Reinforcement Learning for Age of Information Aware Transmission Policies in Slotted ALOHA Channels

Chiara Cavalagli
Dept. of Mathematics
University of Padova, Italy
email: chiara.cavalagli@studenti.unipd.it

Leonardo Badia
Dept. of Information Engineering
University of Padova, Italy
email: leonardo.badia@unipd.it

Andrea Munari
Instit. for Communications and Navigation
German Aerospace Center (DLR), Germany
email: andrea.munari@dlr.de

***Abstract*—We focus on remote monitoring applications, in which a large number of devices send time-stamped status updates over a wireless channel to a common receiver. An uncoordinated medium sharing policy based on ALOHA is considered, and the overall goal is to maintain an up-to-date perception at the receiver, captured via the average age of information (AoI) metric. In this setting, we propose and evaluate a simple reinforcement learning algorithm which is run independently at each node in a fully decentralized fashion. Leaning on a binary success/collision feedback distributed by the receiver, the solution adapts the access behavior of transmitters based on the current value of AoI. We compare the performance of the scheme to that of threshold ALOHA [1], a benchmark protocol that resorts to a central optimization of the access parameters. Interesting insights on the potential of reinforcement learning for AoI improvements in random access channels are derived.**

I. Introduction

Age of information (AoI) has recently emerged as a key indicator to gauge the performance of communications systems that need to deliver information in a timely fashion for computation or actuation. Originally introduced in the context of vechicular networks [2], [3], the metric quantifies the time elapsed since the generation of the last received message from a source of interest, expressing how up-to-date the available knowledge is at the point of collection. Owing to its simple definition and mathematical tractability, AoI has received significant attention in the literature [4], and has been shown to effectively capture some fundamental trade-offs in a variety of settings, providing insights on the design of sampling and transmission strategies [5], as well as of more complex communications network protocols [6].

The notion of information freshness is paramount in a broad set of applications, ranging from asset tracking to distributed control, cyber-physical systems, and industrial automation. A scenario of particular relevance is that of remote source monitoring in the Internet of things (IoT) [7]. In this case, multiple devices sense physical quantities of interest, and report collected data over a shared wireless channel to a common receiver for decision making policies, which benefit from a low AoI. In such settings, grant-based transmission strategies are inefficient due to potentially unpredictable and sporadic traffic generation, together with the large number of often low-power, low-complexity sensing devices that may be deployed. As a result, random access protocols based on variations of ALOHA are employed in commercial solutions [8]. On the other hand, the uncoordinated behavior of nodes and the interference-prone multiplexing render the definition of AoI optimal access strategies non-trivial.

This spurred recent research efforts [9] underpinning some key aspects. Exceedingly large values of AoI are experienced in random access channels for low traffic (sporadic updates) as well as high transmission rates (channel clogged by collisions). In the absence of feedback, initial results [10], [11] showed an inverse proportionality between information freshness and aggregate throughput, with average AoI that scales as $e\,\mathsf{m}$ in a slotted ALOHA system with m contending terminals.

Significant improvements can be attained if a return channel is available to inform nodes about the outcome of each transmission. A simple collision/success feedback makes each terminal aware of its own AoI, enabling policies that give priority to nodes the receiver has a stale perception of. Intuitively, barring access for devices that recently delivered an update reduces channel congestion, and favors the successful communication with the receiver for nodes with higher AoI. Following this approach, different variations of slotted ALOHA that implement a *threshold policy* have been proposed.

Fundamental insights were derived in [1], assessing the performance of a scheme that silences nodes with AoI below a threshold, and lets others contend with a common transmission probability. By jointly optimizing access parameters and threshold, an average AoI that scales as $1.4169\,\mathsf{m}$ was obtained, almost halving the result in the absence of feedback, and incurring a near-negligible throughput loss. A similar solution was introduced in [12], allowing for a dynamic change of the transmission probability and reaching an asymptotic scaling law of $e\,\mathsf{m}/2$ for the average AoI. To this aim, an algorithm akin to Rivest's stabilization [13] is employed, with nodes listening to the feedback after each slot, not only after performing a transmission. Both solutions provide remarkable performance for a random access channel. However, they require knowledge of the network population, and either resort to central optimization or undergo increased complexity in terms of feedback (higher energy expenditure) and local computation. Both aspects may be problematic in large and

A. Munari acknowledges the financial support by the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification number: 16KISK022.

dynamic IoT systems with low-power devices [14].

These remarks trigger the question of whether simpler distributed solutions can be derived [15], [16]. In this paper, we propose an approach based on reinforcement learning [17], in particular, a Q-learning solution that nodes can run in a fully uncoordinated manner, leveraging only feedback received after attempting a packet delivery [18], [19]. This leads each node to choose between slotted ALOHA contention or refraining from channel access based on their current AoI [20]. The algorithm is flanked by a dynamic adaptation of the transmission probability based on the success/collision outcome of own attempts. Both procedures are simple to implement and entail low computational complexity. A threshold-based behavior emerges, providing interesting AoI performance. The initial trends we present are promising, and our work aims to further stimulate results on the application of reinforcement learning to AoI reduction in random access channels.

II. Related Works

An excellent introduction to AoI, with an overview of key results in different settings can be found in [4]. Focusing on random access, the performance of ALOHA-based contention was first explored in slotted systems in [10], [16], later tackling unslotted setups in [11]. These contributions triggered a flourishing line of research, exploring the impact of different aspects on AoI. Among these, slotted ALOHA in the presence of energy harvesting was considered in [21], whereas the role of retransmissions was tackled in [14], [22]. A game theoretic approach to the setting was discussed in [23]. As mentioned in Sec. I, the possibility to leverage feedback was thoroughly studied in [1], [12]. Further improvements were attained in [24] through an additional level of contention among nodes with age above threshold, reaching an average AoI scaling as $0.9641\,\mathsf{m}$. Information freshness in protocols that go beyond simple ALOHA was studied among others in [25] considering framed access, whereas modern random access solutions that combine packet repetitions and interference cancellation at the receiver were studied in [7], [15], [26].

Reinforcement learning has been applied to slotted ALOHA networks, mainly focusing on throughput optimization. Along this line, [18] first proposed a Q-learning strategy for framed access, later refined by [27] considering additional information exchange among nodes, whereas a multi-armed bandit approach was followed in [28]. Deep reinforcement learning techniques were targeted in [19], while an average-payoff method is presented to improve the throughput of delay-constrained ALOHA schemes in [29].

Aiming at AoI minimization, reinforcement learning based scheduling policies were proposed in [20], [30], as well as in [31], considering RF-powered devices, and [32], delving into the impact of hybrid ARQ. In random access channels, research has concentrated on deep reinforcement learning. Along this line, [33] leans on the role of urgency to develop age-aware policies, while [34] jointly optimizes the wireless energy transfer and the system AoI by combining a deep neural network and tabular Q-learning. Moreover, state of the art deep-Q networks have been proposed in [35] with the integration of interference cancellation.

III. System Model

We consider a population of m users (nodes), sharing a wireless channel towards a common receiver (sink). Time is slotted, and the parameters of the system are set such that a data packet fits one slot. All nodes are assumed to be slot synchronized. Medium contention follows a slotted ALOHA policy i.e., in each slot, users independently decide whether to access the channel with probability τ, or to remain silent. This probability may differ among nodes, and change over time. In case of a transmit decision, the user sends over the slot a time-stamped message addressed to the receiver. Following a generate-at-will setting [1], [10], we assume fresh information to be always available for transmission, so that the time stamp of any message is set to the start of the slot it is sent over.

We capture the effect of interference triggered by the uncoordinated access policy via the well-established collision channel model [23]. Specifically, the transmission of two or more packets in the same slot impedes decoding at the receiver, whereas the message of a sole user accessing a slot is successfully retrieved. At the end of each slot, the sink broadcasts a binary feedback, informing all nodes of whether a collision occurred or a packet was decoded. The feedback is modeled as instantaneous and error-free.

Throughout our discussion, we are interested in gauging the ability of the system to maintain an up-to-date knowledge at the sink. To this aim, let $\delta^{(i)}(n)$ denote the current age of information for node i at slot n, i.e.

$$\delta^{(i)}(n) := n - \sigma^{(i)}(n) \tag{1}$$

where $\sigma^{(i)}(n)$ is the time stamp of the last message the receiver collected from node i as of time n. The metric grows linearly over time until the reception of a new message from user i resets it to one slot (i.e., the time needed for the message to be transmitted and decoded by the sink). An example of the time evolution of $\delta^{(i)}(n)$ is reported in Fig. 1. This slot-wise feedback renders each node aware of its own current AoI value. Leaning on (1), we will focus on the average AoI for a node up to time n, defined as

$$\Delta^{(i)}(n) := \frac{1}{n}\sum_{\ell=1}^{n}\delta^{(i)}(\ell). \tag{2}$$

As $n \to \infty$, the value converges to the common definition of average AoI [4]. In addition, we will consider the normalized average network AoI at time n, computed as

$$\bar{\Delta}(n) = \frac{1}{\mathsf{m}}\sum_{i=1}^{\mathsf{m}}\Delta^{(i)}(n).$$

IV. AoI-Based Q-Learning Slotted ALOHA

For the setting under study, we are interested in decentralized policies based on slotted ALOHA that are able to improve the AoI performance [10], [15]. Specifically, we

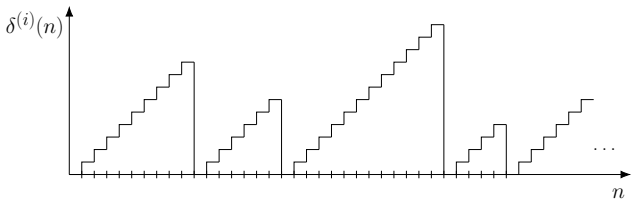

Fig. 1. Example of time evolution for the current AoI of a node.

target distributed algorithms that allow nodes to dynamically tune their channel access solely based on local knowledge gathered through the binary feedback provided by the sink. An optimal solution to the problem is in general elusive, due to the presence of multiple non-coordinated agents contending for the same resource, and in view of the potentially time-correlated interference structure experienced in the system.

To tackle the problem, we thus resort to a reinforcement learning approach, more precisely Q-learning [17] in view of its model-free, off-policy and online control properties. Indeed, in the setup at hand, each agent is unaware of some global channel features such as the number of users or the AoI of other nodes (*model-free*), and shall learn a strategy directly from raw experience (*online, off-policy*). Formally, we define the state of agent i at time $n \in \mathbb{N}$ as its current AoI, capped to a maximum value Θ:

$$s_n^{(i)} := \begin{cases} \delta^{(i)}(n) & \delta^{(i)}(n) \leq \Theta \\ \Theta & \text{otherwise} \end{cases}$$

The truncation of the state space as $\mathcal{S} = \{1, \ldots, \Theta\}$ aims at reducing the computational complexity and speeding up the convergence of the algorithm, with the value of Θ chosen such that nodes experience high AoI values only sporadically. In the remainder, we shall drop the node superscript i when no confusion arises. At each slot, an agent selects an action from space $\mathcal{A} = \{\mathtt{w}, \mathtt{t}\}$. In the former case (`w`), the node remains idle and does not access the channel. When `t` is selected, the agent performs a slotted ALOHA contention to attempt a packet delivery. The outcome of executing $a \in \mathcal{A}$ when in state $s \in \mathcal{S}$ leads to a reward R, which will be presented shortly. Based on this, the algorithm looks for the best policy $\pi(a \,|\, s)$, determining the action to be chosen in each state, through the optimization of the Q-value function $Q(s, a)$ as described for completeness in the pseudo-code at the top of next column.

The procedure is executed independently at each agent. In terms of notation, ε denotes the *exploration rate*, γ the *discount factor*, and α the *learning rate*.

By construction, the algorithm returns a *deterministic* policy, as an agent will consistently choose the action returning the maximum Q value in the current state. We cast this into a slotted ALOHA setting by having each node apply a contention probability whenever action `t` is selected. Specifically, the node actually attempts packet delivery with probability τ, and refrains from accessing the channel otherwise. To complete the proposed approach thus, two aspects have to be specified:

Algorithm 1 Q-learning

Initialize the Q values and the hyper-parameters
- $Q(s, a) = 0$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}$, for every node i
- $\alpha, \gamma \in [0, 1]$, $\varepsilon > 0$

for each slot n **do**
  Given current state s_n, choose action a acting ε-greedy:
  $x \leftarrow$ uniform random number in $[0, 1]$
  if $x \leq \varepsilon$ **then**
    $a \leftarrow$ random action from the state space
  else
    $a \leftarrow \arg\max_{b \in \mathcal{A}(S)} Q(s_n, b)$
  end if
  Take action a, observe s_{n+1}, retrieve R
  Update Q value:
    $m \leftarrow \gamma \max_{a' \in \mathcal{A}} Q(s_{n+1}, a')$
    $Q(s_n, a) \leftarrow Q(s_n, a) + \alpha(R + m - Q(s_n, a))$
end for

the definition of the reward and how to set the value of τ. As to the former, we consider the following strategy:

- a choice to wait provides a reward which solely depends on the current AoI of the agent, regardless of the behavior of other nodes:
$$R(\mathtt{w}) = 1 - \frac{\delta(n)}{\Delta(n)} \tag{3}$$
where we recall that $\Delta(n)$ is the average AoI of the agent up to time n, defined in (2);
- conversely, when the agent selects action `t`, the reward depends on the {success,collision} outcome observed over the current slot, and notified by the receiver via the feedback channel:
$$R(\mathtt{t}) = \begin{cases} \dfrac{\delta(n)}{\Delta(n)} - 1 & \text{if } \textit{success} \\ -1 & \text{if } \textit{collision} \\ 0 & \text{if agent remains } \textit{silent} \end{cases} \tag{4}$$
Here, the last case denotes the reward when the node refrains from accessing the channel (probability $1 - \tau$).

Aiming at a low average AoI for the node, a transmission that leads to successfully delivering an update to the sink is rewarded proportionally to how much the AoI is lowered. This discourages too quick attempts after a reset, which would lead to limited benefit, and incentivizes the agent to become more aggressive when experiencing higher values of AoI. Similarly, a wait action is first positively reinforced (low $\delta(n)$) – preventing an agent from seizing the channel and potentially allowing other nodes in worse condition to communicate –, and later negatively reinforced (high $\delta(n)$). In turn, the introduction of a penalization in case of experiencing a collision mixes the single-agent and the aggregate network objectives in order to promote distributed cooperation. Finally, a neutral reward is foreseen when the node refrains from sending a message after having chosen `t`, irrespectively of the AoI level, as such a

TABLE I
PARAMETERS USED FOR AOI-Q-ALOHA

Parameter	Value
learning rate, α	0.1
discount factor, γ	0.1
exploration rate, ε	0.05
maximum AoI value, Θ	600
adaptation steps for τ	$\omega_s = 0.005$, $\omega_c = 0.005$

behavior is an integral part of the slotted ALOHA contention.

As to the transmission probability, we propose a simple reinforcement approach, which runs in parallel to Q-learning.[1] The value of τ is randomly initialized by each node, and iteratively adjusted over time. After each successful transmission the agent increases τ by a term ω_s, whereas the probability is decreased by ω_c in case a collision is experienced. The presented approach is simple, and can be run by each node only leaning on the feedback received after a transmission. Rewards solely depend on the local AoI (current at time n and average until time n), and neither prior knowledge on the network population, nor centralized optimization of the access parameters or the policy is required.

V. RESULTS AND DISCUSSION

To evaluate the proposed solution, we start focusing on a network of $\mathsf{m} = 100$ nodes. In the remainder, we refer to the presented algorithm as AoI-Q-ALOHA. As a benchmark for performance comparison, we consider the threshold ALOHA policy [1], which resorts to a centralized optimization of both threshold and channel access probability. The parameters employed in our study are reported in Tab. I.

A first interesting result is shown in Fig. 2, reporting the normalized network AoI evolution over time, i.e. $\bar{\Delta}(n)$. The average normalized AoI achieved by threshold ALOHA was obtained by means of dedicated simulations, so as to identify the best threshold and transmission probability pair.[2] For comparison, the performance of a slotted ALOHA system without feedback is also shown. In this case, the transmission probability is set to $1/\mathsf{m}$, obtaining the minimum network AoI $e\,\mathsf{m}$. First, focus on the trend of AoI-Q-ALOHA.

As highlighted by the plot, the Q-learning algorithm reaches convergence rather quickly, allowing nodes to enjoy a relatively low level of AoI in a short time. Furthermore, $\bar{\Delta}$ reaches a value of ~ 1.7, improving by over 40% over the basic slotted ALOHA, and remaining 14% shy of threshold ALOHA. From this standpoint, it is important to remark that the benchmark enforces an optimized transmission policy to all terminals, whereas our uncoordinated approach leads each node to independently adapting its access policy in a fully distributed manner. As such, slightly different trends emerges

[1]This choice allows to reduce the space state and speed-up convergence of the Q-learning algorithm, compared to implementing a joint reinforcement of both the action and channel access probability.

[2]We note that the result is slightly higher than the scaling law $1.4159\,\mathsf{m}$, as the latter only holds for an asymptotically large number of nodes [1].

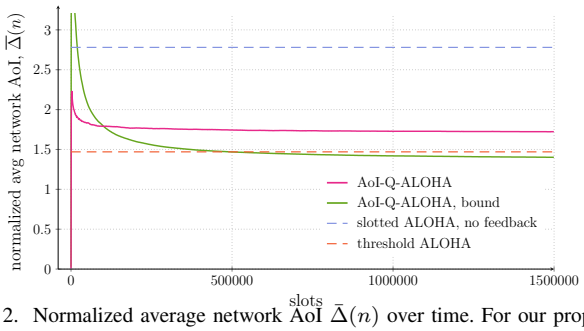

Fig. 2. Normalized average network AoI $\bar{\Delta}(n)$ over time. For our proposed AoI-Q-ALOHA, an exact evaluation is shown, together with a bound for the scheme (discussed later in the section).

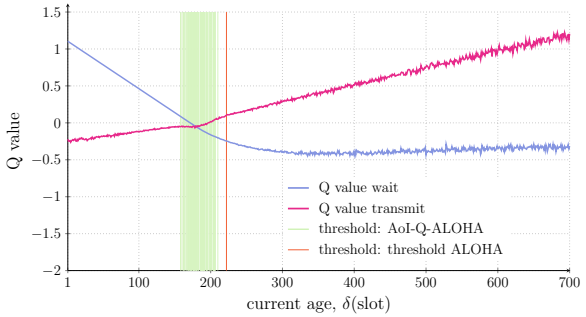

Fig. 3. Average over all nodes of the $Q(\cdot)$ values for the actions, vs the current AoI. The vertical line is the optimal threshold value of threshold ALOHA. Shaded green ranges denote the first value of AoI for which transmitting becomes convenient over waiting for each of the nodes.

in the network, causing suboptimal performance in terms of AoI. Nonetheless, AoI-Q-ALOHA grants a good level of fairness, as Jain's index for AoI among nodes reaches ~ 0.99 at convergence. It is worth noting that the aggregate throughput stabilizes around 0.354 [pkt/slot], with a small degradation compared to the benchmarks.

Further insights on the algorithm are provided by Fig. 3. Here, the average of the Q values seen by the nodes for the transmit and wait actions are reported for each possible value of the state (current AoI). Formally, we show, for each state s, $\frac{1}{\mathsf{m}}\sum_{i=1}^{\mathsf{m}} Q_i(s_n, a)$ for $a \in \{\mathtt{w}, \mathtt{t}\}$ and for n large enough to have reached stable AoI performance (see Fig. 2). A threshold behavior emerges, favored by the definition of the rewards introduced in (3)-(4). Notably, the crossing point for the value of $\mathtt{t}$ and $\mathtt{w}$ actions happens *on average* for a slightly lower AoI value compared to the optimal threshold identified by threshold ALOHA (pink vertical line). For completeness, the plot also reports (green vertical lines) the lowest AoI value at which each of the nodes in the network prefers accessing the channel over waiting. In this perspective, it is important to recall that the learning approach correctly reinforces the idea to refrain from access when the current value of δ is low, and to become more aggressive as AoI grows. However, the algorithm implemented by each node does not converge to a sharp threshold, due to the randomness in the history experienced at the agent. Rather, a transient region emerges for growing AoI, as transmission becomes progressively dominant

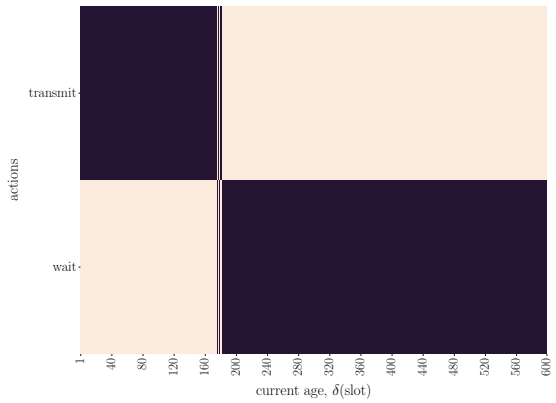

Fig. 4. Example of Q-matrix at convergence for a node in the network. For each current AoI value, on the y-axis it shows whether the wait and transmit action is convenient (white cell, i.e., higher Q value).

over waiting. This is exemplified in Fig. 4, which depicts the Q matrix of one of the agents in the system. On the x-axis, the current AoI value is reported, and, for each δ, a white cell on either the transmit or wait cell indicate a larger Q value for the corresponding action.

Further, Fig. 5 shows the evolution over time of the transmission probability. Also in this case, for each time instant we report the average value of τ among all nodes. The result prompts two remarks. First, we see again a quick rise of the parameter, as nodes dynamically adapt their access probability to the level of contention experienced, in turn affected by the number of nodes experiencing an AoI high enough to prefer `t` over `w` as action of choice. Secondly, it can be observed that the average τ tends to stabilize to a value that is lower than the optimal access probability of threshold ALOHA (red dashed line). This behavior is consistent with the fact that in AoI-Q-ALOHA nodes tend to switch to transmission for lower values of AoI. Accordingly, a stronger level of contention is likely to be experienced, leading to a less aggressive access.

To further delve into the impact of nodes estimating τ in a fully distributed manner, we considered a (non-practical) variation of the scheme, in which each terminal is aware of the current AoI values of other nodes. Upon choosing a transmit action, the agent estimates the number of contenders it might expect over the current slot, assuming that any other node with an AoI value that would lead the agent to transmit (checking its own Q matrix) does so. The transmission probability is in this case chosen as the inverse of the estimated contention level. The decision is made slot by slot. We remark that this solution is indeed simply a useful benchmark, as distributing the AoI of all nodes would not be feasible in large, slotted ALOHA based systems. The performance obtained in this case in terms of normalized average network AoI is shown by the green solid line in Fig. 2. Interestingly, a level of information freshness even lower than the one of threshold ALOHA is attained. The result buttresses the potential of even a simple and distributed Q-learning approach, and calls for further studies of more advanced learning strategies applied

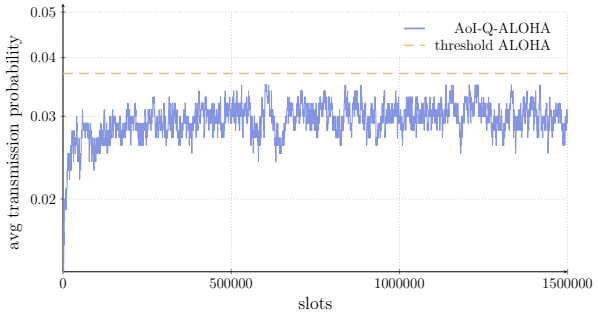

Fig. 5. Evolution over time of the transmission probability τ, averaged over all nodes. The optimal transmission probability for slotted ALOHA is also shown for comparison.

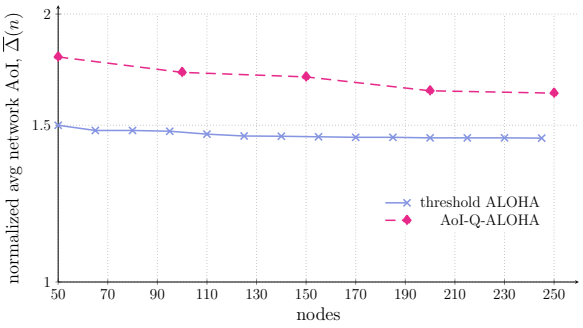

Fig. 6. Normalized average network AoI vs number of nodes in the network.

to the setting under investigation.

In Fig. 6, we report the normalized average AoI obtained by AoI-Q-ALOHA at convergence against the number of nodes in the network. As benchmark we once again report the performance of threshold ALOHA, for which transmission probability and threshold were optimized for each population size by means of dedicated simulations. The plot confirms that the proposed algorithm offers a consistent behavior, and interestingly points out that the gap with the benchmark reduces for larger networks. This hints at a less relevant role played by the behavior of nodes that act with AoI thresholds which are far apart from the average.

We conclude our study by studying the performance of AoI-Q-ALOHA in networks that experience a dynamic change in the number of nodes. The results are shown in Fig. 7, reporting $\bar{\Delta}(n)$ over time. Specifically, we started by operating a network with 100 terminals. After 10^5 slots (leftmost vertical dashed line), additional 50 nodes join the system. No knowledge about this event is distributed among agents, and the newly inserted nodes operate themselves employing AoI-Q-ALOHA without knowing the network cardinality. Right after the change, a sharp decrese in the average AoI is experienced. This is due to the fact that the new agents start by convention with an AoI of 0, thus creating a bias on the average. More interestingly, as time goes by, the system quickly converges again to a stable solution, which performs as it would have if all nodes had been present from the start. At time $n = 10^6$ (rightmost vertical dashed line), we again remove 50 nodes,

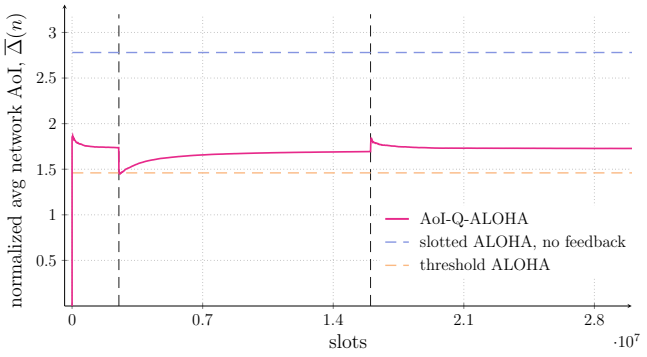

Fig. 7. Evolution over time of $\bar{\Delta}(n)$ for a dynamically changing network population between 100 and 150 nodes.

reverting to the original network population. Also in this case the scheme proves to be robust, as the Q-learning leads nodes to adapt their behavior, achieving the expected performance. These results are quite interesting, as they prove the ability of the approach to seamlessly adapt to dynamic network topologies, which might be encountered in many applications. We remark once more that the algorithm only leverages the feedback each node receives after transmission, and does not require any other form of knowledge on the status of the network. On the other hand, Fig. 7 also highlights that a non-negligible time is required for the considered setting in order for the AoI to converge after a change in the cardinality. This is especially true when the number of nodes is reduced (rightmost part of the plot), and is related to the fact that, in order for the Q-learning approach to update its rewards after having stabilized, a good deal of exploration is required. From this perspective, further studies focusing on the role played by the exploration rate are of particular interest, and will be part of our future works.

REFERENCES

[1] O. Yavaskan and E. Uysal, "Analysis of slotted ALOHA with an age threshold," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, May 2021.

[2] S. Kaul, R. Yates, and M. Gruteser, "On piggybacking in vehicular networks," in *Proc. IEEE GLOBECOM*, Dec 2011.

[3] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE SECON*, June 2011.

[4] R. Yates, Y. Sun, D. Brown, S. 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, May 2021.

[5] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the Wiener process for remote estimation over a channel with random delay," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, Feb. 2020.

[6] T. Shreedhar, S. Kaul, and R. Yates, "ACP+: An age control protocol for the internet," *IEEE/ACM Trans. Netw.*, 2024.

[7] A. Munari, "Modern random access: an age of information perspective on irregular repetition slotted ALOHA," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3572–3585, 2021.

[8] LoRa Alliance, "The LoRa Alliance Wide Area Networks for Internet of Things," www.lora-alliance.org.

[9] A. Munari and E. Uysal, "Information freshness in random access channels for IoT systems," in *Proc. IEEE BalkanCom*, 2021.

[10] R. Yates and S. Kaul, "Status updates over unreliable multiaccess channels," in *Proc. IEEE ISIT*, 2017.

[11] R. Yates and S. K. Kaul, "Age of information in uncoordinated unslotted updating," in *Proc. IEEE ISIT*, 2020.

[12] X. Chen, K. Gatsis, H. Hassani, and S. Bidokhti, "Age of information in random access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 10, pp. 6548–6568, 2022.

[13] R. Rivest, "On self-organizing sequential search heuristics," *Commun. ACM*, vol. 19, no. 2, pp. 63–67, 1976.

[14] A. Munari, "On the value of retransmissions for age of information in random access networks without feedback," in *Proc. IEEE GLOBECOM*, 2022.

[15] L. Badia and A. Munari, "A game theoretic approach to age of information in modern random access systems," in *Proc. IEEE GLOBECOM Wkshps*, 2021.

[16] R. Talak, S. Karaman, and E. Modiano, "Distributed scheduling algorithms for optimizing information freshness in wireless networks," in *Proc. IEEE SPAWC*, June 2018.

[17] R. Sutton and A. Barto, *Reinforcement Learning: an Introduction*, 2nd ed. Cambridge, MA (US): MIT Press, 2018.

[18] Y. Chu, P. Mitchell, and D. Grace, "ALOHA and Q-learning based medium access control for wireless sensor networks," in *Proc. IEEE ISWCS*, 2012.

[19] M. Jadoon, A. Pastore, M. Navarro, and F. Perez-Cruz, "Deep reinforcement learning for random access in machine-type communication," in *Proc. IEEE WCNC*, 2022.

[20] Y.-P. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE ISIT*, 2017.

[21] K.-H. Ngo, G. Durisi, A. G. Amat, A. Munari, and F. Lázaro, "Age of information in slotted ALOHA with energy harvesting," in *Proc. IEEE GLOBECOM*, 2023.

[22] P. Dester, P. Nardelli, F. dos Santos Filho, P. Cardieri, and P. Popovski, "Delay and peak-age-of-information of ALOHA networks with limited retransmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2328–2332, 2021.

[23] L. Badia, A. Zanella, and M. Zorzi, "A game of ages for slotted ALOHA with capture," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4878–4889, 2024.

[24] M. Ahmetoglu, O. Yavascan, and E. Uysal, "MiSTA: An age-optimized slotted ALOHA protocol," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15 484–15 496, 2022.

[25] J. Wang, J. Yu, X. Chen, L. Chen, C. Qiu, and J. An, "Age of information for frame slotted ALOHA," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2121–2135, 2023.

[26] A. Munari, F. Lazaro, G. Durisi, and G. Liva, "The dynamic behavior of frameless ALOHA: Drift analysis, throughput, and age of information," *IEEE Trans. Commun.*, vol. 71, no. 12, pp. 6914–6927, 2023.

[27] M. Zhang, L. de Alfaro, and J. Garcia-Luna-Aceves, "Making slotted ALOHA efficient and fair using reinforcement learning," *Comput. Commun.*, vol. 181, pp. 58–68, 2022.

[28] N. Peng and L. Dai, "Multi-armed-bandit-based framed slotted ALOHA for throughput optimization," *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 847–851, 2024.

[29] L. Deng, D. Wu, J. Deng, P.-N. Chen, and Y. Han, "The story of $1/e$: ALOHA-based and reinforcement-learning-based random access for delay-constrained communications," 2022. [Online]. Available: arxiv.org/abs/2206.09779

[30] S. Leng and A. Yener, "Age of information minimization for wireless ad hoc networks: A deep reinforcement learning approach," in *Proc. IEEE GLOBECOM*, 2019.

[31] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in RF-powered communication systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4747–4760, 2020.

[32] E. Ceran, D. Gündüz, and A. György, "A reinforcement learning approach to age of information in multi-user networks with HARQ," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1412–1426, 2021.

[33] M. Jeong, G. Seo, and E. Hwang, "Age of information optimization by deep reinforcement learning for random access in machine type communication," in *in Proc. IEEE Big Data*, 2022.

[34] M. Abd-Elmagid, H. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in RF-powered communication systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4747–4760, 2020.

[35] H. Zhao, H. Yu, Z. Zhang, M. Zeng, and Z. Fei, "Deep reinforcement learning for the joint AoI and throughput optimization of the random access system," in *in Proc. IEEE WCSP*, 2022.