Regret of Age of Information Bandits for Single and Multiple Sources under Non-stationary Channels

Xiayi Wang, Jianan Zhang, Xiang Cheng, and Yuping Zhao
School of Electronics, Peking University, China

***Abstract*—We study scheduling algorithms to minimize the Age of Information (AoI) regrets under non-stationary channels in a drifting environment, where the probability that a channel can successfully transmit a packet varies over time and the total variation of probabilities across adjacent time slots is bounded. We characterize the AoI regret lower bound for single-source and multi-source systems, and prove that a restarted channel scheduling algorithm achieves an AoI regret within a logarithmic factor gap from the lower bound for a single-source system. We augment the above channel scheduling algorithm with a source scheduling algorithm to support multiple source transmissions where sources are decoupled from channels. In addition, we develop a scheduling algorithm for multiple source transmissions where sources and channels are coupled, and characterize the AoI regret upper bounds.**

***Index Terms*—Scheduling, Age of Information, reinforcement learning, non-stationary channel.**

I. Introduction

With the emergence of time-sensitive applications in autonomous driving and multi-agent systems, real-time information acquisition is crucial to sensing and control. Age of Information (AoI) has been widely studied as a metric for information freshness. Defined as the minimum time elapsed since the generation of the received data packets, AoI captures both transmission delays and data generation intervals. By maintaining a low AoI, systems ensure timely decision-making and responsiveness, thereby enhancing their overall performance and reliability. Current research on AoI spans various domains, encompassing queuing and analysis, sampling and remote estimation, wireless scheduling, and networked control.

Within the wireless scheduling research field, there is a significant emphasis on optimizing scheduling strategies to minimize AoI, thereby ensuring timely data transmission and responsiveness. Existing scheduling efforts focus primarily on designing algorithms under different assumptions for sources, channels, and destinations. For instance, sources of information can involve deterministic or stochastic arrivals, channels can exhibit reliable or unreliable transmission characteristics, and destinations can range from a single node to multiple nodes. Scheduling algorithms for the multi-source single-destination model were developed and their AoI performance was analyzed in [1]. In the presence of wireless interference, a weighted source-destination pair scheduling was studied in [2]. The sources were modeled as interrelated types, and an optimal random scheduling strategy was derived by solving a convex optimization problem in [3]. A Kalman filter-based optimal estimation framework was utilized to design scheduling policies in [4]. Scheduling for heterogeneous multi-source systems was studied, and a generalized round-robin scheduling policy was proposed in [5]. A lightweight scheduler based on an AoI function built with the tight scalar upper bound of the remote estimation error was proposed in [6]. Scheduling algorithms to minimize the average AoI in a specific multi-sensor multi-server Internet-of-Things system was discussed in [7]. The optimization problem to minimize the average AoI while satisfying the timely throughput constraints was formulated in [8]. The diversity in model assumptions has led to the development of numerous scheduling algorithms tailored to the specific scenarios.

To address the challenge of unknown channels' service rates in real-world scenarios where environmental information is not always fully available, recent research has introduced learning-based algorithms that model AoI scheduling as a multi-armed bandit problem. For example, in [9], each channel was modeled as an "arm", and each decision involved pulling an arm to receive a reward. The scheduler estimated the service rates of different channels based on past rewards and selected the channel with the highest estimated value at each time slot. To measure the performance of scheduling algorithms in unknown environments, a common approach is to compare with the AoI value corresponding to the optimal strategy. The difference between the AoI obtained under the current scheduling strategy and the optimal scheduling was defined as *AoI regret*, which quantified the performance loss due to the learning algorithm. In [9], a lower bound on AoI regret was provided for single-source systems, and it was shown that using Upper Confidence Bound (UCB) and Thompson Sampling can achieve order-optimal performance. Building on this work, in [10], the single-source setting was extended to multi-source setting, demonstrating that the upper and lower bounds of AoI regret are proportional to the number of times when suboptimal channels were selected over time T. An order-optimal AoI regret was achieved by a decentralized round-robin algorithm in [11]. However, all these studies assumed that channels' service rates were stationary, learning a time-invariant unknown quantity for scheduling.

In reality, channels' service rates are time-varying, and far less previous research has considered non-stationary chan-

This work was supported in part by the National Natural Science Foundation of China under Grants 62341101, 62301011 and 62125101; and in part by the New Cornerstone Science Foundation through the Xplorer Prize.

nels. AoI-oriented scheduling under non-stationary channels in *switching environments* was studied in [12]. The main characteristic of a switching environment is that environmental changes occur abruptly at decision epochs, with the severity of channel state changes described by constraining the total number of changes. Upper and lower bounds on AoI regret under three non-stationary models for a single-source system were provided in [12]. Another line of work investigated non-stationary network control under *drifting environments* [13] [14]. Unlike switching environments, drifting environments do not constrain the number of changes but instead limit the total variation of service rates over time T. This assumption aligns better with the patterns of real-world environmental changes, such as the gradual and continuous variations in channel service rates over time, rather than limiting the total number of changes. Queue scheduling algorithms were examined under various total variation assumptions in [15], while a new Max Weight-UCB algorithm that employed the maximum weight strategy and sliding window confidence for generalized wireless network scheduling was introduced in [16]. However, both [15] and [16] focused on throughput optimization, without addressing scheduling in drifting environments from a timeliness perspective.

We study the AoI regret minimization problems in both single-source and multi-source systems that attempt to communicate through a common, limited, and unknown-rate pool of channels. We consider the scheduling of both sources and channels while allowing for continuous variations in channels' service rates. Specifically, our contributions are as follows.

- We develop an AoI regret lower bound for single-source and multi-source transmission models under non-stationary channels in the drifting environment.
- For the single-source transmission model, we apply the REXP3 algorithm [17] as a restarted channel scheduling algorithm to minimize AoI and prove that the algorithm achieves an AoI regret upper bound that is within a logarithmic factor from the lower bound.
- For the multi-source transmission models, we study the problem with decoupled or coupled sources and channels. We propose the Max Age REXP3 and Max Weight Age UCB scheduling algorithms for these scenarios, respectively, and prove AoI regret upper bounds of the algorithms.

The rest of the paper is organized as follows. We introduce the model in Section II. In Section III, we study single-source systems, develop an AoI regret lower bound, and study the performance of an algorithm achieving AoI regret upper bound that is within a logarithmic factor from the lower bound. In Section IV, we study multi-source systems, develop algorithms, and prove their AoI regret upper bounds. Section V presents numerical simulations, and Section VI concludes the paper.

II. System Model

In this section, we present models for the transmission source and channel, and formulate the AoI regret minimization problems.

A. Source Model

We adopt a discrete time-slotted system. At the beginning of each time slot, a source generates a new data packet. To ensure the freshness of the transmitted information, only the most recently generated data packet is retained in the source queue, and outdated packets are discarded according to the Last Come First Served queuing discipline. Depending on the number of sources, the model can be categorized into single-source and multi-source systems. In a single-source system, only channel scheduling needs to be considered, whereas in a multi-source system, both source and channel scheduling needs to be addressed.

B. Drifting Channel Model

Information from the sources needs to be transmitted to the destination through unreliable channels. We use a binary ON/OFF model to represent the current state of a channel. When a channel is ON, choosing the channel can successfully transmit information at the current time. Conversely, when a channel is OFF, the channel is of poor quality and choosing this channel would result in a transmission failure. We use $\mu_k(t)$ to represent the probability that channel k is ON at time t, through which a data packet can be successfully transmitted. The probability $\mu_k(t)$ is referred to as the *service rate* of the channel k at time t. Different channels have different service rates, and the rates vary over time even for the same channel.

Channels' service rates are unknown and can only be estimated using past scheduling decisions and the history of transmission successes and failures. Consider non-stationary channels in a drifting environment, where service rates change adhering to a given total variation budget within a time period T.

Assumption 1. (Drifting Dynamic). The total variation over T is upper-bounded by the variation budget:

$$\sum_{t=1}^{T-1} \sup_{k \in K} |\mu_k(t+1) - \mu_k(t)| \leq V_T,$$

where $V_T = \frac{1}{K} T^{\alpha}$ with $\alpha \in (0, 1)$.

Such a non-stationary channel model has applications in various contexts. For instance, in wireless communication systems, it describes fluctuations in channel quality, aiding the design of scheduling and resource allocation algorithms. In traffic management, it models dynamic changes in traffic flow to optimize signal control and route planning. In network security, this assumption helps model variations in network traffic to enhance defense capabilities.

C. Coupling Relationship for Multiple Source Transmission

In a multi-source system, both source scheduling and channel scheduling need to be considered. The sources and channels can be either coupled or decoupled. If they are coupled, selecting one source uniquely determines the corresponding channel. If they are decoupled, the selections of sources and channels are independent.

Taking the vehicular network scenario as an example, the coupled model corresponds to the scenario in which multiple distant vehicles transmit information to a single designated road side unit. In this case, selecting the vehicle for transmission (i.e., source) determines the transmission channel. On the other hand, the decoupled model corresponds to a scenario where a single vehicle is equipped with multiple types of sensors and the sensory data are transmitted among adjacent vehicles. It is possible to independently schedule a specific type of sensory data and select a channel to transmit data.

D. Performance Metrics

The AoI is a metric that measures the freshness of received information. It is defined as: $h(t) = t - g(t)$, where t represents the current time, and $g(t)$ denotes the generation time of the most recently received data packet up to the current time.

For a single-source system, let $h^\pi(t)$ denote the destination AoI under the current policy π. The optimization objective is to minimize the sum of the AoI over time T: $\sum_{t=1}^{T} h^\pi(t)$. For a multi-source system, let $h_m^\pi(t)$ denote the destination AoI of source m under the current policy π. The optimization objective is to minimize the sum of the AoI for all sources over time T: $\sum_{m=1}^{M}\sum_{t=1}^{T} h_m^\pi(t)$.

Suppose that an oracle knows channels' service rates at each decision epoch and can make optimal scheduling decisions to minimize AoI $h^*(t)$. We define the regret metric to measure the difference in AoI between strategy π and the optimal strategy. The regret for a single-source system is

$$R^\pi(T) = \mathbb{E}\left[\sum_{t=1}^{T} h^\pi(t) - \sum_{t=1}^{T} h^*(t)\right]. \tag{1}$$

The regret for a multi-source system is

$$R^\pi(T) = \mathbb{E}\left[\sum_{m=1}^{M}\left(\sum_{t=1}^{T} h_m^\pi(t) - \sum_{t=1}^{T} h_m^*(t)\right)\right]. \tag{2}$$

A smaller AoI regret implies that the information reaching the destination is fresher, which aligns with the optimization objective of minimizing AoI.

E. Problem Formulation

We consider a centralized scheduling problem for single-source and multi-source systems operating over non-stationary channels. For the scheduler, the number of sources and the coupling relationship between the current sources and channels are known, while the channels' service rates at each time slot remain unknown. The scheduler must learn the channel service rates based on past scheduling decisions and their corresponding transmission success or failure outcomes, aiming to output current scheduling decisions that minimize the system's AoI regret.

III. Channel Scheduling for Single-source Systems

Single-source systems only require the decision for channel scheduling. To investigate the properties of AoI regret in the drifting environment under Assumption 1, we first establish a lower bound for AoI regret. Subsequently, we study the performance of the REXP3 algorithm for channel scheduling, and analyze its AoI regret upper bound.

A. Lower Bound for AoI Regret in Single-source Systems

Consider a single-source system as depicted in Fig. 1, where fresh data are generated at each time slot awaiting transmission. There are K available channels in the environment, following a drifting model of variation. Theorem 1 provides a lower bound on AoI regret in terms of the number of channels K, variation budget V_T, and total time T under the drifting environment.

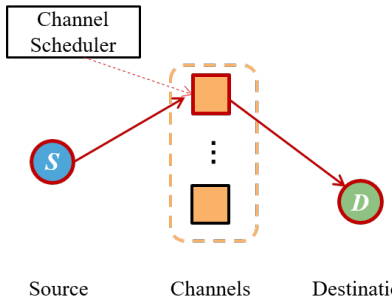

Fig. 1. Single source transmission in the drifting environment.

Theorem 1. *With non-stationary channels under Assumption 1, for any number of channels $K \geq 2$ and time horizon $T \geq 1$, there exists a distribution over the assignment of channel states such that the AoI regret of any policy is $\Omega((KV_T)^{1/3}T^{2/3})$.*

Proof. We extend the regret proof from the switching environment in [12] to the drifting environment by grouping the variations in service rates of non-stationary channels across time slots. We provide a proof outline and highlight the differences.

First, model each decision moment as following a Bernoulli distribution, assuming that the mean of a specific channel k^* is $\mu_{k^*} = \frac{1}{2} + \epsilon$, while all other channels have a mean of $\mu_k = \frac{1}{2}$. This setup is based on the worst-case scenario assumption.

Then, define the indicator function $s(t)$ to represent the success of current scheduling at time t, and $s^*(t)$ to denote the success of optimal scheduling. Use $s(t) - s^*(t)$ to quantify AoI regret. Adopting the approach in [9], a coupling process is introduced to express $\mathbb{E}[s(t) - s^*(t)]$. This leads to Eq.(7) in [12]. For an arbitray policy π, the AoI regret over time T is represented as

$$R_\pi(T) \geq \frac{2\epsilon}{K(1+2\epsilon)} \sum_{k^*=1}^{K} (T - \mathbb{E}_{k^*}[N_{k^*}(T)]),$$

where k^* is the best channel, and $N_{k^*}(T)$ represents number of correct choices.

Time in a drifting environment is segmented into intervals of Δ_T time slots, assuming that the value of channel service

rates remain unchanged in a group. The best channel will only change between the intervals. Under Assumption 1,

$$\sum_{t=1}^{T-1} \sup_{k\in K} |\mu_k(t+1) - \mu_k(t)|$$
$$\leq \sum_{j=1}^{g-1} \epsilon = \left(\left\lceil \frac{T}{\Delta_T} \right\rceil - 1 \right) \cdot \epsilon \leq \frac{T}{\Delta_T} \cdot \epsilon \leq V_T,$$

where $g = \left\lceil \frac{T}{\Delta_T} \right\rceil$ represents the number of groups.

From [18] Theorem A.2,

$$\sum_{k^*=1}^{K} \mathbb{E}\left[N_{k^*}(T_j)\right] \leq |T_j| + \frac{|T_j|}{2} \cdot \sqrt{K|T_j| \log \frac{1}{1-4\epsilon^2}}. \quad (3)$$

By summing over m groups, we obtain

$$\begin{aligned} R_\pi(T) &\geq \textstyle\sum_{j=1}^{g} \frac{\epsilon}{K(\frac{1}{2}+\epsilon)} \cdot \sum_{k^*=1}^{K} (|T_j| - \mathbb{E}\left[N_{k^*}(T_j)\right]) \\ &\geq T \cdot \tfrac{\epsilon}{K(\frac{1}{2}+\epsilon)} \cdot K - \tfrac{\epsilon}{K(\frac{1}{2}+\epsilon)} \textstyle\sum_{j=1}^{g} \sum_{k^*=1}^{K} \mathbb{E}\left[N_{k^*}(T_j)\right]) \\ &\geq T \cdot \tfrac{\epsilon}{\frac{1}{2}+\epsilon} - \tfrac{\epsilon}{K(\frac{1}{2}+\epsilon)} T - \tfrac{T}{2} \cdot \sqrt{K\Delta_T \log \tfrac{1}{1-4\epsilon^2}}, \end{aligned}$$

where the second inequality follows from $\sum_{j=1}^{g} |T_j| = T$ and the last inequality follows from $|T_j| \leq \Delta_T$.

Using the fact that $\log \frac{1}{1-y} \leq 4 \log\left(\frac{4}{3}\right) y$ for $y \in \left[0, \frac{1}{4}\right]$ and $K \geq 2$ from [12],

$$R_\pi(T) \geq T \cdot \frac{\epsilon}{2(\frac{1}{2}+\epsilon)} - \frac{T\epsilon^2}{K(\frac{1}{2}+\epsilon)} \cdot \sqrt{K\Delta_T \log \frac{4}{3}}.$$

Finally, by setting $\Delta_T = \left\lceil K^{\frac{1}{3}} \left(\frac{T}{v_T}\right)^{\frac{2}{3}} \right\rceil$, and $\epsilon = \min\left[\frac{1}{4}\sqrt{\frac{K}{\Delta_T}}, V_T \frac{\Delta_T}{T}\right]$, we obtain

$$R_\pi(T) \geq \frac{1}{4\sqrt{2}} \left(K V_T\right)^{\frac{1}{3}} T^{\frac{2}{3}}.$$

□

B. REXP3 Algorithm and AoI Regret Upper Bound

The scheduling is carried out using the REXP3 algorithm [17] which was originally developed to solve a reward maximization problem. The algorithm aims to address the issue of non-stationary changes by repeatedly restarting the classical EXP3 algorithm in a multi-armed bandit setting. We apply the same algorithm and prove its performance for AoI regret in Theorem 2. For completeness, the algorithm is described as follows.

Theorem 2. *With non-stationary channels under Assumption 1, let π be the REXP3 policy with a batch size $\Delta_T = \lceil (K \log K)^{1/3} (T/V_T)^{2/3} \rceil$ and with $\gamma = \min\left\{1, \sqrt{\frac{K \log K}{(e-1)\Delta_T}}\right\}$. Then, for every $T \geq 1$, $K \geq 2$, the AoI regret is $O\left((K \log K \cdot V_T)^{1/3} T^{2/3}\right)$.*

The proof of Theorem 2 relies on constructing two equivalent worst-case systems. We extend the method in [9], originally applied to stationary channels, to non-stationary channels, providing a new definition for the alternative system in the non-stationary setting and a new general formula for the upper bound of AoI regret. Finally, by incorporating results from the REXP3 algorithm, we complete the proof. The restart scheme presents two key challenges in the proof: the complexity of non-stationary segmented regret summation and the derivation of performance bounds dependent on specific window sizes. These challenges highlight the intricacies involved in adapting the restart strategy to non-stationary environments.

Algorithm 1 REXP3

Input: a positive number γ, and a batch size Δ_T
Output: channel scheduling decision
Set batch index $j = 1$.
while $j \leq \lceil T/\Delta_T \rceil$ **do**
  Set $\tau = (j-1)\Delta_T$.
  Initialization: for any $k \in \mathcal{K}$ set $\omega_t^k = 1$
  for $t = \tau+1, ..., \min\{T, \tau+\Delta_T\}$ **do**
    for $k \in \mathcal{K}$ **do**
      Set $p_t^k = (1-\gamma)\frac{w_t^k}{\sum_{k'=1}^{K} w_t^{k'}} + \frac{\gamma}{K}$.
    end
    Select the channel k' from $\mathcal{K}$ according to the distribution $\{p_t^k\}_{k=1}^K$.
    Receive a reward $X_t^{k'}$.
    For k' set $\hat{X}_t^{k'} = X_t^{k'}/p_t^{k'}$.
    For any $k \neq k'$ set $\hat{X}_t^k = 0$.
    For all $k \in \mathcal{K}$ update:

$$w_{t+1}^k = w_t^k \exp\left\{\frac{\gamma \hat{X}_t^k}{K}\right\}.$$

  end
  Set $j = j+1$.
end
return

Proof. Define an alternative scheme A, representing all cases in the original schedule where the suboptimal channels are replaced by the current **worst channel set**, i.e., the set for channels with service rates $\mu_{\min}(t)$. Let the original scheme be O, then

$$\text{Regret}_A \geq \text{Regret}_O.$$

For scheme A, when the worst channels are aggregated from $t = 1$, the AoI increases [9]. For stationary channels, this conclusion holds because, when calculating the AoI for a specific scheduling, it is represented as the sum of the probabilities of all scheduling errors, which is the product of $1 - \mu_{min}$ or $1 - \mu^*$. Since the product for earlier errors is lower, moving the scheduling with μ_{min} to the front increases the cumulative error probability.

However, directly applying this method in non-stationary channels would result in errors. This is because, at different times, the worst service rates $\mu_{min}(t)$ are not equal. To

ensure that the AoI increases after clustering, we assume that N errors occur, with the times of erroneous scheduling corresponding exactly to the **largest** N service rates in the set $\{\mu_{\min}(t), t = 1, ...T\}$. Under these conditions, the inequality

$$1 - \mu_{min}(t_1) \geq 1 - \mu_{min}(t_2)$$

holds for all $t_1 \leq t_2$. Under this assumption, we define the clustering of erroneous schedules in the first N time slots as Scheme B. Then,

$$\text{Regret}_B \geq \text{Regret}_A.$$

Therefore,

$$\text{Regret}_B \geq \text{Regret}_O.$$

Analyzing the AoI corresponding to scheme B, from Eq.(16)-(18) in [9], we obtain

$$\text{Regret}_B = \sum_{t=1}^{T} \mathbb{E}[h(t)]$$

$$\leq \frac{T}{\mu^{*\min}} + \frac{1-\mu^{*\min}}{\mu^{*\min} p} + \left(\frac{1}{p} - \frac{1}{\mu^{*\min}}\right) \cdot \mathbb{E}[N(T)],$$

where $\mu^{*\min}$ represents the minimum $\mu_{\max}(t)$ across all slots, and $p \leq \mu_{\min}(t)$ is a constant greater than zero. We use this constraint to ensure that no channel is completely shut down, which also aligns with real-world environmental variations. Substituting the bound $\mathbb{E}[N(T)]$ of the REXP3 for the number of incorrect selections by the time T [17], we obtain

$$\text{Regret}_O = O\left(K \log K \cdot V_T\right)^{\frac{1}{3}} T^{\frac{2}{3}}.$$

□

For channel scheduling in a single-source system, we demonstrate that it is possible to learn the variations in channel service rates in drifting environments by employing multi-armed bandit algorithms and adaptive algorithms for handling non-stationary problems. Comparing Theorem 1 and Theorem 2, it is evident that the application of the REXP3 algorithm achieves an AoI regret within a logarithmic factor gap from the lower bound.

IV. Source and Channel Scheduling for Multi-source Systems

Multi-source systems require scheduling for both sources and channels. We discuss decoupled and coupled multi-source systems, and develop the Max Age REXP3 algorithm and Max Weight Age UCB algorithm, respectively. We analyze the upper bounds of AoI regret for both algorithms.

A. Decoupled Source and Channel Scheduling

For the decoupled scheduling system shown in Fig. 2, assume there are M sources and K channels awaiting scheduling. The scheduler outputs source scheduling decisions and channel scheduling decisions independently.

A centralized scheduler knows the instantanous AoI of every source. Therefore, the AoI regret during scheduling arises entirely from suboptimal channel scheduling, and the lower bound for AoI regret is given by Theorem 3.

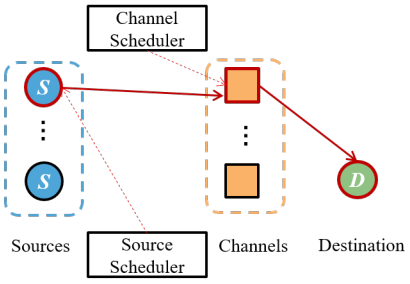

Fig. 2. Decoupled sources and channels.

Theorem 3. *Under Assumption 1, for any $M \geq 2, K \geq 2, T \geq 1$ and decoupled sources and channels, there exists a distribution over the assignment of channel states such that the AoI regret of any policy is $\Omega((KV_T)^{1/3}T^{2/3})$.*

Proof. From Eq. (1) and (2), it can be observed that the AoI regret for multiple sources is the sum of the AoI regrets for individual sources. Let $r_m(t)$ indicate AoI regret for source m at time t. Then, the AoI regret for source m can be represented as:

$$R_m(T) = \mathbb{E}[\sum_{t=1}^{T-1} r_m(t)].$$

The AoI regret for the entire system is expressed as:

$$R(T) = \sum_{m=1}^{M} R_m(t).$$

Under centralized scheduling, both an arbitrary policy π and the optimal policy π^* select the same source in each round.

$$R(T) = \sum_{m=1}^{M} \mathbb{E}[\sum_{t=1}^{T-1} r_m(t)].$$

Moreover, let $h^*_m(t)$ represents AoI of source m under the optimal policy. From the definition of $r_m(t)$:

$$r_m(t) = h_m(t) - h^*_m(t).$$

According to the proof of Theorem 1, the single-source AoI regret is proportional to the number of incorrect channel selections $N(T)$. In the case of multiple sources, the AoI regret is proportional to the sum of the incorrect channel selection counts $N(T)$ for each source. Since the source scheduler is centralized, it can select the best source in each round. For a specific source m at the current time t, let $N_m(t)$ represent the total number of incorrect channel selections up to time t. Then,

(1) If source m is the best source at time t, the increase in the number of incorrect selections satisfies the following:

$$N_m(t) - N_m(t-1) = 1 - \mathbb{I}(C(t) = C^*(t)).$$

(2) If source m is not the best source at time t, then by not scheduling source m at time t, both the optimal strategy

and the current strategy will increase the AoI of source m by 1, resulting in a total number of incorrect channel selections increase of 0.

$$N_m(t) - N_m(t-1) = 0.$$

Summing over all sources M, the regret increment between the current time t and the previous time $t-1$ depends on whether the optimal channel was selected at the current time t. Summing over time T, the lower bound is determined by the number of incorrect selections made over time steps T. Then,

$$\begin{aligned} N(T) &= \sum_{m=1}^{M} \mathbb{E}[\sum_{t=1}^{T-1} (1 - \mathbb{I}(C(t) = C^*(t)) P(S(t) = m)] \\ &= \sum_{t=1}^{T-1} [1 - \mathbb{I}(C(t) = C^*(t))]. \end{aligned}$$

Meanwhile,

$$R(T) = \Omega\left(\mathbb{E}\left[N(T)\right]\right).$$

Substituting $N_{k^*}(T_j)$ from Eq.(3) of Theorem 1 completes the proof. □

If only one source can be selected for transmission at each time slot, and the objective function is the sum of AoI for all sources, the optimal source for scheduling is the one that, when its AoI is reset to 1, results in the greatest reduction of AoI. Based on this observation, we propose the Max Age REXP3 algorithm in Algorithm 2. The performance is given by Theorem 4.

Theorem 4. *Under Assumption 1, for any $M \geq 2, K \geq 2, T \geq 1$ and decoupled source-channel selection, the AoI regret of Max Age REXP3 policy is $O\left((K \log K \cdot V_T)^{1/3} T^{2/3}\right)$.*

The AoI regret in centralized control arises entirely from channel scheduling. Therefore, the proof strategy begins by establishing the optimality of selecting the Max Age source, which results in no AoI regret. Next, the performance of REXP3 is analyzed, yielding an outcome equivalent to Theorem 2. Due to space constraints, detailed proof steps are omitted.

B. Coupled Source and Channel Scheduling

In the coupled system shown in Fig. 3, there is a one-to-one pairing between source and channel, and the number of pairs is K. Since the scheduling decisions made by the scheduler select specific source-channel pairs, it is necessary to consider both the current AoI of the sources and the transmission success rates of the channels.

Assume that an oracle knows the current channels' service rates, the Max-Weight Age scheduling algorithm [1] minimizes AoI by selecting the source-channel pair with the largest $\mu_i h_i(t)\left(h_i(t)+2\right)$. For the problem where the channels' service rates are unknown, we use a reinforcement learning algorithm to learn the channel service rate $\hat{\mu}_i(t)$ at the current time. Subsequently, we select the scheduling decision corresponding to the largest $\hat{\mu}_i(t) h_i(t)\left(h_i(t)+2\right)$. We name the algorithm Max Weight Age UCB in Algorithm 3, and prove its performance in Theorem 5.

Algorithm 2 Max Age REXP3

Input: a positive number γ, and a batch size Δ_T
Output: source-channel scheduling decision
Set batch index $j = 1$.
while $j \leq \lceil T/\Delta_T \rceil$ **do**
  Set $\tau = (j-1)\Delta_T$.
  Initialization: for any $k \in \mathcal{K}$ set $\omega_t^k = 1$
  for $t = \tau+1, ..., \min\{T, \tau+\Delta_T\}$ **do**
    For any $m \in \mathcal{M}$, let $h_m(t)$ denotes the current AoI of the source m.
    Select a source m' such that:

$$m' = \text{argmax}_{m \in \mathcal{M}} \ h_m(t).$$

    for $k \in \mathcal{K}$ **do**
      Set $p_t^k = (1-\gamma)\frac{w_t^k}{\sum_{k'=1}^{K} w_t^{k'}} + \frac{\gamma}{K}$.
    end
    Select the channel k' from $\mathcal{K}$ according to the distribution $\{p_t^k\}_{k=1}^K$.
    Receive a reward $X_t^{k'}$.
    For k' set $\hat{X}_t^{k'} = X_t^{k'}/p_t^{k'}$.
    For any $k \neq k'$ set $\hat{X}_t^k = 0$.
    For all $k \in \mathcal{K}$ update:

$$w_{t+1}^k = w_t^k \exp\left\{\frac{\gamma \hat{X}_t^k}{K}\right\}.$$

  end
  Set $j = j+1$.
end
return

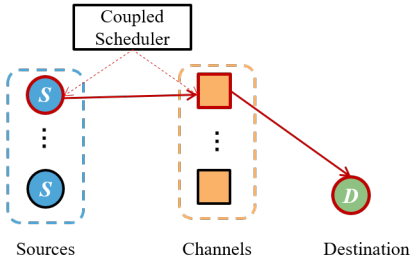

Fig. 3. Coupled sources and channels.

Theorem 5. *Under Assumption 1, for any $M = K \geq 2, T \geq 1$ and coupled source-channel selection, the AoI regret of Max Weight Age UCB policy is $O\left((K \log K \cdot V_T)^{1/3} T^{2/3}\right)$.*

Proof. (sketch) We aim to find a scheduling policy that solves:

$$\max_{i \in I} \mathbb{E}\left[\bar{W}_i(t) \mid h_i(t)\right].$$

As indicated by [16], this problem can be characterized as a

Algorithm 3 Max Weight Age UCB

Input: restart period $\tau = \left\lceil (K \log K)^{1/3} (T/V_T)^{2/3} \right\rceil$, window size $d = \tau$

Output: source-channel pair scheduling decision

For any $i \in \mathcal{I}$, let $h_i(t)$ denotes the current AoI of the pair i.

if $t = \tau_j \in \mathcal{T} = \{\tau_0, \tau_1, ..., \tau_K\}$ **then**

  Initialize pair i reward $\phi_i(\tau_j) = 0$, pair i selected number of times $N_i(\tau_j) = 0$, pair i estimated service rate $\hat{\mu}_i(t) = 0$, $\forall i \in \mathcal{I}$.

  Reset the weights for pair i: $w_i(\tau_j) = \frac{h_i(\tau_j)(h_i(\tau_j)+2)}{\|h(\tau_j)(h(\tau_j)+2)\|_\infty}$.

end

if $t \in (\tau_j, \tau_{j+1})$ **then**

  for $i \in \mathcal{I}$ **do**

    Define the indicator function $s(t)$ to represent the success of current scheduling at time t. Update pair i estimated service rate $\hat{\mu}_i(t)$ as:

    $\phi_i(t) = \phi_i(t-1) + \mathbb{I}(x(t-1) = i)s(t-1)$;

    $N_i(t) = N_i(t-1) + \mathbb{I}(x(t-1) = i)$;

    $\hat{\mu}_i(t) = \frac{\phi_i(t)}{N_i(t)}$;

    $w_i(t) = w_i(\tau_j)$.

  end

end

$\rho_i(t) = \sqrt{\frac{3\log(\tau)}{2N_i(t)}}$(or ∞ if $N_i(t) = 0$), $\forall i \in \mathcal{I}$.

$\bar{W}_i(t) = \min\{w_i(t)\hat{\mu}_i(t) + \rho_i(t), 1\}, \forall i \in \mathcal{I}$.

Activate the source-channel scheduling decision:

$$s(t) = \text{argmax}_{i \in \mathcal{I}} \ \bar{W}_i(t).$$

Update current AoI h_i for any $i \in \mathcal{I}$.

return

stochastic combinatorial multi-armed bandit problem in non-stationary environment and solved via the combinatorial UCB with a sliding window algorithm. In view of the requirements, the reward function satisfies the l_1 triggering probability modulated bounded smoothness assumption and the monotonicity assumption.

The regret within group τ_j is related to the restart time, window size, and degree of variation.

Substituting $\tau = \left\lceil (K \log K)^{1/3} (T/V_T)^{2/3} \right\rceil$ and $d = \tau$ into Lemma 2 in [16], the proof is complete. □

We have addressed the problem of source and channel scheduling in unknown drifting environments for multi-source systems by proposing algorithms and studying their performance. The upper bound performance of both decoupled and coupled source-channel scheduling are consistent with that of single-source scheduling, which remains at the order of $(K \log K \cdot V_T)^{1/3} T^{2/3}$.

V. Simulation

We compare the performance of several algorithms from the perspectives of channel scheduling and source scheduling, corresponding to the single-source transmission in Section 3 and the decoupled multi-source transmission scheduling problem in Section 4. Due to the lack of comparison schemes for the coupled scheduling system, we compare the algorithm's performance under different non-stationary conditions. All results are presented using AoI regret as the performance metric.

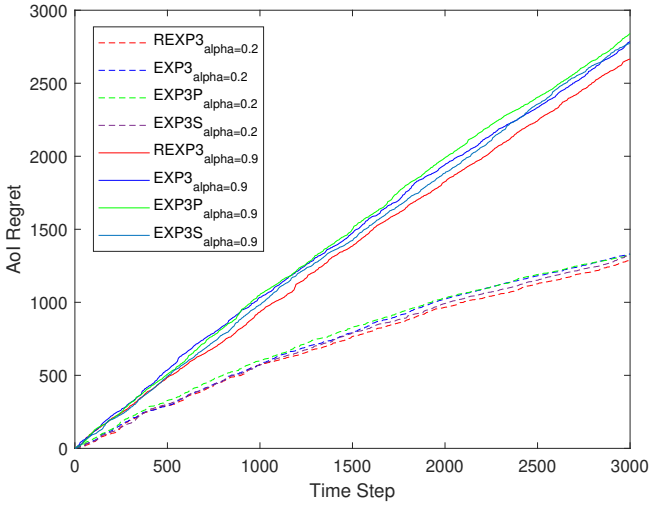

Fig. 4. Comparision of channel scheduling algorithms in single-source system.

We analyze a single-source system that selects the channel with the highest service rate from three options. The initial rates are 0.5, 0.6, and 0.7, with evaluation over 3,000 time slots, using $V_{T1} = T^{0.20}/3$ and $V_{T2} = T^{0.90}/3$ in Figure 4. Baseline performance is compared against EXP algorithms, with dashed lines ($\alpha = 0.2$) and solid lines ($\alpha = 0.9$) in different colors. The results show that smaller V_T reduces environmental variation and AoI regret. The red line, representing our REXP3 algorithm, consistently outperforms the baselines for both α values, demonstrating REXP3's superior performance in dynamic environments.

In Figure 5, we consider a system with three sources and three channels, adding three comparison source selection schemes to REXP3: random scheduling, round-robin and Max Age scheduling. Random scheduling selects one source randomly for scheduling at each time slot, while round-robin scheduling traverses the sources in a specified order to ensure fairness in source scheduling. Comparing the performance of the same algorithm in two drifting environments, all three algorithms exhibit lower AoI regret in the environment with a smaller variation budget. Comparing the performance of these three decoupled scheduling schemes, we observe that random scheduling exhibits the worst AoI performance, followed by round-robin scheduling, whereas the Max Age algorithm we employed achieves the lowest AoI regret.

Coupled scheduling in multi-source transmission lacks comparative schemes; thus, we examine the algorithm performance under various parameters for non-stationary channels. For a model with three source-channel pairs, we conduct simulations with $\mu_{\min}$= 0.05, 0.10, and 0.20. As shown in Figure 8, smaller

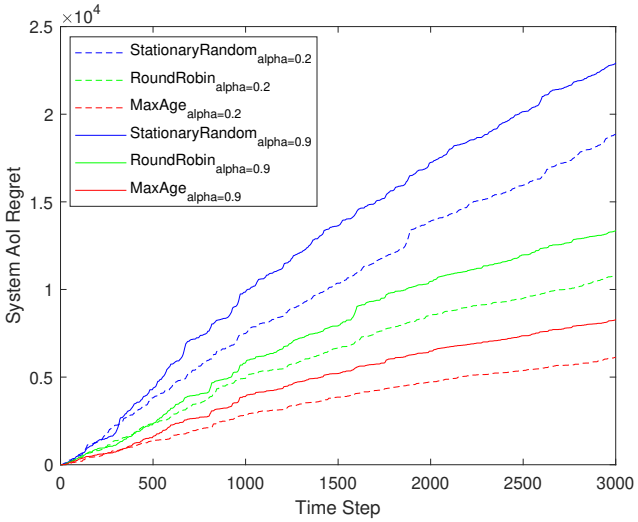

Fig. 5. Comparision of source scheduling algorithms in decoupled multi-source system.

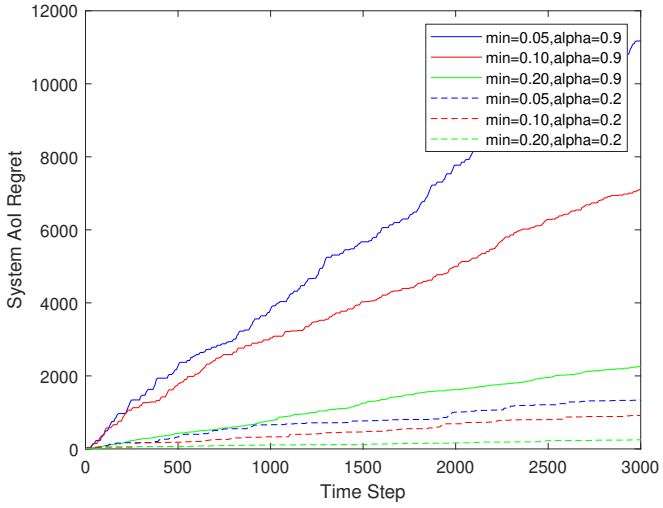

Fig. 6. Comparision of source scheduling algorithms in coupled multi-source system.

α leads to a lower regret. When $\alpha = 0.9$, a decrease in $\mu_{\min}$ leads to a noticeable increase in AoI regret. This phenomenon can be explained by the fact that a smaller $\mu_{\min}$ allows for greater non-stationary variation in the channels, making it more challenging to learn accurate current decisions based on the past scheduling history.

VI. Conclusion

We address scheduling problems in single and multi-source systems under unknown and non-stationary channel conditions, aiming to minimize the system's AoI. Under the drifting environment assumption for non-stationarity, we develop AoI regret lower bounds. We apply REXP3 algorithm for channel scheduling in a single-source system to minimize AoI and prove that the AoI regret upper bound is within a logarithmic factor from the lower bound. Subsequently, we propose the Max Age REXP3 and Max Weight Age UCB scheduling algorithms for decoupled and coupled multi-source systems, respectively, and analyze the upper bounds on AoI regret. Simulation results validate the effectiveness of these algorithms, showing that the proposed algorithms outperform baselines.

References

[1] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.

[2] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2018, pp. 61–70.

[3] V. Tripathi and E. Modiano, "Optimizing age of information with correlated sources," in *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2022, pp. 41–50.

[4] R. V. Ramakanth, V. Tripathi, and E. Modiano, "Monitoring correlated sources: Aoi-based scheduling is nearly optimal," *arXiv preprint arXiv:2312.16813*, 2023.

[5] K.-Y. Lin, Y.-C. Huang, and Y.-P. Hsu, "Scheduling for periodic multi-source systems with peak-age violation guarantees," *IEEE Transactions on Communications*, 2023.

[6] T. Chang, X. Cao, and W. X. Zheng, "A lightweight sensor scheduler based on aoi function for remote state estimation over lossy wireless channels," *IEEE Transactions on Automatic Control*, 2023.

[7] X. Xie, H. Wang, and X. Liu, "Scheduling for minimizing the age of information in multisensor multiserver industrial internet of things systems," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 573–582, 2023.

[8] E. Fountoulakis, T. Charalambous, A. Ephremides, and N. Pappas, "Scheduling policies for aoi minimization with timely throughput constraints," *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 3905–3917, 2023.

[9] S. Fatale, K. Bhandari, U. Narula, S. Moharir, and M. K. Hanawal, "Regret of age-of-information bandits," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 87–100, 2021.

[10] E. U. Atay, I. Kadota, and E. Modiano, "Aging wireless bandits: Regret analysis and order-optimal learning algorithm," in *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*. IEEE, 2021, pp. 1–8.

[11] A. Prasad, V. Jain, and S. Moharir, "Decentralized age-of-information bandits," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–6.

[12] Z. Song, T. Yang, X. Wu, H. Feng, and B. Hu, "Regret of age-of-information bandits in nonstationary wireless networks," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2415–2419, 2022.

[13] W. C. Cheung, D. Simchi-Levi, and R. Zhu, "Learning to optimize under non-stationarity," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1079–1087.

[14] H. Luo, C.-Y. Wei, A. Agarwal, and J. Langford, "Efficient contextual bandits in non-stationary worlds," in *Conference On Learning Theory*. PMLR, 2018, pp. 1739–1776.

[15] J. Huang, L. Golubchik, and L. Huang, "When lyapunov drift based queue scheduling meets adversarial bandit learning," *IEEE/ACM Transactions on Networking*, 2024.

[16] Q. M. Nguyen and E. Modiano, "Learning to schedule in non-stationary wireless networks with unknown statistics," in *Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2023, pp. 181–190.

[17] O. Besbes, Y. Gur, and A. Zeevi, "Stochastic multi-armed-bandit problem with non-stationary rewards," *Advances in neural information processing systems*, vol. 27, 2014.

[18] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The non-stochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.