem corresponds to the reward function (18) of the MDP. Therefore, given the initial state s_0, we can rewrite problem (P1) as:

$$(\text{P2}): \quad \min_{\pi \in \Pi} \limsup_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} G(s_n, x_n)$$

2.3 Relative Value Iteration Algorithm

Since problem (P2) is a finite-state finite-action MDP problem, there exists an optimal deterministic stationary policy [20]. Moreover, since the policy is stationary, the time index can be omitted hereafter. According to [22], for average-cost MDP problems, the optimal policy can be obtained by solving the following Bellman equation:

$$V(s) + A^* = \min_{x \in \chi(s)} \left\{ G(s,x) + \sum_{s' \in \mathcal{S}} P(s'|s,x) V(s') \right\}, \quad \forall s \in \mathcal{S}$$

where A^* is the optimal long-term average AoI and $V(s)$ is the relative value function, defined as:

$$V(s) = \lim_{N \to \infty} \mathbb{E}^{\pi} \left[\sum_{n=0}^{N-1} (G(s_n, x_n) - A^*) \mid s_0 = s \right]$$

Therefore, the optimal policy for long-term average AoI can be obtained by solving:

$$\pi^*(s) = \arg \min_{x \in \chi(s)} \left\{ G(s,x) + \sum_{s' \in \mathcal{S}} P(s'|s,x) V(s') \right\}$$

To obtain A^* and $V(s)$, we adopt the relative value iteration algorithm (RVIA) [22] to iteratively solve the Bellman equation (19) when the channel transition probabilities are known. Specifically, for any initial state s_0, in the k-th iteration of RVIA, we have:

$$V_{k+1}(s) = \min_{x \in \chi(s)} \left\{ G(s,x) + \sum_{s' \in \mathcal{S}} P(s'|s,x) V_k(s') \right\} - \min_{x \in \chi(s_0)} \left\{ G(s_0,x) + \sum_{s' \in \mathcal{S}} P(s'|s_0,x) V_k(s') \right\}$$

The algorithm converges to the per-stage optimal average cost A^* when the Bellman error in the k-th iteration satisfies $|c_{max}^{k+1} - c_{min}^{k+1}| \leq \epsilon$, where c_{max}^{k+1} and c_{min}^{k+1} are defined as:

$$c_{max}^{k+1} = \max_{s \in \mathcal{S}} \{V_{k+1}(s) - V_k(s)\}, \quad c_{min}^{k+1} = \min_{s \in \mathcal{S}} \{V_{k+1}(s) - V_k(s)\}$$

At this point, the corresponding optimal policy can be obtained through equation (22). The detailed steps of the algorithm are shown in Algorithm 1.

Algorithm 1: Relative Value Iteration Algorithm

Input: Initial system state s_0, and Bellman error threshold ϵ.

Output: A^*, and optimal policy π^*.

a) Initialize $V_0(s) = 0$ for all $s \in \mathcal{S}$, and set iteration index $k = 0$.

b) While $|c_{max}^k - c_{min}^k| > \epsilon$, repeat the following steps:

c) For each state $s \in \mathcal{S}$, compute $V_{k+1}(s)$ using equation (21).

d) Update c_{max}^{k+1} and c_{min}^{k+1}, and increment iteration index $k = k + 1$.

e) After convergence, compute $A^* = \frac{c_{max}^k + c_{min}^k}{2}$ and obtain the optimal policy π^* using equation (22).

2.4 Q-Learning Algorithm

In practical environments, channel state transition probabilities are typically difficult to obtain. Therefore, we adopt a model-free Q-learning online algorithm [18] to solve problem (P2) by iteratively finding the optimal policy. Specifically, in the Q-learning process, the source continuously interacts with the environment through trial-and-error to estimate and learn the optimal action-value function. The source then selects actions in the current state based on the learned Q-values. To ensure that the estimated action-value function eventually converges to the optimal action-value function, we use the ϵ-greedy policy to balance exploration and exploitation. This guarantees that sufficiently rich environmental states are explored while utilizing the explored state information to minimize the system' s long-term average AoI. Thus, in each time slot, the source selects a random action with probability ϵ and the optimal action with probability $1 - \epsilon$.

Mathematically, action selection following the ϵ-greedy policy can be expressed as:

$$x_n = \begin{cases} \arg\min_{x \in \chi(s_n)} Q(s_n, x), & \text{if } p_{rand} \leq 1 - \epsilon \\ \text{a random action } x \in \chi(s_n), & \text{otherwise} \end{cases}$$

where $p_{rand} \in (0, 1)$ is a randomly generated probability in the current time slot, and $Q(s_n, x_n)$ represents the action-value function. Specifically, given a state-action pair (s_n, x_n), the iterative update formula for Q-learning can be expressed as:

$$Q(s_n, x_n) \leftarrow Q(s_n, x_n) + \gamma_n \left[G(s_n, x_n) + \min_{x' \in \chi(s_{n+1})} Q(s_{n+1}, x') - Q(s_n, x_n) \right]$$

where γ_n is the learning rate at time slot n. To accelerate the learning speed of the Q-learning algorithm and ensure the source explores sufficient state information, we typically set larger learning rate γ_n and greedy rate ϵ_n in the initial

iteration period. On the other hand, as the number of iterations increases, we gradually reduce the learning rate and greedy rate so that the estimated action-value function can converge quickly and smoothly to the optimal action-value function.

The detailed steps of the Q-learning algorithm are shown in Algorithm 2.

Algorithm 2: Q-Learning Algorithm

Input: Initial system state s_0, learning rate γ_n and greedy rate ϵ_n.

Output: Learned policy π^*.

a) Initialize $Q(s,x)=0$ for all $s \in \mathcal{S}$ and $x \in \chi(s)$, and randomly select an initial state s_0.

b) While time slot n is less than a preset value, repeat the following steps:

c) In the current state s_n, select action x_n according to the ϵ-greedy policy: choose a random action with probability ϵ_n and the optimal action with probability $1-\epsilon_n$.

d) Take action x_n, interact with the environment to obtain reward $G(s_n,x_n)$ and next system state s_{n+1}.

e) Update the action-value $Q(s_n,x_n)$ using equation (29), increment time slot $n=n+1$, and go to step b.

f) Finally, compute $\pi^*(s)=\arg\min_{x\in\chi(s)} Q(s,x)$ to obtain the learned policy.

3 Simulation Results and Performance Analysis

In this section, we analyze the performance of the proposed joint sampling and hybrid backscatter communication updating policy. To evaluate its performance, we compare it with the joint sampling and WPC updating policy (denoted as Policy A) [12] and the joint sampling and BC updating policy (denoted as Policy B). The simulation results demonstrate the performance of Algorithm 1 when channel dynamic information is known, and the performance of the Q-learning algorithm proposed in Algorithm 2 when channel dynamic information is lacking.

3.1 Simulation Parameter Settings

In the simulations, we set the source' s energy harvesting efficiency to $\eta=0.7$ and the distance between source S and destination D to $d_{SD}=10$ m. The noise power at the destination is $\delta^2=-95$ dBm [12]. The distance between energy transmitter ET and source S is $d_{ES}=10$ m. Path loss is modeled as $L=20\log_{10}(d/d_0)$ [17, 23], where d is the channel link distance. Each time slot duration is set to 1 second, and the transmit power of the energy transmitter is $P_{max}=25$ dBm [17]. The average channel gain of the source' s uplink is

$\bar{h} = 10^{-3}$ [17]. The circuit energy consumption for backscatter communication and active information transmission are set to $P_c^{BC} = 8.9$ μW and $P_c^{IT} = 113$ μW, respectively [17, 24]. The energy cost of status sampling is $E_s = 18$ μJ [12]. The source' s backscatter coefficient α_n is discretized into 5 levels, while all other state and action variables are discretized into 10 levels. Specifically, since channel gains are partitioned using the equal-probability method, the channel state transition probability is $P(h'|h) = P(g'|g) = 1/K$.

3.2 Performance Analysis

The simulation results in [Figure 3: see original paper]-[Figure 5: see original paper] show the performance of the relative value iteration algorithm when channel dynamic information is known. [Figure 3: see original paper] displays the achievable optimal long-term average AoI for different policies as the transmit power of ET varies, with update packet size $M = 8$ Mbits. It can be observed that regardless of ET's transmit power, the proposed policy significantly outperforms both the joint sampling and WPC updating policy and the joint sampling and BC updating policy. This is because the proposed policy combines the low-power characteristics of BC mode and the high-rate characteristics of active IT mode, enabling it to select the optimal update packet transmission mode under different channel conditions. Specifically, under the proposed policy, when ET' s transmit power is small, the source stores less energy in its battery and can choose BC mode or combination modes like BC-IT for urgent update packet transmission. When ET' s transmit power is large, the source can store more energy in its battery, thus having more opportunities to transmit update packets to the destination even when channel conditions are poor.

Additionally, it can be observed that when ET' s transmit power is low, Policy B achieves lower AoI than Policy A, while when ET' s transmit power is high, Policy A achieves lower average AoI than Policy B. This is because Policy A requires higher update energy costs. When ET' s transmit power is low, the source does not have sufficient energy for timely update transmission, resulting in higher achievable optimal average AoI compared to Policy B. However, as ET' s transmit power increases, the source harvests more energy, and due to the higher transmission rate of active IT mode compared to BC mode, Policy A' s achievable optimal average AoI gradually becomes lower than that of Policy B.

[Figure 4: see original paper] compares the achievable optimal long-term average AoI of different policies as the update packet size M varies. The performance of the proposed policy is superior to both Policy A and Policy B, and the optimal average AoI of all policies increases monotonically as the status update packet size increases. It can also be observed that when update packets are small, Policy B' s average AoI performance is significantly better than Policy A' s. However, when update packets are large, Policy A' s average AoI performance is better than Policy B' s, because active IT mode has a faster transmission rate than BC mode and can transmit larger update packets.

[Figure 5: see original paper] plots the optimal long-term average AoI versus update packet size for different sampling costs E_s and battery capacities B_{max}. Specifically, since the unit energy quantum in the parameter setting with $B_{max} = 0.6$ mJ is twice that in the setting with $B_{max} = 0.3$ mJ, to ensure equal sampling energy costs when comparing different battery capacities, we set $E_s = 1$ when $B_{max} = 0.3$ mJ and $E_s = 2$ when $B_{max} = 0.6$ mJ. The simulation results clearly show that the system' s optimal long-term average AoI decreases as E_s decreases or B_{max} increases. This is because a smaller E_s allows the source to save more energy, and a larger B_{max} allows the source to store more energy, both of which increase the possibility of the source' s continuous operation in the future. Additionally, since increasing battery capacity enables the transmission of larger status update packets, increasing battery capacity improves AoI performance more than reducing sampling energy cost, especially when update packets are large.

[Figure 6: see original paper] shows the system average AoI performance achieved by the model-based relative value iteration algorithm and the model-free Q-learning algorithm after convergence over 10^4 time slots. Specifically, since the relative value iteration algorithm knows the precise statistical model of the environment (such as channel state transition probabilities), it serves as a performance lower bound (optimal performance) for the Q-learning algorithm. It can be observed that the average AoI of both algorithms decreases as ET' s transmit power increases, and the performance of the Q-learning algorithm is very close to that of the relative value iteration algorithm. Specifically, the Q-learning algorithm' s performance approaches 96.23% of the relative value iteration algorithm' s performance overall. Therefore, even when the source lacks channel dynamic information, adopting the Q-learning algorithm can still achieve high system AoI performance.

4 Conclusion

This paper studied the long-term average AoI minimization problem in a backscatter-assisted wireless powered communication system. To improve the system' s AoI performance, we proposed a joint sampling and HBC updating policy where the source can dynamically select the sensor' s sampling action and the transmitter' s update mode. To obtain the optimal policy, we first modeled the problem as an infinite-horizon average-cost MDP with finite states and actions. Then, in scenarios where channel dynamic information is known, we solved the problem iteratively through a relative value iteration algorithm. In scenarios where channel dynamic information is unknown, we adopted a model-free Q-learning algorithm to learn the optimal policy. Numerical results demonstrated that the proposed policy significantly outperforms both the joint sampling and WPC updating policy and the joint sampling and BC updating policy. Additionally, we found that even without channel dynamic information, the Q-learning algorithm can achieve high AoI performance through trial-and-error interaction and learning. In future work, we will

consider a multi-source dual-hop relay network scenario for backscatter-assisted wireless powered communication and seek age-optimal policies through deep reinforcement learning algorithms to optimize the system' s AoI performance.

References

[1] Abd-Elmagid M A, Pappas N, Dhillon H S. On the role of age of information in the internet of things [J]. IEEE Communications Magazine, 2019, 57 (12): 72-77.

[2] Kaul S, Yates R, Gruteser M. Real-time status: how often should one update? [C]// Proc of IEEE INFOCOM. Piscataway, NJ: IEEE Press, 2012: 2731-2735.

[3] Ma D, Lan G, Hassan M, et al. Sensing, computing, and communications for energy harvesting IoTs: a survey [J]. IEEE Communications Surveys & Tutorials, 2020, 22 (2): 1222-1250.

[4] Ponnimbaduge Perera T D, Jayakody D N K, Sharma S K, et al. Simultaneous wireless information and power transfer (SWIPT): recent advances and future challenges [J]. IEEE Communications Surveys & Tutorials, 2018, 20 (1): 264-302.

[5] 孙径舟, 王乐涵, 孙宇璇, 等. 面向 6G 网络的信息时效性度量及研究进展 [J]. 电信科学, 2021, 37 (6): 3-13. (Sun Jingzhou, Wang Lehan, Sun Yuxuan, et al. Information timeliness metrics and research progress for 6G network [J]. Telecommunications Science, 2021, 37 (6): 3-13.)

[6] Ponnimbaduge Perera T D, Jayakody D N K, Pitas I, et al. Age of information in SWIPT-enabled wireless communication system for 5GB [J]. IEEE Wireless Communications, 2020, 27 (5): 162-167.

[7] Arafa A, Yang Jing, Ulukus S, et al. Age-minimal transmission for energy harvesting sensors with finite batteries: online policies [J]. IEEE Trans on Information Theory, 2020, 66 (1): 534-556.

[8] Leng Shiyang, Yener A. Age of information minimization for an energy harvesting cognitive radio [J]. IEEE Trans on Cognitive Communications and Networking, 2019, 5 (2): 427-439.

[9] Krikidis I. Average age of information in wireless powered sensor networks [J]. IEEE Communications Letters, 2019, 8 (2): 628-631.

[10] Abd-Elmagid M A, Dhillon H S, Pappas N. A reinforcement learning framework for optimizing age of information in RF-powered communication systems [J]. IEEE Trans on Communications, 2020, 68 (8): 4747-4760.

[11] 刘玲珊, 熊轲, 张煜, 等. 信息年龄受限下最小化无人机辅助无线供能网络的能耗: 一种基于 DQN 的方法 [J]. 南京大学学报: 自然科学, 2021, 57 (5): 847-856. (Liu Lingshan, Xiong Ke, Zhang Yu, et al. Energy minimization in UAV-assisted wireless powered sensor networks with AoI constraints: A DQN-based approach [J]. Journal of Nanjing University: Natural Science, 2021, 57 (5): 847-856.)

[12] Abd-Elmagid M A, Dhillon H S, Pappas N. AoI-optimal joint sampling and updating for wireless powered communication systems [J]. IEEE Trans on Vehicular Technology, 2020, 69 (11): 14110-14115.

[13] Liu V, Parks A, Talla V, et al. Ambient backscatter: wireless communication out of thin air [J]. ACM SIGCOMM Computer Communication Review, 2013, 43 (4): 39-50.

[14] Lu Xiao, Niyato D, Jiang Hai, et al. Ambient backscatter assisted wireless powered communications [J]. IEEE Wireless Communications, 2018, 25 (2): 170-177.

[15] Li Dong, Peng Wei, Liang Yingchang. Hybrid ambient backscatter communication systems with harvest-then-transmit protocols [J]. IEEE Access, 2018, 6: 45288-45298.

[16] 叶迎晖, 施丽琴, 卢光跃. 反向散射辅助的无线供能通信网络中用户能效公平性研究 [J]. 通信学报, 2020, 41 (7): 84-94. (Ye Yinghui, Shi Liqin, Lu Guangyue. User-centric energy efficiency fairness in backscatter-assisted wireless powered communication network [J]. Journal on Communications, 2020, 41 (7): 84-94.)

[17] Long Yusi, Huang Gaofei, Tang Dong, et al. Achieving high throughput in wireless networks with hybrid backscatter and wireless-powered communications [J]. IEEE Internet of Things Journal, 2021, 8 (13): 10658-10671.

[18] Sutton R S, Barto A G. Reinforcement Learning: An Introduction [M]. Cambridge, MA: MIT Press, 2018.

[19] Zhou Bo, Saad W. Joint status sampling and updating for minimizing age of information in the internet of things [J]. IEEE Trans on Communications, 2019, 67 (11): 7468-7482.

[20] Puterman M L. Markov decision processes: discrete stochastic dynamic programming [M]. New York: Wiley, 1994.

[21] Sadeghi P, Kennedy R A, Rapajic P B, et al. Finite-state Markov modeling of fading channels-a survey of principles and applications [J]. IEEE Signal Processing Magazine, 2008, 25 (5): 57-80.

[22] Bertsekas D P. Dynamic programming and optimal control [M]. Belmont, MA: Athena Scientific, 2005.

[23] Zhou Xun, Zhang Rui, Ho C K. Wireless information and power transfer: architecture design and rate-energy tradeoff [J]. IEEE Trans on Communications, 2013, 61 (11): 4754-4767.

[24] Lu Xiao, Jiang Hai, Niyato D, et al. Wireless powered device to device communications with ambient backscattering: performance modeling and analysis [J]. IEEE Trans on Wireless Communications, 2018, 17 (3): 1869-1884.

Note: Figure translations are in progress. See original paper for figures.