

Friedrich Pyttel, Wanja de Sombre, Andrea Ortiz and Anja Klein

"Age of Information Minimization in Status Update Systems with Imperfect Feedback Channel", in *IEEE International Conference on Communications (ICC), Denver, USA, June 2024*.

©2024 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this works must be obtained from the IEEE.

# Age of Information Minimization in Status Update Systems with Imperfect Feedback Channel

Friedrich Pyttel, Wanja de Sombre, Andrea Ortiz, Anja Klein  
Communications Engineering Lab, Technical University of Darmstadt, Germany.  
{w.sombre, a.ortiz, a.klein}@nt.tu-darmstadt.de, friedrich.pyttel@stud.tu-darmstadt.de

**Abstract**—Status Update System (SUS) are monitoring applications of Internet of Things (IoT). They are formed by a sender that monitors a remote process and sends status updates to a receiver over a wireless channel. For successful monitoring, the sender must keep the status updates at the receiver fresh. This freshness is generally measured using the Age of Information (AoI) metric. The aim of the sender is to find a monitoring and transmission strategy that minimizes the AoI. To find the optimal strategy, the sender needs to accurately track the AoI at the receiver, i.e., it needs to perfectly know whether a transmitted status update is correctly received or not. This knowledge can be achieved by using a feedback channel between receiver and sender to send acknowledge (ACK) or negative acknowledge (NACK) messages. However, in real applications, the feedback channel is not perfect, and the transmission of ACK/NACK messages might fail. This means, the monitoring and transmission decisions have to be made under uncertainty about the receiver's AoI. To overcome this challenge, we introduce the concept of a so-called belief distribution and propose a joint monitoring and transmission strategy at the sender based on reinforcement learning. Our approach, termed Belief Learning, exploits the belief distribution to minimize the AoI at the receiver. Through numerical simulations we show that Belief Learning enables the sender to achieve near-optimal performance with respect to the perfect feedback channel case.

## I. INTRODUCTION

Modern Internet of Things (IoT) devices enable widespread monitoring, e.g., of remote environmental processes [1], [2] or industrial facilities [3]. Such monitoring applications of IoT devices are commonly known as Status Update Systems (SUSs). A SUS is formed by a sender and a receiver. The sender performs the monitoring and sends status updates to the receiver over a wireless communication channel. For successful monitoring, the sender must keep the status updates at the receiver fresh. This freshness of the status updates can be evaluated using different metrics depending on the considered scenario, e.g., Age of Information (AoI), Age of Incorrect Information (AoII) or Query Age of Information (QAoI). An extensive overview on the topic is given in [4].

Among the available metrics, one of the most popular is AoI. AoI was first introduced in [5], [6], and measures the time elapsed since the generation of a status update [5]. In

order to keep the AoI at the receiver low, and thus information about the monitored process fresh, the sender needs to devise monitoring and transmission strategies and to track the correct reception of the transmitted status updates at the receiver. Using its monitoring and transmission strategy, the sender decides when to monitor the remote process and when to transmit status updates. The challenge in devising such strategies for SUS comes from the fact that the sender is usually battery operated. As a result, the available energy needs to be efficiently allocated for monitoring and transmission in order to ensure the freshness of status updates at the receiver.

Existing works on AoI in SUS focus on the design of transmission strategies that minimize the AoI at the receiver under different assumptions. In [7], [8], transmission strategies minimizing the AoI are proposed for a sender which is constantly monitoring the remote process. The authors assume the sender is battery-operated with a fixed and limited amount of available energy. In [9]–[12] it is assumed that the battery-operated sender has no control over the monitoring. Instead, the generation of status updates at the sender is modeled as a random process, where status updates are generated at random time instants. In this setting, different transmission strategies are proposed depending on the amount of energy available for transmission and the knowledge the sender has about the behaviour of the SUS. Such knowledge can include the channel quality between the sender and the receiver, or the probability of a status update generation. Joint monitoring and transmission strategies are investigated in [10], [13]–[16]. In this case, the sender actively decides when to monitor the remote process and when to transmit a status update to the receiver. Moreover, to overcome the energy limitation of battery-operated senders, these works consider energy harvesting to recharge the sender's battery.

The transmission strategies proposed by the aforementioned works assume a perfect feedback channel between sender and receiver for the transmission of acknowledge (ACK) or negative acknowledge (NACK) messages. This means, the sender perfectly knows whether a transmitted status update is correctly received or not. Thus, it is able to accurately track the AoI at the receiver. However, in real applications the feedback channel is not perfect and the feedback might not be always received at the sender. This poses an additional challenge for the design of transmission strategies because the monitoring and transmission decisions have to be made under

This work has been funded by the German Research Foundation (DFG) as a part of the projects C1 and B3 within the Collaborative Research Center (CRC) 1053 - MAKI (Nr. 210487104) and has been supported by the BMBF project Open6GHub under grant 16KISKO14, by DAAD with funds from the German Federal Ministry of Education and Research (BMBF) and by the LOEWE Center emergenCity.

uncertainty about the receiver's AoI. The authors in [11] take a first step to investigate the impact of imperfect feedback channels in SUS. However, they focus on the derivation of closed-form expressions for AoI under different error models for the channel. Furthermore, they do not consider the sender's limited energy nor the energy cost associated with monitoring the remote process and the transmission of status updates.

In this paper, we investigate the design of joint monitoring and transmission strategies for SUS with an imperfect feedback channel. We model both, the channel between the sender and the receiver as well as the feedback channel as packet erasure channels. This means that status updates are correctly received at the receiver with a fixed probability. Otherwise they are lost. The same holds for the feedback channel. An ACK/NACK message is correctly transmitted back to the sender with a fixed probability. Otherwise the feedback is lost. We further assume the sender is capable of harvesting energy in order to recharge its battery. As in real applications, we assume the amounts of harvested energy and the variations of the remote monitored process are not known in advance. The main contributions of this paper can be summarized as follows

- We propose a joint monitoring and transmission strategy at the sender based on reinforcement learning to minimize the AoI at the receiver using an imperfect feedback channel. Our proposed approach, which we term Belief Learning, enables the sender to achieve near-optimal performance compared to the perfect feedback channel case, while efficiently using the available harvested energy.
- We introduce the concept of a so-called *belief distribution* to handle the uncertainty the sender has about the AoI at the receiver caused by the imperfect feedback channel. Moreover, we propose an algorithm to determine the current belief distribution based on the sender's available information.
- Through extensive numerical simulations, we show that our proposed Belief Learning approach yields a lower AoI compared to state-of-the-art transmission strategies for AoI minimization in SUS.

The rest of the paper is organized as follows. In Sec. II we introduce the system model. The AoI minimization problem is formulated as a Markov Decision Process (MDP) in Sec. III. Our proposed solution is presented in Sec. IV, followed by numerical results in Sec. V. Sec. VI concludes the paper.

## II. SYSTEM MODEL

The considered SUS is depicted in Fig. 1. It consists of a battery-operated sender, a receiver and two wireless channels connecting them, i.e., a data channel for the transmission of status updates, and a feedback channel for the transmission of ACK and NACK messages. We consider a time slotted system where a finite time horizon  $T$  is divided into time slots of equal length and indexed by  $t \in \mathbb{N}$ .

In each time slot  $t$ , the sender decides on one of four possible actions. They are formed by the combination of monitoring and transmitting, denoted by  $m_t, l_t \in \{0, 1\}$ , respectively. Every time the sender decides to monitor the

remote process, i.e.,  $m_t = 1$ , the generated status update is placed in a data buffer at the sender. The data buffer has a size of one, meaning that only the last generated status update is stored<sup>1</sup>. The resulting actions  $(m_t, l_t)$  are: monitor the remote process  $(1, 0)$ , monitor the remote process and transmit the newly generated status update  $(1, 1)$ , transmit the stored status update  $(0, 1)$ , or remain idle  $(0, 0)$ . The sender utilizes the energy stored in its battery to perform each of these actions. Monitoring the remote process requires  $\mu \in \mathbb{N}$  energy units while transmitting a status update requires  $\nu \in \mathbb{N}$  energy units. The sender's battery is assumed to have a finite capacity  $B_{\max} \in \mathbb{N}$  and we denote the current battery level as  $b_t \in \{0, 1, \dots, B_{\max}\}$ . The battery's recharging is done through an energy harvesting process modeled by the discrete random variable  $H$ .  $H$  is uniformly distributed over the set  $\mathcal{H} = \{0, 1, \dots, h_{\max}\}$  with  $h_{\max} \in \mathbb{N}$ . At the beginning of each time slot, a realization  $h_t \in \mathcal{H}$  of  $H$  denotes the number of energy units harvested in the previous time slot. The battery level  $b_t$  denotes the total number of energy units available in time slot  $t$  and is updated in each time slot as

$$b_{t+1} = \min\{B_{\max}, b_t - m_t\mu - l_t\nu + h_t\}. \quad (1)$$

The status update at the sender's data buffer is characterized by the time stamp  $\tau_S$ , which indicates the time slot of generation. At any time slot  $t$ , the freshness of the status update in the sender's data buffer is measured using the AoI  $\Delta_{S,t}$  at the sender defined as

$$\Delta_{S,t} := t - \tau_S. \quad (2)$$

For the transmission of the status updates, i.e.,  $l_t = 1$ , we model the wireless data channel between the sender and the receiver as an erasure channel. Using this model, the transmitted status update is successfully decoded at the receiver with a probability  $p_D \in (0, 1]$ , which is assumed to be known at the sender and depends on the channel quality. The receiver stores a successfully decoded status update in its data buffer. As in the sender's case, we assume the receiver's data buffer has a size of one. Similarly, we characterize the status update at the receiver's data buffer by its time stamp  $\tau_R$ . We define the AoI  $\Delta_{R,t}$  at the receiver as

$$\Delta_{R,t} := t - \tau_R. \quad (3)$$

In every time slot, the receiver provides feedback to the sender over an imperfect feedback channel. If a status update is successfully decoded, the receiver transmits an ACK. If not, a NACK is sent. We model the feedback channel as an erasure channel, i.e., the feedback message is successfully decoded at the sender with probability  $p_F \in [0, 1]$ .

Our goal is to design a monitoring and transmission strategy at the sender that minimizes the average AoI  $\bar{\Delta}_R$  at the receiver defined as

$$\bar{\Delta}_R = \frac{1}{T} \sum_{i=0}^{T-1} \Delta_{R,i}. \quad (4)$$

<sup>1</sup>Note that we aim at keeping the status updates fresh, so having a larger data buffer to store older status updates does not improve the performance.

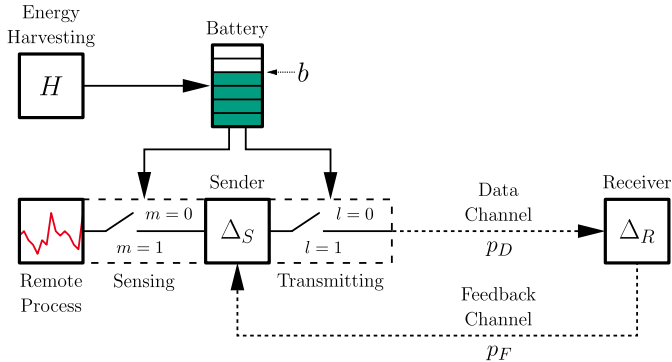


Fig. 1: Considered SUS formed by a battery-operated sender and a receiver.

### III. PROBLEM FORMULATION

The sender's monitoring and transmission strategy allows it to decide which action  $(m_t, l_t)$  to perform in each time slot. In this section, we formulate this decision-making problem as an MDP  $\mathcal{M}$ .  $\mathcal{M}$  is formed by a state space  $\mathcal{S}$ , an action space  $\mathcal{A}$ , a cost function  $c$  and a transition probability function  $P$ .

In time slot  $t$ , the state  $s_t = (\Delta_{S,t}, \Delta_{R,t}, b_t) \in \mathcal{S}$  consists of the AoI  $\Delta_{S,t}$  at the sender, the AoI  $\Delta_{R,t}$  at the receiver and the sender's battery level  $b_t$ . We consider a finite state space  $\mathcal{S} := \{0, 1, \dots, \hat{\Delta}\} \times \{0, 1, \dots, \hat{\Delta}\} \times \{0, 1, \dots, B_{\max}\}$  in which the AoI values at the sender and the receiver are limited by a maximum value  $\hat{\Delta}$ . The action space  $\mathcal{A} := \{(0, 0), (0, 1), (1, 0), (1, 1)\}$  contains the sender's possible actions  $a_t = (m_t, l_t)$ . The cost function  $c$  assigns a cost to each state transition from  $s_t$  to  $s_{t+1}$  under an action  $a_t$ . We define the cost as

$$c(s_t, a_t, s_{t+1}) = C_t := \Delta_{R,t+1}. \quad (5)$$

The transition probability function  $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$  assigns a probability to every state transition under an action  $a_t$ .  $P$  incorporates the random characteristics of the SUS, i.e.,  $p_D$  and  $H$ . Note that in real SUSs, knowledge about  $P$  is not available in advance.

A strategy  $\pi : \mathcal{S} \rightarrow \mathcal{A}$  is a solution of the MDP. It deterministically assigns an action  $a_t$  to every state  $s_t$ . Our goal is to minimize the average AoI  $\bar{\Delta}_R$  at the receiver. However, it has to be considered that decisions made in any time step  $t$  have an impact on the future system states and AoIs. To this aim, we introduce the discount factor  $\gamma$  and reformulate our goal as the minimization of the cumulative discounted cost

$$G := \sum_{k=1}^{\infty} \gamma^k C_t. \quad (6)$$

The optimal policy that minimizes  $G$  is denoted by  $\pi^*$ . The challenge in finding  $\pi^*$  comes from the fact that, since  $p_F \in (0, 1]$ , the feedback is not always correctly received. As a consequence, the sender is uncertain about the current AoI  $\Delta_{R,t}$  at the receiver and the cost  $C_t$ . Since the sender's current state  $s_t$  contains  $\Delta_{R,t}$ , the sender is also uncertain about  $s_t$ .

### IV. BELIEF LEARNING

In this section we propose a joint monitoring and transmission strategy based on reinforcement learning to find a policy that minimizes  $G$  under uncertainty about  $\Delta_{R,t}$ . Our strategy, termed Belief Learning, is based on the idea of building a *belief distribution* to track the state of the system in a probabilistic manner. In the following subsections, we formally define the belief distribution and describe how to update it based on monitoring and transmission decisions. Next, we present Belief Learning.

#### A. Belief Distribution

**Definition 1.** Let  $\mathcal{S}$  be the state space,  $\hat{\Delta} \in \mathbb{N}$  be the maximum value for the AoI and  $B_{\max} \in \mathbb{N}$  be the size of the battery. The belief distribution in time step  $t$  is then defined as an array

$$B_t \in \mathcal{B} = \mathbb{R}^{(\hat{\Delta}+1) \times (\hat{\Delta}+1) \times (B_{\max}+1)}, \quad (7)$$

with

$$B_t = (\beta_{i,j,k}^t)_{i \in \{0, \dots, \hat{\Delta}\}, j \in \{0, \dots, \hat{\Delta}\}, k \in \{0, \dots, B_{\max}\}}. \quad (8)$$

Moreover,  $B_t$  satisfies

$$\sum_{i=0}^{\hat{\Delta}} \sum_{j=0}^{\hat{\Delta}} \sum_{k=0}^{B_{\max}} \beta_{i,j,k}^t = 1 \quad (9)$$

and

$$\beta_{i,j,k}^t \in [0, 1] \forall i, j, k. \quad (10)$$

The belief distribution  $B_t$  indicates how likely it is for the system to be in state  $s_t = (\Delta_{S,t}, \Delta_{R,t}, b_t)$  in time slot  $t$ . We introduce the shorthand notation  $B_t(s_t) = \beta_{\Delta_{S,t}, \Delta_{R,t}, b_t}^t$ .  $B_t$  has one of two possible structures, a concentrated and a distributed one. The so-called concentrated belief distribution occurs when the current state  $s_t = (\Delta_{S,t}, \Delta_{R,t}, b_t)$  is known to the sender. In such cases,  $B_t$  follows

$$\beta_{i,j,k}^t = \begin{cases} 1 & \text{if } i = \Delta_{S,t}, j = \Delta_{R,t}, k = b_t \\ 0 & \text{else.} \end{cases} \quad (11)$$

When the sender is uncertain about the system state, the belief distribution has a distributed structure where all its values  $\beta_{i,j,k}^t$  are strictly less than one, i.e.,  $\beta_{i,j,k}^t < 1, \forall i, j, k$ .

#### B. Belief Distribution Update

The system's state evolves based on the harvested energy, the channel, and the actions selected. To track the state,  $B_t$  is updated in each time slot using the information available at the sender, i.e.,  $p_D, (m_t, l_t), b_t, \Delta_{S,t}$  and successfully decoded ACK/NACK feedback.

As mentioned above, the uncertainty about the system's state depends on the successful reception of an ACK/NACK from the receiver. There are three different cases of how the uncertainty about the current state evolves.

- 1) If an ACK is decoded at the sender, the sender has complete information about its state  $s_t$ , irrespective of any uncertainty about its state in previous time slots.

---

**Algorithm 1** Updating  $B_t$ 

---

```
1: if an ACK is received then
2:   update  $\Delta_{R,t+1}$  ▷ Eq. (3)
3:    $B_{t+1} \leftarrow (0, \dots, 0)$ 
4:    $\beta_{\Delta_{S,t+1}, \Delta_{R,t+1}, b_{t+1}}^{t+1} \leftarrow 1$  ▷ Concentrated belief distr., Eq. (11)
5: else
6:   if  $m_t = 1$  then ▷ The sender monitors the remote process
7:      $\beta_{0,j,k}^{t+1} \leftarrow \sum_{i=0}^{\hat{\Delta}} \beta_{i,j,k}^t, \forall j, k,$ 
8:      $\beta_{i,j,k}^{t+1} \leftarrow 0, \forall i \in \{1, \dots, \hat{\Delta}\}, j, k$ 
9:   end if
10:  if  $l_t = 1$  and no NACK is received then ▷ The sender transmits
11:     $\beta_{i,j,k}^{t+1} \leftarrow (1 - p_D) \beta_{i,j,k}^t + \mathbb{1}_{i=j} (p_D \sum_{l=0}^{\hat{\Delta}} \beta_{i,l,k}^t), \forall i, j, k,$ 
12:  end if
13:  if  $b_t \neq b_{t+1}$  then
14:     $\beta_{i,j,b_{t+1}}^{t+1} \leftarrow \beta_{i,j,b_t}^t, \forall i, j,$ 
15:     $\beta_{i,j,b_t}^{t+1} \leftarrow 0, \forall i, j,$  ▷  $b_t$  evolves to  $b_{t+1}$ 
16:  end if
17:  if  $\Delta_{S,t} \neq \Delta_{S,t+1}$  then
18:     $\beta_{\Delta_{S,t+1}, j, b_{t+1}}^{t+1} \leftarrow \beta_{\Delta_{S,t}, j, b_{t+1}}^t, \forall j$ 
19:     $\beta_{\Delta_{S,t}, j, b_{t+1}}^{t+1} \leftarrow 0, \forall j$  ▷  $\Delta_{S,t}$  evolves to  $\Delta_{S,t+1}$ 
20:  end if
21:  for  $i \in \{0, \dots, \hat{\Delta}\}, j \in \{0, \dots, \hat{\Delta} - 1\}$  and  $k \in \{0, \dots, B_{\max}\}$  do
22:     $\beta_{i,j,k}^{t+1} \leftarrow \beta_{i,j-1,k}^t$  ▷ Shift in the  $j$ -dimension
23:  end for
24:   $\beta_{i,\hat{\Delta},k}^{t+1} \leftarrow \beta_{i,\hat{\Delta},k}^t + \beta_{i,\hat{\Delta},k}^t, \forall i, k$ 
25: end if
26: return  $B_{t+1}$ 
```

---

- 2) If a NACK is decoded at the sender, the sender can deduce that the current transmission attempt was not successful and that the AoI at the receiver rises. In this case, previous uncertainty about the AoI at the receiver remains.
- 3) If the feedback is lost, the sender has no information whether the current transmission was successful. Therefore, in addition to previous uncertainty, a new layer of uncertainty is added regarding the latest transmission attempt and the resulting AoI at the receiver.

These three cases are reflected in Alg. 1, which summarizes the  $B_t$  update procedure. If an ACK is decoded at the sender,  $\Delta_{R,t+1}$  can be determined based on  $\Delta_{S,t}$ . In this case, the belief distribution takes a concentrated structure (lines 1-4). Otherwise, the belief distribution is updated using the available information, i.e.,  $p_D, (m_t, l_t), b_t$ , and  $\Delta_{S,t}$ . If the sender monitors, i.e.,  $m_t = 1$ , then  $\Delta_{S,t+1} = 1$  because the sender has a fresh status update of the remote process. In this case, the belief distribution is updated considering that only the entries  $\beta_{0,j,k}^t$  are non-zero (lines 6-9). In cases when the sender transmits the status update, i.e.,  $l_t = 1$ , the update of the belief distribution depends on the quality  $p_D$  of the data channel and the reception of a NACK. If the sender attempts to transmit and a NACK is not received, there is a probability of  $p_D$  that the transmission attempt is successful and  $\Delta_{R,t+1}$  is temporarily set to  $\Delta_{S,t+1}$ . With a probability of  $(1 - p_D)$ ,  $\Delta_{R,t+1}$  is temporarily set to  $\Delta_{R,t}$ . Taking  $p_D$

---

**Algorithm 2** Belief Learning

---

```
1: initialize  $\alpha_0, \varepsilon_0$ , discount factor  $\gamma$ 
2: initialize  $Q$  and  $B$  with zeros
3: set  $V(s) = \min_{a \in \mathcal{A}} Q(s, a), \forall s \in \mathcal{S}$  ▷ State value function
4: observe initial state  $s = s_0$ 
5: update  $B$  based on  $s_0$  ▷ Alg. 1
6: set  $\pi(s) = \arg \min_{a \in \mathcal{A}} Q(s, a) \forall s \in \mathcal{S}$ 
7: while  $t \leq T$  do
8:   select an action  $a_t = (m_t, l_t)$  ▷  $\varepsilon$ -greedy, Eq. (12)
9:   perform  $a_t$ 
10:  observe  $b_{t+1}, \Delta_{S,t+1}$  and calculate  $B_{t+1}$  ▷ Alg. 1
11:  if  $B_t$  has a concentrated structure then
12:    update  $Q$  ▷ Eq. (13)
13:    update  $V(s) \leftarrow \min_{a \in \mathcal{A}} Q(s, a), \forall s \in \mathcal{S}$ 
14:    update  $\pi(s) \leftarrow \arg \min_{a \in \mathcal{A}} Q(s, a), \forall s \in \mathcal{S}$ 
15:  end if
16:  update  $B_t \leftarrow B_{t+1}$ 
17: end while
18: return  $\pi$ 
```

---

into account, the belief distribution is updated, resulting in a distributed structure (lines 10-12). Next, considering that the battery levels  $b_t$  and  $b_{t+1}$  and the AoI values  $\Delta_{S,t}$  and  $\Delta_{S,t+1}$  are perfectly known at the sender, the distributed belief distribution is updated for these values (lines 13-20). To consider the fact that the AoI at the receiver  $\Delta_{R,t}$  increases each time slot, the values of  $\beta_{i,j,k}^t$  are shifted by one in the  $j^{\text{th}}$  dimension (lines 21-28). With all values  $\beta_{i,j,k}^t$  updated, the algorithm terminates and returns the new  $B$ .

### C. Belief Learning

In this section, we present our proposed Belief Learning approach to find a strategy  $\pi^{\text{BL}}$  that minimizes the cumulative discounted cost  $G$ . Belief Learning is based on  $\varepsilon$ -greedy  $Q$ -learning. However, and in contrast to this traditional approach, it is able to handle the uncertainty about the sender's state. To this aim, Belief Learning uses a novel modified update rule for the action value function  $Q(s, a)$  which exploits the belief distribution  $B$ . Our algorithm is summarized in Alg. 2.

As in standard  $Q$ -learning, Belief Learning selects actions that minimize  $G$  based on  $Q(s, a)$ . The values of  $Q(s, a)$  are updated according to the selected actions, the observed states, and the belief distribution  $B$ . We first initialize the learning parameters, as well as  $Q$ , and  $B$  (lines 1-2). Additionally, we initialize the state value function  $V$  (line 3). Next, we observe the initial state  $s_0$  and update  $B$  using Alg. 1 (lines 4-5). The policy  $\pi$  is initialized using the state-value function  $Q$  (line 6). In every time slot  $t$ , the action  $a_t = (m_t, l_t)$  is selected based on the policy  $\pi$  and the belief distribution  $B$  following the  $\varepsilon$ -greedy mechanism, i.e., with probability  $\varepsilon_t$  the algorithm explores by randomly selecting an action  $a_t \in \mathcal{A}$ , whilst with probability  $1 - \varepsilon_t$  the algorithm exploits the past experience by selecting the action  $a_t$  as

$$a_t = \arg \max_{a \in \mathcal{A}} \sum_{s=(\Delta_{S,t}, \Delta_{R,t}, b) \in \mathcal{S}} B_t(s) \mathbb{1}_{\pi(s)=a}. \quad (12)$$

To balance exploration and exploitation, we linearly decay  $\varepsilon_t$  over time. Using the available information at the sender, we calculate  $B_{t+1}$  as the update of  $B_t$  using Alg. 1 (line 10).

TABLE I: Simulation Parameters

Parameter	Value	Parameter	Value
$N$	100	$\widehat{\Delta}$	40
$T$	$50 \cdot 10^3$	$\alpha_t$	$0.1 + \frac{0.001-0.1}{T_{\text{learn}}}t$
$\mu$	3	$\epsilon_t$	$0.9 + \frac{0.01-0.9}{T_{\text{learn}}}t$
$\nu$	1	$\gamma$	0.9
$h_{\max}$	1	$\phi$	1
$B_{\max}$	5	$p_D$	0.4

Next, we evaluate if  $B_t$  has a concentrated structure and in that case, update  $Q$ ,  $V$  and  $\pi$  (lines 11-15). The action value function  $Q$  is updated using the belief distribution  $B_{t+1}$  as

$$Q(s_t, a_t) \leftarrow (1 - \alpha_t)Q(s_t, a_t) + \alpha_t \sum_{s' \in \mathcal{S}} B_{t+1}(s') (c(s_t, a_t, s') + \gamma V(s')), \quad (13)$$

where  $\alpha_t$  is the learning rate. Note that we do not update  $Q$ ,  $V$  and  $\pi$  when there is uncertainty about  $s_t$ , i.e.,  $B_t$  has a distributed structure. The procedure described in Alg. 2 yields  $Q$ -values which are near optimal for frequently visited states.

## V. NUMERICAL RESULTS

### A. Reference Strategies

To compare the performance of our proposed Belief Learning, we consider four strategies.

**Value Iteration:** This strategy provides the optimal monitoring and transmission strategy when perfect knowledge about  $\mathcal{M}$  is available.

**Threshold based [10]:** The sender decides to jointly monitor and transmit  $(m_t, l_t) = (1, 1)$  every time the AoI at the receiver  $\Delta_{R,t}$  exceeds an optimal threshold  $\phi$  as derived in [10]. In any other case, it idles.

**Periodic:** This strategy periodically monitors the remote process and transmit the status update  $(m_t, l_t) = (1, 1)$ . The period  $T_p$  is matched to the energy harvesting process, such that  $T_p = \left\lceil \frac{2(\mu+\nu)}{h_{\max}} \right\rceil$ .

**Random:** This strategy monitors the remote process and transmit the status update  $(m_t, l_t) = (1, 1)$  with probability  $p_R$ . As in the periodic case, we match  $p_R$  to the energy harvesting process, such that  $p_R = \frac{h_{\max}}{2(\mu+\nu)}$ .

### B. Simulation Setup

We select system parameters to resemble a general SUS, as introduced in Section I. The considered values are given in Table I. They are used unless otherwise specified. Our proposed Belief Learning is trained using  $T_{\text{learn}} = 5 \cdot 10^6$  time slots. For the evaluation, each strategy is tested for  $T = 5 \cdot 10^4$  time slots. The presented results are obtained by averaging the results of  $N = 100$  independent realizations.

The considered value iteration and threshold-based approaches require a perfect feedback channel. For a fair comparison when  $p_F < 1$ , we derive their respective policies  $\pi^{\text{VI}}$  and

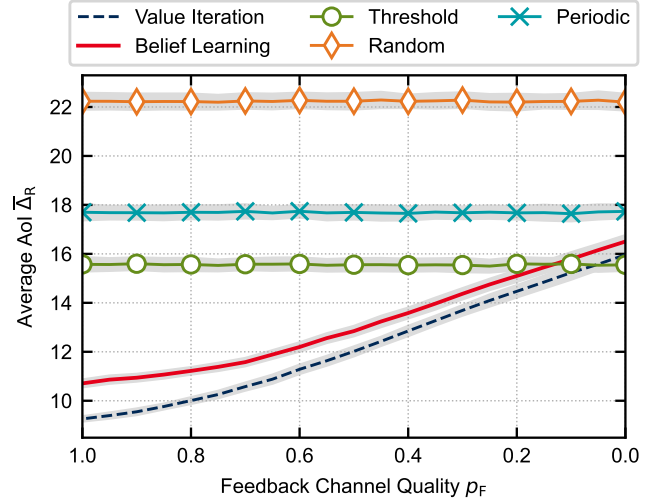


Fig. 2: Average AoI at the receiver  $\overline{\Delta}_R$  versus the feedback channel quality  $p_F$ .

$\pi^{\text{TH}}$  offline and use the selection criteria in (12) to choose the actions. This means, these approaches build their own belief distributions  $B$  based on the information available at the sender using Alg. 1. In each time slot  $t$ ,  $B_t$  is updated and the action  $a_t$  is selected based on their own policies,  $\pi^{\text{VI}}$  and  $\pi^{\text{TH}}$ , according to (12).

### C. Simulation Results

Fig. 2 shows the average AoI at the receiver  $\overline{\Delta}_R$  for different values of  $p_F$ . The gray area around each of the lines represents the standard deviation. The lowest  $\overline{\Delta}_R$  is achieved by Value Iteration by requiring perfect knowledge about the MDP  $\mathcal{M}$ . The performance of our proposed Belief Learning follows that of Value Iteration and achieves a  $\overline{\Delta}_R$  only approx. 7% higher for  $p_F = 0.5$  and without this strict requirement. Note that as the quality of the channel decreases, i.e., lower  $p_F$ , the uncertainty about the system state increases but the performance of our proposed Belief Learning gets closer to Value Iteration. For  $p_F = 0.2$ , the  $\overline{\Delta}_R$  achieved by Belief Learning is only approx. 4% higher than Value Iteration. Even though the Threshold-based approach exploits the belief distribution  $B_t$ , it has a constant behaviour because the optimal considered threshold  $\phi = 1$  results in a greedy approach that always attempts to monitor and transmit. As the Periodic and Random strategies do not consider the system state  $s_t$ , their behavior is not affected by the quality of the feedback channel. Moreover, note that all strategies which utilize a belief distribution  $B_t$ , i.e., Belief Learning, Value Iteration and Threshold-based, perform better than those which do not, i.e., Periodic and Random, even when  $p_F = 0$ . This shows that building a belief distribution of the system state improves the performance as it allows us to exploit the available feedback.

We evaluate the learning speed of our proposed Belief Learning in Fig. 3, where we show  $\overline{\Delta}_R$  versus the number of learning time steps  $T_{\text{learn}}$  for different values of  $p_F$ . In this case,

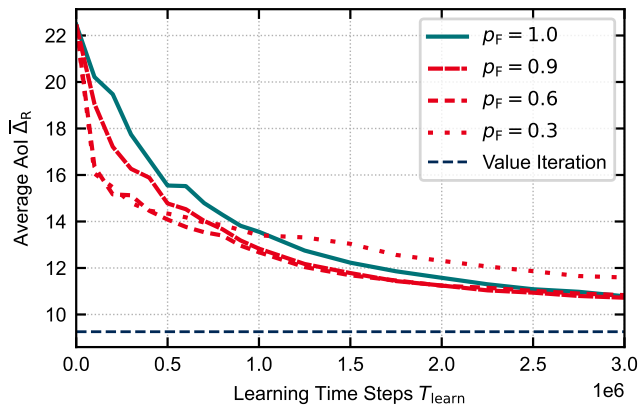


Fig. 3: Average AoI at the receiver  $\bar{\Delta}_R$  versus the amount of learned time steps  $T_{\text{learn}}$ .

we consider  $N = 300$ . For every data point, we separately run Alg. 2 with that specific  $T_{\text{learn}}$ . In this way, no bias from  $\varepsilon$ -greedy occurs. Furthermore, for a fair comparison, we test the learned strategy on a system with  $p_F = 1$ . We see that as  $T_{\text{learn}}$  increases, the algorithm converges to the same  $\bar{\Delta}_R$  regardless of the value of  $p_F$ . For  $T_{\text{learn}} < 10^6$ , the performance quickly increases for all values of  $p_F$ . It can be seen that the learning speed is larger for lower values of  $p_F$ . However, as  $T_{\text{learn}}$  increases, the learning speed decreases. This can be explained by the usage of the belief distribution in the update rule of Belief Learning (13). Consider two consecutive time slots  $t$  and  $t + 1$ . If  $B_t$  and  $B_{t+1}$  are both concentrated, (13) is equivalent to the traditional  $Q$ -learning update rule. If however,  $B_t$  is concentrated and  $B_{t+1}$  is distributed, the usage of this update rule leverages the information in  $B_{t+1}$  by taking all possible outcomes and their probability into account. This allows for a more accurate estimate of the expected cost of choosing action  $a_t$  in state  $s_t$ . When  $p_F$  is lower, the transition from a concentrated  $B_t$  to a distributed  $B_{t+1}$  occurs more often. Recall that Belief Learning only updates the  $Q$ -values, if  $B_t$  is concentrated. As we transition to a distributed  $B_{t+1}$  more often as  $p_F$  decreases, the ratio of time slots effectively learned decreases as well. This also explains the slower convergence in the case  $p_F = 0.3$ .

## VI. CONCLUSIONS

We considered a SUS in which a sender transmits status updates of a monitored process to a receiver over a wireless channel. To measure the freshness of the status update at the receiver, we considered the AoI. The optimal monitoring and transmission strategy at the sender requires knowledge about the receiver's AoI. Such knowledge can be obtained by means of a wireless feedback channel between receiver and sender.

Considering that in real applications, the feedback channel is not perfect, we investigated the design of a monitoring and transmission strategy at the sender under uncertainty about the receiver's AoI. For this purpose, we introduced the concept of a so-called Belief Distribution and proposed a monitoring and transmission strategy based on reinforcement learning, termed Belief Learning. We showed that Belief Learning allows the sender to exploit the received ACK/NACK to estimate the receiver's AoI and make informed monitoring and transmission decisions. Through numerical simulations, we showed that Belief Learning achieves near-optimal performance with respect to the perfect feedback channel case.

## REFERENCES

- [1] H. Li, X. Liu *et al.*, "Aquiculture remote monitoring system based on IoT Android platform," *Transactions of the Chinese Society of Agricultural Engineering*, 2013.
- [2] S. Abraham, J. Beard *et al.*, "Remote environmental monitoring using Internet of Things (IoT)," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, 2017.
- [3] S. Adhya, D. Saha *et al.*, "An IoT based smart solar photovoltaic remote monitoring and control unit," in *2016 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC)*, 2016.
- [4] R. D. Yates, Y. Sun *et al.*, "Age of Information: An Introduction and Survey," *IEEE Journal on Selected Areas in Communications*, 2021.
- [5] S. Kaul, M. Gruteser *et al.*, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2011.
- [6] S. Kaul, R. Yates *et al.*, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, 2012.
- [7] E. T. Ceran, D. Gündüz *et al.*, "Average Age of Information With Hybrid ARQ Under a Resource Constraint," *IEEE Transactions on Wireless Communications*, 2019.
- [8] A. Maatouk, S. Kriouile *et al.*, "The Age of Incorrect Information: A New Performance Metric for Status Updates," *IEEE/ACM Transactions on Networking*, Oct. 2020.
- [9] B. Zhou, W. Saad *et al.*, "Risk-Aware Optimization of Age of Information in the Internet of Things," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020.
- [10] W. de Sombre, F. Marques *et al.*, "A unified approach to learn transmission strategies using age-based metrics in point-to-point wireless communication," *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, 2023.
- [11] S. Rezasoltani and C. Assi, "Real-Time Status Updates in Wireless HARQ With Imperfect Feedback Channel," *IEEE Transactions on Wireless Communications*, 2022.
- [12] W. de Sombre, A. Ortiz *et al.*, "Risk-sensitive optimization and learning for minimizing age of information in point-to-point wireless communications," *IEEE International Conference on Communications (ICC)*, 2023.
- [13] S. Feng and J. Yang, "Age of Information Minimization for an Energy Harvesting Source With Updating Erasures: Without and With Feedback," *IEEE Transactions on Communications*, 2021.
- [14] E. T. Ceran, D. Gunduz *et al.*, "Learning to Minimize Age of Information over an Unreliable Channel with Energy Harvesting," *Tech. Rep.*, 2021.
- [15] F. Chiariotti, J. Holm *et al.*, "Query Age of Information: Freshness in Pull-Based Communication," *IEEE Transactions on Communications*, 2022.
- [16] O. Ozel and P. Rafiee, "Intermittent Status Updating Through Joint Scheduling of Sensing and Retransmissions," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2021.