

# Online Pricing Incentive to Sample Fresh Information

Hongbo Li<sup>1</sup> and Lingjie Duan<sup>2</sup>, *Senior Member, IEEE*

**Abstract**—Today mobile users such as drivers are invited by content providers (e.g., Tripadvisor) to sample fresh information of diverse paths to control the age of information (AoI). However, selfish drivers prefer to travel through the shortest path instead of the others with extra costs in time and gas. To motivate drivers to route and sample diverse paths, this paper is the first to propose online pricing for a provider to economically reward drivers for diverse routing and control the actual AoI dynamics over time and spatial path domains. This online pricing optimization problem should be solved without knowing drivers' costs and even arrivals, and is intractable due to the curse of dimensionality in both time and space. If there is only one non-shortest path, we leverage the Markov decision process (MDP) techniques to analyze the problem. Accordingly, we design a linear-time algorithm for returning optimal online pricing, where a higher pricing reward is needed for a larger AoI. If there are a number of non-shortest paths, we prove that pricing one path at a time is optimal, yet it is not optimal to choose the path with the largest current AoI. Then we propose a new backward-clustered computation method and develop an approximation algorithm to alternate different paths to price over time. Perhaps surprisingly, our analysis of approximation ratio suggests that our algorithm's performance approaches closer to optimum given more paths.

**Index Terms**—Age of information, approximation error, online multi-path pricing, polynomial-time algorithms.

## I. INTRODUCTION

TODAY content providers (e.g., Tripadvisor, Yelp and Google Maps) prefer not to deploy expensive dedicated sensor networks to cover the whole city or nation. Instead, they invite mobile users such as drivers to sample fresh information on diverse paths especially those infrequently visited in the past [1]. The sampled live information on the way include air quality data, shopping promotion and location, and traffic condition [1], [2], [3]. However, selfish drivers are not willing to travel through non-shortest paths to sample fresh information if there are no rewards to offset their extra costs in time and gas [4]. It is shown in [5] that the network performance degrades significantly due to drivers' selfish routing behaviors. To leverage the power of the mobile crowd, it is

critical for providers to properly offer monetary rewards to drivers to change their myopic routing decisions for sampling fresh information along diverse paths. Such incentive mechanisms should be designed in an online version to adapt to the actual variations of information freshness over paths and time.

Recent crowdsourcing works study optimal sensing policies by recruiting mobile vehicles to sense data (e.g., [6], [7], [8], [9]). For example, [6] takes advantage of the mobility of vehicles to provide location-based services in large scale areas. [7] defined spatial and temporal coverage as two metrics for crowdsourcing quality to design greedy and genetic approximation algorithm. [8] incentivizes a crowd of vehicle drivers to sense and sample the desired target regions in one-shot. [9] allows allocated vehicles to follow their origin and destination routes while maximize the overall sensing benefit. However, all of these studies overlook the freshness of sampled data.

To model the information freshness, [10] proposes the concept of age-of-information (AoI) before a new information update is received. Following this, both time-average and peak AoI are developed to measure the average and maximum ages in the time domain, respectively [11], [12]. Numerous works have been analyzing AoI statistics and designing scheduling policies to minimize these two performance metrics [13], [14], [15], [16]. For example, by formulating the AoI state updating as a Markov decision process (MDP), [13] develops approximation algorithms to schedule information packets from multi-sources to end-users for minimizing the average AoI. [14] proposes online scheduling policies to minimize the AoI under different metrics in multi-flow, multi-server systems. [15] compares several scheduling policies to find that transmitting the packet with the largest current AoI is optimal. [17] studies link scheduling to optimize of min-max peak AoI in wireless networks. However, these works only consider the passive packet arrival process for transmission protocols design, and overlook the opportunity to design the sampling process.

To actively sample fresh data, several works study the path planning problem in UAV-assisted IoT to minimize AoI [18], [19], [20]. [18] formulates an MDP to capture the dynamics of UAV locations and applies reinforcement learning to solve the problem. [19] jointly considers energy consumption and AoI evolution to study the data acquisition problem. However, these works only consider the fully controlled UAV to route but overlook vehicle users' selfish behaviors to disobey. As today many mobile users are invited to sample information, it

Manuscript received 14 January 2022; revised 25 September 2022; accepted 19 October 2022. Date of publication 25 October 2022; date of current version 6 January 2023. The work was supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 Grant under Award MOE-T2EP20121-0001. Recommended for acceptance by Dr. Rachid El Azouzi. (Corresponding author: Lingjie Duan.)

The authors are with the Pillar of Engineering Systems and Design, Singapore University of Technology and Design, Singapore, Singapore 487372 (e-mail: hongbo\_li@mymail.sutd.edu.sg; lingjie\_duan@sutd.edu.sg).

Digital Object Identifier 10.1109/TNSE.2022.3216856

2327-4697 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

is important to leverage human-involved sampling or crowdsourcing to control AoI for content providers.

Only a few mobile crowdsourcing works have taken the economics effect of AoI into consideration for information update or reward maximization. [21] considers sampling costs for users to decide when to self-update local data, without considering any incentive design from the content providers. [22] studies how an information customer requests and pays for the fresh information to be updated by the source. [23] studies a two-stage game model for a fresh data market to maximize a platform's profit, where the platform provides data with different AoI to dynamically arriving users. [24] jointly designs optimal upload strategy and incentivizes users to offload data in order to control the AoI. However, these works do not study how to motivate the power of the crowd for sampling fresh data to control AoI. The most related work to this paper is [25], which proposes an offline algorithm to provide sampling rewards to mobile users for controlling expected AoI. Yet this work cannot adapt to unexpected AoI change for dynamic incentive design, and it only looks at a single path instead of a road network for information sampling.

To our best knowledge, this paper is the first work to study how a content provider designs its optimal pricing to reward drivers online for diverse routing and fresh information sampling. To best adapt to the actual variations of AoI over paths and time, we formulate our online optimization problem as a stochastic dynamic program. However, we need to overcome the following technical challenges.

- *Incomplete information about drivers' cost and arrival pattern:* To control the actual AoI evolution in real time, our incentive pricing as compensation should be designed according to drivers' actual arrivals and extra travel costs on non-shortest paths. Yet in practice, such information are private to drivers and unavailable for the content provider when deciding the online pricing. This makes it infeasible to apply online control methods such as Hamilton-Jacobi-Bellman equations and neural networks approximation here [26], [27].
- *Curse of dimensionality in both time and spatial path domains:* The optimal online pricing should be time- and path-dependent, making the number of system states exponentially increase with the time horizon  $T$  and path number  $N$  in backward induction [28], [29]. Some recent work finds special feature of system states (e.g., periodicity) to simplify the backward/forward induction yet cannot apply to our AoI problem (e.g., [30], [31]).

The existing AoI works design offline scheduling policies for multi-channel network by dynamic programming (e.g., [25], [32]). There are some recent work on online scheduling, yet they assume the system space to be countable and linearly increasing with number of channels and time (e.g., [14], [15], [33]). MDP techniques are widely used to model the dynamic pricing problem (e.g., [34], [35], [36]). [34] discretizes the state space into the set of finite intervals to greatly reduce its size, which is not applicable in our multi-path problem with possibly unbounded AoI. [35] proposes to aggregate similar

states to reduce the computational times, but it cannot provide any rigorous performance analysis such as proving approximation error in the worst case. [36] applies heuristic methods to solve the MDP under completed cost information, which did not analyze the error bound or prove structural properties. Their algorithm design and performance analysis cannot apply to our online pricing problem to sample fresh information in a large-scale road network over time.

Our paper aims to overcome the above technical challenges for online pricing design, and our key novelty and main contributions in this paper are summarized as follows.

- *Novel online pricing to sample fresh information over time and space:* To motivate the power of the mobile crowd, this paper is the first to propose online pricing incentive for a content provider to economically reward drivers to sample fresh information through diverse routing. We optimize pricing reward on human-involved sampling to affect the actual AoI dynamics over time and spatial paths, without knowing drivers' hidden extra travel costs and even arrivals.
- *Linear complexity algorithm to solve online pricing for one non-shortest path:* To minimize the state space of the our formulated MDP problem, we exploit the dynamics of AoI feature to jointly apply a fixed look-up table to greatly simplify the backward induction process. Then we design a linear-time algorithm for returning optimal online pricing. We prove structural properties of our unique pricing solution, and show that a higher pricing reward is needed for a larger AoI. We also show that our algorithm is applicable to infinite time horizon.
- *Approximation algorithm to solve online pricing for multiple non-shortest paths:* We propose a new backward-clustered computation method to overcome the curse of dimensionality in the path number. We first prove that it is optimal to only price one path at a time, while it is not optimal to myopically choose the path with the largest current AoI. Based on the backward-clustered method, we develop a new approximation algorithm to alternate different paths to price over time. This algorithm has only polynomial-time complexity, and its complexity does not depend on the number of paths. Perhaps surprisingly, our analysis of approximation ratio suggests that this algorithm's performance approaches closer to the optimum if more paths are involved to sample.

The rest of the paper is organized as follows. In Section II, we overview the system model to sample fresh information and introduce the problem formulation for online pricing. In Sections III and IV, we prove structural properties of our unique pricing solution, and then propose a linear-time algorithm to return optimal online pricing if there is only one non-shortest path. In Section V, we develop a new approximation algorithm to alternate different paths to price over time if there are a number of non-shortest paths, which has only polynomial-time complexity and does not depend on the path number. Finally, we conclude this paper in Section VI.

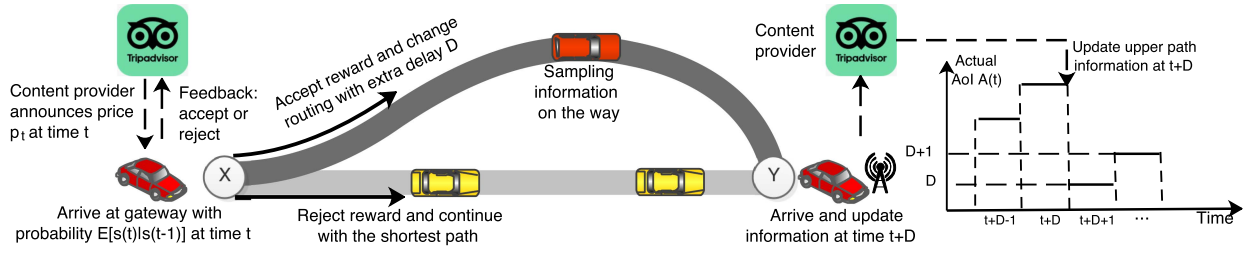


Fig. 1. System model of pricing-driven sampling, where a random flow of drivers sequentially arrive at the gateway X and make routing decisions from X to destination Y. The road network includes the shortest path with normalized zero travel cost and another distant upper path with extra travel delay  $D$ .

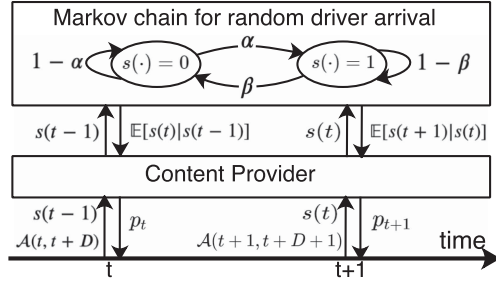


Fig. 2. Information observation and decision process by the online provider.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

As illustrated in Fig. 1, a random flow of drivers sequentially arrive at the gateway X in the discrete time horizon of in total  $T$  rounds. At any time slot  $t = 0, \dots, T$ , a driver (if any) needs to make routing decision from X to destination Y. Here, the road network includes the shortest path with normalized zero travel cost and another distant path with extra travel delay  $D$  in the upper part of Fig. 1. For ease of exposition, we first focus on this case with only one non-shortest path, and then extend our analysis and pricing design to an arbitrary number of non-shortest paths later in Section V.

Selfish drivers hesitate to choose the non-shortest upper path to incur extra travel cost in time and gas, resulting in under-sampling of this path (e.g., to collect air quality data, shopping promotion and location, or traffic condition [1], [2], [3]). There are always enough drivers to cover the shortest path and we do not need to consider the AoI there. Our problem is how to motivate the randomly arriving drivers to sample the non-shortest path to control AoI there. At each time  $t$ , the content provider observes the actual AoI  $A(t)$  of the non-shortest path (see Fig. 1), and compensates an arrived driver at the gateway with pricing reward  $p_t$  to possibly change his routing to the upper path to return the sampled information of the whole path at time  $t + D$ .

At each time slot  $t$ , the content provider can predict the ongoing AoI evolution till  $t + D$ , and its pricing decision needs to be adaptive to foreseeable AoI evolution in set

$$\mathcal{A}(t, t + D) = \{A(\tau) | t \leq \tau \leq t + D\}.$$

We can thus rewrite price  $p_t$  at time  $t$  as  $p_t(\mathcal{A}(t, t + D))$  or simply  $p_t(A(t + D))$ . Note that the content provider will not

decide any price after time  $T - D$ , as a driver can no longer return the sampled information timely before the end time  $T$ .

Next we first introduce the driver's model and then present the online pricing problem formulation.

### A. Driver's Arrival and Cost Model for Sampling

Following the traffic control literature (e.g., [4], [37]), we model drivers' random arrivals at the gateway X over time as a Markov chain. As shown in Fig. 2, if a driver arrives at the beginning of time slot  $t$ , we denote it as  $s(t) = 1$ , and otherwise  $s(t) = 0$ . Each time slot's duration is set small enough such that there is at most one arrival at a time, and we practically model the correlation between arrivals across neighboring time slots. That is, if  $s(t) = 1$ ,

$$s(t + 1) = \begin{cases} 1, & \text{with probability } 1 - \beta, \\ 0, & \text{with probability } \beta. \end{cases}$$

Similarly, if  $s(t) = 0$ , we update  $s(t + 1)$  by replacing the two probabilities  $1 - \beta$  and  $\beta$  above by  $\alpha$  and  $1 - \alpha$ . At the beginning of time slot  $t$ , the content provider only knows past arrival information  $s(t - 1)$  in Fig. 2, and expects arrival probability during time slot  $t$  by using the Markov chain, i.e.,

$$\mathbb{E}[s(t) | s(t - 1)] = s(t - 1)(1 - \beta) + (1 - s(t - 1))\alpha. \quad (1)$$

Besides the random arrival process, we consider the challenging incomplete information scenario that the content provider does not know each driver's actual cost when deciding pricing  $p_t(A(t + D))$  before observing  $s(t)$ . A driver has extra cost to travel on the non-shortest path with delay  $D$ , and different drivers have different cost sensitivities. Let  $x \in [0, 1]$  be the normalized cost sensitivity of a driver, then we model its cost as  $\text{cost} = xD$ , which is proportional to the delay with individual cost sensitivity  $x$ . Thus, a driver with greater  $x$  is more cost sensitive and less likely to accept the price offer to change route. Yet, the provider only knows that  $x$  in a normalized range  $[0, 1]$  randomly follows cumulative distribution function (CDF)  $F(\cdot)$ , which can be obtained by fitting different historical data's frequencies into a histogram and converting it to CDF as [38]. The maximum travel cost is thus  $D$  and the content provider will not decide any pricing reward  $p_t$  beyond  $D$  to over-pay any driver, and we expect

$$0 \leq p_t(A(t + D)) \leq D.$$

The cost sensitivity distribution in certain applications is considered to be truncated normal or logistic random distribution [38], [39], which are used for our simulations later. In this paper, we assume that the i.i.d. random distributions  $F(\cdot)$  of drivers' cost sensitivities satisfies the following assumption, which is the regular value distribution widely assumed in the mechanism design literature [40].

*Assumption 1:* Assume that the following function

$$H(x) := x + \frac{F(x)}{F'(x)} \quad (2)$$

increases monotonically in  $x \in [0, 1]$ , where  $F(x)$  and  $F'(x)$  are the CDF and PDF of a driver's random cost sensitivity.

According to [40], the second term  $\frac{F(x)}{F'(x)}$  of (2) is the complementary hazard rate and the monotonicity of  $H(x)$  tells the Myerson's regularity. Assumption 1 is not strong and it is satisfied by multiple common distributions such as uniform, exponential and logistic distributions. Later in Section III we will relax Assumption 1 and extend our analysis and results to some other distributions (e.g., truncated normal distribution) in Corollary 1.

### B. Online Pricing Problem Formulation

Upon arrival at time  $t$  with  $s(t) = 1$ , a driver observes the current pricing reward  $p_t(A(t+D))$ . He decides to accept this offer or not, by checking if his utility

$$U_t(A(t+D)) = p_t(A(t+D)) - xD \quad (3)$$

of travelling on the non-shortest path is positive or not. If  $s(t) = 1$  and  $U_t(p_t(A(t+D))) \geq 0$  at time  $t$ , the driver accepts the offer at the gateway X, and updates the whole path's information after travel delay  $D$ , helping reduce the AoI  $A(t+D)$  of this path to  $D$ , as shown in the right part of Fig. 1. Otherwise, the next foreseen AoI  $A(t+D+1)$  at time  $t+D+1$  increases from  $A(t+D)$  by one slot to  $A(t+D)+1$ . Thus, the dynamics of the actual AoI is given as

$$A(t+D+1) = \begin{cases} D, & \text{if } U_t(A(t+D)) \geq 0 \\ & \text{and } s(t) = 1; \\ A(t+D) + 1, & \text{otherwise,} \end{cases} \quad (4)$$

where  $U_t(A(t+D))$  is defined in (3), and the pricing reward is accepted in the former case with probability

$$Q(t) = \mathbb{E}[s(t)|s(t-1)]F\left(\frac{p_t(A(t+D))}{D}\right) \quad (5)$$

with  $\mathbb{E}[s(t)|s(t-1)]$  in (1) from the content provider's point of view. The expectation of the final payment to the possibly arrived driver at time  $t$  is thus  $Q(t)p_t(A(t+D))$ , and the online pricing  $p_t(A(t+D))$  needs to well balance the next AoI evolution in (4) and the expected payment.

Finally, we are ready to formulate the provider's online pricing problem in Fig. 2 by the Markov decision process (MDP) techniques with four components [28], [41] to best

adapt to the actual AoI evolution  $A(t+D)$  and past arrival observation  $s(t-1)$  under the memoryless property.

- *States:* We define the state of the MDP in slot  $t$  by the tuple  $\mathbf{S}_t = (A(t+D), s(t-1))$ . Note that  $s(-1) = 0$  at initial time slot  $t = 0$ .
- *Actions:* The action of the MDP in slot  $t$  is the price  $p_t(A(t+D)) \in [0, D]$ . Note that the continuous action space size is infinite.
- *Transition probabilities:* According to the conditional arrival probability in (1) and AoI dynamics in (4), there can be three possible states at the next time slot  $t+1$ . The path will be sampled by an arrival driver with probability  $Q(t)$  in (5). If there is no arrival at time  $t$ , the state  $\mathbf{S}_{t+1}$  at  $t+1$  will be  $(A(t+D)+1, 0)$  with probability

$$Q_0(t) = 1 - \mathbb{E}[s(t)|s(t-1)]. \quad (6)$$

If there is an arrival at time  $t$  but the driver does not accept the price, the state  $\mathbf{S}_{t+1}$  at  $t+1$  will be  $(A(t+D)+1, 1)$  instead with probability

$$Q_1(t) = \mathbb{E}[s(t)|s(t-1)]\left(1 - F\left(\frac{p_t(A(t+D))}{D}\right)\right). \quad (7)$$

In summary, we can obtain the following state transitions of  $\mathbf{S}_{t+1}$ :

$$\begin{cases} \bar{\mathbf{S}}_{t+1} = (D, 1), & \text{with probability } Q(t) \text{ in (5),} \\ \mathbf{S}_{t+1}^0 = (A(t+D)+1, 0), & \text{with probability } Q_0(t) \text{ in (6),} \\ \mathbf{S}_{t+1}^1 = (A(t+D)+1, 1), & \text{with probability } Q_1(t) \text{ in (7),} \end{cases} \quad (8)$$

which correspond to the three outcomes: arrival to sample, no current arrival, and arrival to not sample.

- *Cost:* Let  $V(\mathbf{S}_t, p_t(A(t+D)))$  be the immediate cost of the MDP if action  $p_t(A(t+D))$  is taken in slot  $t$  under state  $\mathbf{S}_t$ , which is defined as the summation of the actual AoI and expected economic payment:

$$V(\mathbf{S}_t, p_t(A(t+D))) = A(t+D) + Q(t)p_t(A(t+D)). \quad (9)$$

Considering the discount factor  $\rho \in (0, 1)$  under discrete time horizon, the objective of the MDP is to find the optimal pricing function  $p_t(A(t+D))$  at current time slot  $t$  that minimizes the long-term total  $\rho$ -discounted cost:

$$C_t^*(\mathbf{S}_t) = \min_{p_t(A(t+D)) \in [0, D]} \sum_{\tau=t}^T \rho^{\tau-t} V(\mathbf{S}_\tau, p_\tau(A(\tau+D))), \quad (10)$$

s.t. (1), (3), (4) and (9),

which is a non-convex problem due to the non-convex AoI dynamics constraint. The current price  $p_t(A(t+D))$  to announce affects the dynamics of AoI since  $t+1$ . Note that the current pricing decision can only affect the AoI after a time delay  $D$ , thus like  $p_t(A(t+D))$  we also use the foreseen

**Algorithm 1:** The computation of look-up table  $\mathbf{C}^*(\bar{\mathbf{S}})$  by backward induction.

---

```

1: Store  $C_{T-D}^*(D, 1) = D$  in  $\mathbf{C}^*(\bar{\mathbf{S}})$  table;
2: for  $i = 1$  to  $i = T - D$  do
3:   Compute  $C_{T-D}^*(D + i, 1)$  and  $C_{T-D}^*(D + i, 0)$  according to (14);
4:   for  $j = 1$  to  $i$  do
5:     Obtain  $C_{T-D-j+1}^*(D, 1)$  from  $\mathbf{C}^*(\bar{\mathbf{S}})$  table;
6:     Compute  $p_{T-D-j}^*(D + i - j)$  according to (12) using binary search with error  $\varepsilon$ ;
7:     Compute  $C_{T-D-j}^*(D + i - j, 1)$  and  $C_{T-D-j}^*(D + i - j, 0)$  according to (11) from last decision slot;
8:   end for
9:   Store  $C_{T-D-i}^*(D, 1)$  in  $\mathbf{C}^*(\bar{\mathbf{S}})$  table;
10: end for
11: return  $\mathbf{C}^*(\bar{\mathbf{S}})$  table

```

---

AoI  $A(t + D)$  in the cost objective function  $C_t^*(\mathbf{S}_t)$  in (10) above.

### C. Characterization of Dynamic Program

For any time  $t = \{0, \dots, T - D\}$  in the finite time horizon, problem (10) can be written as [28]:

$$C_t^*(\mathbf{S}_t) = \min_{p_t(A(t+D))} V(\mathbf{S}_t, p_t(A(t+D))) + \rho \mathbb{E}_{\mathbf{S}_{t+1}} [C_{t+1}^*(\mathbf{S}_{t+1})], \quad (11)$$

where the cost-to-go is

$$\mathbb{E}_{\mathbf{S}_{t+1}} [C_{t+1}^*(\mathbf{S}_{t+1})] = \sum_{j=0}^1 Q_j(t) C_{t+1}^*(\mathbf{S}_{t+1}^j) + Q(t) C_{t+1}^*(\bar{\mathbf{S}}_{t+1})$$

according to the transition probabilities (8).

The optimal pricing satisfies the first-order necessary condition of (11) with respect to  $p_t(A(t + D))$ , i.e.,

$$\frac{p_t^*(A(t + D))}{D} + \frac{F\left(\frac{p_t^*(A(t+D))}{D}\right)}{F'\left(\frac{p_t^*(A(t+D))}{D}\right)} - \frac{\rho}{D} \Delta C_{t+1}^* = 0, \quad (12)$$

where

$$\Delta C_{t+1}^* = C_{t+1}^*(\mathbf{S}_{t+1}^1) - C_{t+1}^*(\bar{\mathbf{S}}_{t+1}). \quad (13)$$

To solve (12), we need the inputs of  $C_{t+1}^*(\mathbf{S}_{t+1}^1)$  and  $C_{t+1}^*(\bar{\mathbf{S}}_{t+1})$ . For each long-term cost function since time  $t$ , it takes  $O(\log_2(\frac{D}{\varepsilon}))$  complexity to obtain its pricing solution to (12) using binary search with error  $\varepsilon$  [42].

Lacking the information about drivers' hidden sampling costs and even arrivals, the transition probabilities are related with the pricing  $p_t^*(A(t + D))$  in (11). Therefore, it is infeasible to apply online control methods such as Hamilton-Jacobi-Bellman equations and neural networks approximation here [26], [27], [43].

As it is too late to price after  $t = T - D$  for returning timely information before end-time  $T$ , we obtain from (11) that

**Algorithm 2:** The linear-time computation of online optimal pricing  $p_t^*(A(t + D))$  to (11) at any time  $t \in \{0, \dots, T - D\}$ .

---

```

1: Input:  $\mathbf{C}^*(\bar{\mathbf{S}})$  table, binary search error  $\varepsilon$ ;
2: Observe  $A(t)$  and predict foreseen AoI  $A(t + D)$ ;
3: Compute  $C_{T-D}^*(\mathbf{S}_{T-D}^0)$  and  $C_{T-D}^*(\mathbf{S}_{T-D}^1)$  according to (14);
4: for  $1 \leq i \leq T - D - t - 1$  do
5:   Obtain  $C_{T-D-i}^*(\bar{\mathbf{S}}_{T-D-i})$  from  $\mathbf{C}^*(\bar{\mathbf{S}})$  table;
6:   Compute  $C_{T-D-i}^*(\mathbf{S}_{T-D-i}^0)$ ,  $C_{T-D-i}^*(\mathbf{S}_{T-D-i}^1)$  according to (11) from last decision slot  $T - D - i + 1$ ;
7:   Compute  $p_{T-D-1-i}^*(A(t + D) + T - D - 1 - i)$  according to (12) using binary search with error  $\varepsilon$ ;
8: end for
9: return  $p_t^*(A(t + D))$  for current time  $t$ 

```

---

$$C_{T-D}^*(\mathbf{S}_{T-D}) = A(T). \quad (14)$$

Based on the state transitions in (8) and equation (14), we find that the state space increases polynomially with time horizon  $T$  in (11) for the single-path pricing problem. Thus, one can use backward induction to solve the problem (11) in polynomial time [28], [30], which may be not small for a large  $T$ . Thus, we will further exploit the unique AoI feature and propose Algorithms 1 and 2 to reduce to linear complexity later in Section IV. On the other hand, we will show later in Section V that the state space increases exponentially in  $T$  and path number  $N$ , which forbids to apply traditional MDP methods [44].

Before that, we first prove key structural properties of the optimal (long-term) cost functions in (11) and optimal online pricing in next section. For ease of reading, we summarize all the key notations in Table I.

### III. PROVED STRUCTURAL PROPERTIES OF OPTIMAL ONLINE PRICING

Since the price  $p_t(A(t + D))$  is located in a compact interval  $[0, D]$ , there always exists optimal pricing solutions to the problem (11). However, since drivers' information about costs and random arrivals are incomplete, the monotonicity of cost function with respect to  $A(t + D)$  and the uniqueness of optimal pricing solution to (12) cannot be guaranteed [45] and [28]. In this section, at any time  $t$ , we first examine the monotonic relationship of cost function and optimal pricing with respect to the foreseen AoI  $A(t + D)$ . Then we prove the uniqueness of optimal online solution to problem (11).

By observing (12), we find the optimal pricing solution  $p_t^*(A(t + D))$  is determined by the the cost functions  $C_{t+1}^*(\mathbf{S}_{t+1}^1)$  without information update,  $C_{t+1}^*(\bar{\mathbf{S}}_{t+1})$  with update, and the CDF  $F(\cdot)$ . To examine the optimal pricing's relationship with  $A(t + D)$ , we first need to examine the relationship between cost function  $C_t^*(\mathbf{S}_t)$  in (11) and foreseen AoI  $A(t + D)$  [28].

**Proposition 1:** The optimal cost function  $C_t^*(\mathbf{S}_t)$  in (11) at any time  $t \in \{0, 1, \dots, T - D\}$  increases monotonically with the foreseen AoI  $A(t + D)$ .



TABLE I  
KEY NOTATIONS USED IN THE PAPER

Notation	Definition
$D$	The extra travel delay of the single non-shortest path
$T$	The time horizon
$A(t+D)$	The foreseen AoI with the travel delay $D$ at the beginning of time slot $t$
$p_t(\cdot)$	The pricing reward to compensate an arrived driver at time $t$
$s(t-1)$	Driver's arrival information of the last time slot $t-1$
$\mathbb{E}[s(t) s(t-1)]$	The conditional probability of a driver's arrival during time slot $t$ .
$x$	The cost sensitivity of a driver
$F(\cdot)$	The CDF of a driver's random cost sensitivity
$U_t(\cdot)$	The utility of a driver at time slot $t$
$\mathbf{S}_t$	The state $(A(t+D), s(t-1))$ at time $t$ in the single path problem
$Q(t)$	Transition probabilities at time slot $t$ in MDP
$V(\cdot)$	The immediate cost of the MDP
$\rho$	The discount factor
$C_t(\cdot)$	The $\rho$ -discounted long-term cost function at time $t$
$\varepsilon$	The error of binary search
$C^*(\tilde{\mathbf{S}})$	The look-up table to store cost function $C_t^*(\tilde{\mathbf{S}}_t)$ for any $t$
$N$	The total number of non-shortest paths
$D_i$	The travel delay of the $i$ -th non-shortest path
$\hat{t}_t$	The index of the path with the maximum net payoff at time slot $t$
$\mathbb{A}(t)$	The foreseen AoI set of all the non-shortest paths at the beginning of time slot $t$
$\mathbb{S}_t$	The state $(\mathbb{A}(t), \hat{t}_{t-1}, s(t-1))$ at time $t$ in the multi-path problem

*Proof:* Suppose that there are two states  $\mathbf{S}_t^a$  and  $\mathbf{S}_t^b$  with their actual AoI satisfying  $A^a(t+D) > A^b(t+D)$  at any time slot  $t \in \{0, \dots, T-D\}$ . In the following, we apply both mathematical induction and backward induction to prove

$$C_t^*(\mathbf{S}_t^a) > C_t^*(\mathbf{S}_t^b), \quad (15)$$

which tells the monotonicity of the optimal cost function.

We first prove the base case at the last time slot  $t = T-D$ . According to (14), the two cost functions satisfy

$$C_{T-D}^*(\mathbf{S}_{T-D}^a) = A^a(T) > A^b(T) = C_{T-D}^*(\mathbf{S}_{T-D}^b),$$

which holds for (15).

Next, we assume the induction hypothesis that  $C_{t_0+1}^*(\mathbf{S}_{t_0+1}^a) > C_{t_0+1}^*(\mathbf{S}_{t_0+1}^b)$  is true for a particular time slot  $t_0+1$ . It follows to show that  $C_{t_0}^*(\mathbf{S}_{t_0}^a) > C_{t_0}^*(\mathbf{S}_{t_0}^b)$  also holds. Denote by  $p_{t_0}^*(A^a(t_0+D))$  and  $p_{t_0}^*(A^b(t_0+D))$  the optimal pricing to  $A^a(t_0+D)$  and  $A^b(t_0+D)$ , respectively. Consider another non-optimal pricing to  $A^b(t_0+D)$ :

$$\hat{p}_{t_0}(A^b(t_0+D)) = p_{t_0}^*(A^a(t_0+D)), \quad (16)$$

whose corresponding non-optimal cost function must satisfy  $\hat{C}_{t_0}^*(\mathbf{S}_{t_0}^b) \geq C_{t_0}^*(\mathbf{S}_{t_0}^b)$ . Then we can derive the induction step:

$$\begin{aligned} C_{t_0}^*(\mathbf{S}_{t_0}^a) - C_{t_0}^*(\mathbf{S}_{t_0}^b) &\geq C_{t_0}^*(\mathbf{S}_{t_0}^a) - \hat{C}_{t_0}^*(\mathbf{S}_{t_0}^b) \\ &= A^a(t+D) - A^b(t+D) \\ &\quad + \rho \mathbb{E}_{\mathbf{S}_{t_0+1}^a} [C_{t_0+1}^*(\mathbf{S}_{t_0+1}^a)] - \rho \mathbb{E}_{\mathbf{S}_{t_0+1}^b} [C_{t_0+1}^*(\mathbf{S}_{t_0+1}^b)], \end{aligned}$$

which is larger than 0 due to the fact that  $A^a(t+D) > A^b(t+D)$ ,  $C_{t_0+1}^*(\mathbf{S}_{t_0+1}^a) \geq C_{t_0+1}^*(\mathbf{S}_{t_0+1}^b)$  and (16).

As we have proven that both the base case and the induction step are true, (15) holds for each  $t \in \{0, \dots, T-D\}$  by mathematical induction. This further tells the monotonicity of the optimal cost function  $C_t^*(\mathbf{S}_t)$  in (11). ■

Intuitively, a larger foreseen AoI  $A(t+D)$  leads to larger long-term cost. Based on the monotonicity of the long-term cost function in Proposition 1, next we are ready to prove the uniqueness of optimal online pricing solution to problem (11) at any time  $t$ .

*Proposition 2:* Under Assumption 1, the optimal online pricing solution  $p_t^*(A(t+D))$  to problem (11) exists and is unique at any time  $t \in \{0, \dots, T-D\}$ .

The proof of Proposition 2 is given in our technical report [46]. We can generalize our Proposition 2 to fit some other distributions such as truncated normal distribution.

*Corollary 1:* If there exists a cutoff point  $x_{th} \in [0, 1]$ , such that  $H(x) < 0$  for  $x \in [0, x_{th}]$  and  $H(x)$  increases with  $x \in [x_{th}, 1]$ , which holds for truncated normal distribution, then the optimal pricing solution to problem (11) exists and is unique.

Based on Propositions 1, 2 and the expression (12), we can also derive the monotonicity of optimal online pricing  $p_t^*(A(t+D))$  with respect to  $A(t+D)$  [28].

*Corollary 2:* The unique optimal online pricing  $p_t^*(A(t+D))$  at any time  $t \in \{0, \dots, T-D\}$  increases monotonically with the foreseen AoI  $A(t+D)$ .

Corollary 2 tells that as the foreseen AoI is large to dominate in the long-term cost objective (11), we should set a high price to immediately motivate drivers to sample and control the AoI evolution.

#### IV. LINEAR-TIME ALGORITHM FOR OPTIMAL ONLINE PRICING

Recall that backward induction can be used to solve problem (11) in polynomial time. In this section, we exploit the dynamic feature of AoI to jointly use a fixed look-up table and backward induction to further reduce the increasing number of cost functions in the time domain. Furthermore, we show our designed algorithm is also applicable to infinite time horizon.

According to (12), at any time  $t$  we first need to compute long-term cost functions  $C_{t+1}^*(\mathbf{S}_{t+1}^1)$  and  $C_{t+1}^*(\tilde{\mathbf{S}}_{t+1})$  before solving optimal pricing  $p_t^*(A(t+D))$ . To obtain  $C_{t+1}^*(\mathbf{S}_{t+1}^1)$  for example, as discussed at the end of Section II, we further need to apply iteration computation to derive another three cost functions  $C_{t+2}^*(\mathbf{S}_{t+2}^0)$ ,  $C_{t+2}^*(\mathbf{S}_{t+2}^1)$  and  $C_{t+2}^*(\tilde{\mathbf{S}}_{t+2})$  at time  $t+2$ . After careful examination of this branching network, we find out that a large number of cost functions repeatedly appear due to the linearly increasing feature of AoI over time. Take the transition of  $(A(t+D), 0)$  at time  $t=0$  with

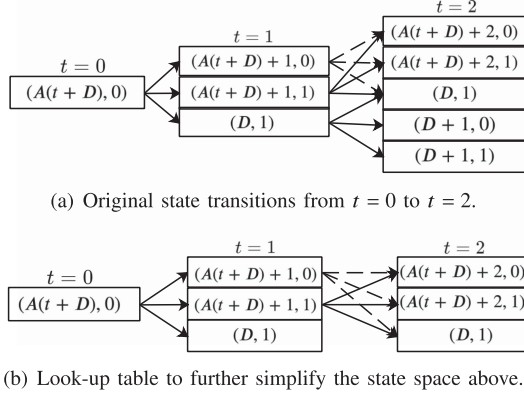


Fig. 3. An illustrative example of  $T - D = 2$  for the look-up table approach to reduce state space. (a)Original state transitions from  $t = 0$  to  $t = 2$  (b)Look-up table to further simplify the state space above.

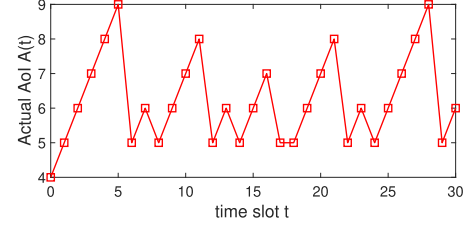
$T - D = 2$  as an illustrative example, which implies three decision slots  $t = 0, 1, 2$ . Fig. 3(a) shows the polynomial increase of state space for iterative computation from  $t = 0$  to  $t = 2$ . At time  $t = 1$ ,  $(A(t+D) + 1, 0)$  with no driver arrival during last time slot  $t = 0$  (i.e.,  $s(0) = 0$ ) and  $(A(t+D) + 1, 1)$  with last arrival have the same branches to  $(A(t+D) + 2, 0)$ ,  $(A(t+D) + 2, 1)$  and  $(D, 1)$  at  $t = 2$ . As there are actually 5 states at time  $t = 2$ , we can apply backward induction as [44] and [30] to derive the optimal online pricing at current time  $t$  in polynomial time.

Moreover, as inspired by [47] and [48], we want to further simplify the state space by applying the look-up table approach. As shown in Fig. 3(a), we find the state  $\bar{\mathbf{S}}_t = (D, 1)$  appears at each time slot  $t \in \{0, \dots, T - D\}$  as it is possible to find the driver arrival to sample and update information at any time. Since the inputs  $D$  and 1 of these functions are constant, we do not need to wait till our observation of such constants to decide the pricing online. Instead, we propose to compute any  $C_t^*(\bar{\mathbf{S}}_t)$  before  $t = 0$  and store them in a fixed look-up table  $\mathbf{C}^*(\bar{\mathbf{S}})$  for any online use later. Then we no longer need to expand the state space from  $\bar{\mathbf{S}}_t = (D, 1)$  at any time  $t$ . As illustrated in Fig. 3(b), by jointly applying the look-up table, we end up with only three cost functions to compute online at both  $t = 1$  and  $t = 2$ . Thus, we successfully reduce the number of cost functions to compute online from quadratic term  $(T - D)^2$  to linear item  $3(T - D)$  at each time  $t$ .

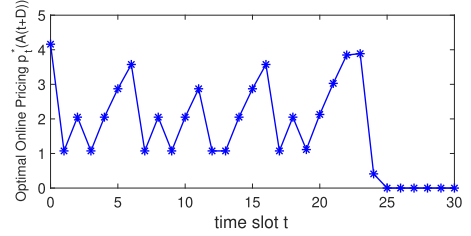
Thanks to the look-up table approach that successfully reduces the polynomially increasing number of cost functions to linear number, we are ready to propose Algorithm 1 to compute the look-up table  $\mathbf{C}^*(\bar{\mathbf{S}})$  and Algorithm 2 to optimally solve problem (11) for online pricing at any time  $t = \{0, \dots, T - D\}$ .

In Algorithm 1, we apply backward induction to calculate cost functions  $C_t^*(D, 1)$  from  $t = T - D$  to  $t = 0$ :

- In step 1, we first obtain the value of  $C_{T-D}^*(D, 1)$  in time slot  $T - D$  according to (14).
- From step 2 to 10, we use the loop to calculate  $C_t^*(D, 1)$  backward from time slot  $t = T - D - 1$  to  $t = 0$ , and store each value in the  $\mathbf{C}^*(\bar{\mathbf{S}})$  table in step 9.



(a) Evolution of actual AoI under the optimal online pricing  $p_t^*(A(t+D))$ ,  $t \in \{0, \dots, T = 30\}$ .



(b) Online optimal pricing policy returned by Algorithm 2.

Fig. 4. The evolution of actual AoI  $A(t)$  and online optimal price  $p_t^*(A(t+D))$  returned by Algorithm 2 from  $t = 0$  to 30. Here we set  $A(0) = 4$ ,  $T = 30$ ,  $D = 5$ ,  $\rho = 0.85$ , probabilities  $\alpha = 0.8$ ,  $\beta = 0.6$  for the Markov chain in Fig. 2, and each driver's cost CDF  $F(x)$  satisfies a truncated normal distribution with mean 0.6 and variance 0.7. (a)Evolution of actual AoI under the optimal online pricing  $p_t^*(A(t+D))$ ,  $t \in \{0, \dots, T = 30\}$  (b)Online optimal pricing policy returned by Algorithm 2.

- From step 4 to 8, to calculate each  $C_{T-D-i}^*(D, 1)$  in the  $i$ -th loop, we calculate all the possible long-term cost functions backward from time slot  $t = T - D$  to  $t = T - D - i + 1$ , according to Fig. 3(a).

Note that here we use the state tuple  $(D, 1)$  in  $C_t^*(\cdot)$  to distinguish from the state space  $\bar{\mathbf{S}}_t$  elsewhere. After returning the fixed look-up table  $\mathbf{C}^*(\bar{\mathbf{S}})$ , we are ready to apply it to solve  $p^*(A(t+D))$  in Algorithm 2:

- In step 3, we first obtain the known value of  $C_{T-D}^*(\bar{\mathbf{S}}_{T-D})$  and  $C_{T-D}^*(\bar{\mathbf{S}}_{T-D}^1)$  in time slot  $T - D$  according to (14), where  $\bar{\mathbf{S}}_{T-D}^0$  and  $\bar{\mathbf{S}}_{T-D}^1$  are  $(A(D) + T - D, 0)$  and  $(A(D) + T - D, 1)$ , respectively.
- From step 4 to 8, we calculate all the possible long-term cost functions and the corresponding optimal price backward from time slot  $T - D - t - 1$  to  $t + 1$ .
- We finally obtain the optimal price  $p_t^*(A(t+D))$  during the last loop  $i = T - D - t - 1$ .

As the pricing action space  $[0, D]$  is continuous, we equally discretize it into  $(\frac{D}{\varepsilon} + 1)$  many points with  $\varepsilon$  as the gap. Here  $\varepsilon$  can also be viewed as the error of binary search to find the optimal online pricing  $p_t^*$  according to (12).

Based on the above analysis, we propose the following theorem and further analyze the complexity of Algorithm 2.

**Theorem 1:** Algorithm 2 returns the optimal online pricing  $p_t^*(A(t+D))$  with the linear complexity  $O((T - D)\log_2(\frac{D}{\varepsilon}))$  at any time  $t = \{0, \dots, T - D\}$ , where  $\varepsilon$  is the binary search error in step 7 of Algorithm 2.

The proof of Theorem 1 is given in our technical report [46]. Note that storing any other cost functions into a look-up table cannot help to reduce the complexity anymore, because the initial AoI at time  $t = 0$  is unknown and can be unbounded.

Thus, the linear time complexity of Algorithm 2 is the minimum and cannot be further improved [28], [29].

In the following experiment, we create a typical instance to run the online optimal pricing  $p_t^*(A(t+D))$  returned by Algorithm 2 and corresponding AoI  $A(t)$ ,  $\forall t \in \{0, \dots, T=30\}$  in Fig. 4. Here we set the travel delay  $D=5$  for the non-shortest path in Fig. 1 with initial AoI  $A(0)=4$ , and the cost distribution of each driver to follow truncated normal distribution with mean 0.6 and variance 0.7. Fig. 4 shows that the online pricing to announce follows a delayed pattern of the actual AoI observation over time. This is consistent with the monotonicity of the pricing with respect to delayed AoI by  $D$  in Corollary 2. We also notice the online pricing converges to zero since  $t=T-D$ , as any outdated update after  $T$  cannot help.

*Remark 1:* Actually, we can revise our algorithm to fit the infinite time horizon  $T \rightarrow \infty$ , by replacing the  $T$  time window by a future finite  $T_{th}$  window in Algorithm 1 and 2 for efficient computation (with constant complexity) at any time  $t$ . We can first prove that the optimal cost functions are bounded under  $T \rightarrow \infty$ , and then prove that the performance error of this approximation algorithm (revised from Algorithm 2) exponentially reduces to zero as  $T_{th}$  increases.

## V. GENERALIZATION OF ONLINE PRICING TO MULTI-PATH SAMPLING SCENARIO

For ease of exposition, we only consider to sample one non-shortest path in Fig. 1 and propose a linear algorithm to solve the problem. In this section, we generalize to sample an arbitrary number  $N$  of different non-shortest paths and introduce the generalized system model first. Different from the single-path problem, the state space here increases exponentially with time horizon  $T$  and path number  $N$ . To overcome the curse of dimensionality in the spatial path domain (i.e., the path number), we first prove that it is optimal to only price one path at a time, which is not necessarily the path with the largest current AoI. We then propose a new backward-clustered computation method across paths and design an approximation algorithm of polynomial complexity to alternate different paths to price over time.

### A. Model Extension to Sample Multiple Diverse Paths

Upon arrival at the gateway  $X$  by following the Markov chain, a driver now faces an arbitrary number  $N$  of non-shortest paths from  $X$  to destination point  $Y$  for routing. We consider diverse paths such that each path  $i \in \{1, \dots, N\}$  has a unique travel delay  $D_i$  and a driver with personalized cost sensitivity  $x \sim F(\cdot)$  incurs

$$\text{cost} = xD_i$$

to travel on this path to sample fresh information there. We aim to design online pricing to all the paths to help sample fresh information globally to control the maximum AoI. We summarize the actual AoI evolutions in all the  $N$  paths in set

$$\mathbb{A}(t) = \{A_i(t+D_i) | i \in \{1, \dots, N\}\} \quad (17)$$

to tell the maximum foreseen AoI from time  $t$ , depending on which path  $\hat{i}_{t-1}$  was sent by the last driver (if any) at time  $t-1$  to sample. We next see how does  $\mathbb{A}(t)$  update for all the paths.

At the beginning of time  $t \in \{0, \dots, T\}$ , the provider decides online pricing set

$$\mathbb{P}(\mathbb{A}(t)) = \{p_{i,t}(\mathbb{A}(t)) \in [0, D_i] | i \in \{1, 2, \dots, N\}\} \quad (18)$$

based on the foreseen AoI set  $\mathbb{A}(t)$ . Given the online prices to different paths, a driver with cost sensitivity  $x$  (if arrives at time  $t$ ) finds path  $i$  appealing as long as the price there can justify his travel cost, i.e., his utility of travelling path  $i$

$$U_{i,t}(\mathbb{A}(t)) = p_{i,t}(\mathbb{A}(t)) - xD_i \quad (19)$$

is no less than 0, which depends on the pricing set  $\mathbb{P}(\mathbb{A}(t))$  and the driver's private travel cost sensitivity  $x$ . If he finds multiple paths appealing, he will optimally accept the path  $\hat{i}_t$  with maximum net payoff, i.e.,

$$\hat{i}_t = \begin{cases} \arg \max_{i \in \{1, \dots, N\}} U_{i,t}(\mathbb{A}(t)), & \text{if } \max_i U_{i,t}(\mathbb{A}(t)) \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

If there is no driver arrival at time  $t$  or this driver does not consider to sample any of the  $N$  non-shortest paths, we set  $\hat{i}_t = 0$ . Based on (20), we update all the paths' next foreseen AoI as:

$$A_i(t+D_i+1) = \begin{cases} D_i, & \text{if } i = \hat{i}_t, \\ A_i(t+D_i)+1, & \text{if } i \neq \hat{i}_t. \end{cases} \quad (21)$$

Based on the above model extension, we are ready to apply the MDP formulation as [28] and [41] again to formulate the provider's online pricing problem below.

- *States:* We define the state of the new MDP in slot  $t$  by  $\mathbb{S}_t = (\mathbb{A}(t), \hat{i}_{t-1}, s(t-1))$ . Similarly,  $s(-1) = 0$  and  $\hat{i}_{-1} = 0$  at time  $t = 0$ .
- *Actions:* The action of the MDP in slot  $t$  is the price set  $\mathbb{P}(\mathbb{A}(t))$  defined in (18).
- *Transition probabilities:* Compared to the state transitions (8) in single-path, the state space of  $\mathbb{S}_t$  now increases exponentially with  $N$ , because every path has probability to be sampled at each time slot. Denote the probability that path  $i$  is accepted by an arrived driver to sample (i.e.,  $i = \hat{i}$ ) by  $q_i(t)$ . The state  $\mathbb{S}_{t+1} = (\mathbb{A}(t+1), \hat{i}_t, s(t))$  at time slot  $t+1$  can change into

$$\begin{cases} \mathbb{S}_{t+1}^a = (\mathbb{A}(t+1), 0, 0), & \text{with probability } \mathbf{Q}_a(t) = 1 - \mathbb{E}[s(t)|s(t-1)], \\ \mathbb{S}_{t+1}^b = (\mathbb{A}(t+1), 0, 1), & \text{with probability } \mathbf{Q}_b(t) = \mathbb{E}[s(t)|s(t-1)] \left(1 - \sum_{i=1}^N q_i(t)\right), \\ \mathbb{S}_{t+1}^i = (\mathbb{A}(t+1), i, 1), & \text{with probability } \mathbf{Q}_i(t) = \mathbb{E}[s(t)|s(t-1)] q_i(t), \end{cases} \quad (22)$$

where the dynamics of  $\mathbb{A}(t+1)$  is defined in (21). There are totally  $N+2$  outcomes: no current arrival,



arrival to not sample, and arrival to sample path  $i \in \{1, \dots, N\}$ .

- *Cost*: Let  $V(S_t, P(A(t)))$  be the immediate cost of the MDP if action  $P(A(t))$  is taken in slot  $t$  under state  $S_t$ , which concerns about the maximum foreseen AoI among all the paths as [15] and [17]:

$$V(S_t, P(A(t))) = \max_i \{A_i(t + D_i)\} + \sum_{i=1}^N Q_i(t) p_{i,t}(A(t)). \quad (23)$$

Thus, we extent problem (10) by considering multiple paths: at the beginning of time slot  $t$ ,

$$C_t^*(S_t) = \min_{P(A(t))} \sum_{\tau=t}^T \rho^{\tau-t} V(S_\tau, P(A(\tau))), \quad (24)$$

s.t. (1), (17) – (21) and (23).

Similar to (11), (24) can be rewritten as

$$C_t^*(S_t) = \min_{P(A(t))} V(S_t, P(A(t))) + \rho E_{S_{t+1}} [C_{t+1}^*(S_{t+1})], \quad (25)$$

where the cost-to-go is

$$E_{S_{t+1}} [C_{t+1}^*(S_{t+1})] = Q_a(t) C_{t+1}^*(S_{t+1}^a) + Q_b(t) C_{t+1}^*(S_{t+1}^b) + \sum_{i=1}^N Q_i(t) C_{t+1}^*(S_{t+1}^i)$$

according to transition probabilities (22). Note that drivers' random choice model in  $q_i(t)$  among  $N$  paths incurs totally  $(N+2)^{T-\max_i\{D_i\}}$  cost functions to iteratively (25) much more difficult than (11).

Existing works (e.g., [14], [15], [33]) also formulate dynamic program to design optimal scheduling policies for multi-channel network to minimize peak or average weighted AoI. However, the system space of actions only linearly increases in the path number  $N$  and time horizon  $T$ . However, in our problem (25), the system space increases exponentially. In the following, we first reduce multi-path pricing to single-path at a time and then propose our backward-clustered computation to design low-complexity algorithm.

### B. Reducing Multi-Path Pricing to Single-Path At a Time

Note that Propositions 1 and 2 can be extended for this multi-path model and ensure existence and uniqueness of the online pricing solution to problem (25). Here we need to overcome the curse of dimensionality in both the path number  $N$  and time horizon  $T$ . In this subsection, we prove that we can reduce the  $N$  spatial searching dimensions greatly by reducing multi-path pricing to single-path pricing at each time  $t$ .

*Proposition 3*: To solve problem (25), it is optimal to only price path  $\hat{i}_t^*$  out of  $N$  paths at each time  $t$ , where

$$\hat{i}_t^* = \arg \max_{i \in \{1, \dots, N\}} \{A_i(t + D_i)\}. \quad (26)$$

The proof of Proposition 3 is given in our technical report [46]. Proposition 3 shows that it is optimal to only price a single path at a time. Intuitively, if positive prices are given to more than one path, the driver will still choose only one to sample. Then the provider need to give higher price for the target path given the competitive prices from others, which incurs unnecessary cost.

Further, Proposition 3 tells that we should target at the path with the maximum foreseen AoI  $\max_i \{A_i(t + D_i)\}$  instead of largest current AoI  $\max_i \{A_i(t)\}$ . This result is different from [14] and [15], because the current pricing decision in our problem can only help reduce the AoI after a path-dependent time delay.

Thanks to Proposition 3, we only need to design the positive price  $p_{\hat{i}_t^*, t}(A(t))$  for the target path  $\hat{i}_t^*$  in (26) instead of  $P(A(t))$  at each time  $t$ . Then the cost function in (23) is simplified to:

$$V(S_t, p_{\hat{i}_t^*, t}(A(t))) = A_{\hat{i}_t^*}(t + D_{\hat{i}_t^*}) + Q_{\hat{i}_t^*}(t) p_{\hat{i}_t^*, t}(A(t)). \quad (27)$$

Accordingly, (25) can also be simplified as

$$C_t^*(S_t) = \min_{p_{\hat{i}_t^*, t}(A(t))} V(S_t, p_{\hat{i}_t^*, t}(A(t))) + \rho E_{S_{t+1}} [C_{t+1}^*(S_{t+1})], \quad (28)$$

where the cost-to-go is

$$E_{S_{t+1}} [C_{t+1}^*(S_{t+1})] = Q_a(t) C_{t+1}^*(S_{t+1}^a) + Q_b(t) C_{t+1}^*(S_{t+1}^b) + Q_{\hat{i}_t^*}(t) C_{t+1}^*(S_{t+1}^{\hat{i}_t^*}) \quad (29)$$

by letting  $q_i(t) = 0$  for all  $i \in \{0, \dots, N\}$  and  $i \neq \hat{i}_t^*$  in (22). Though the total number of cost functions has been greatly reduced from  $(N+2)^{T-\max_i\{D_i\}}$  to  $3^{T-\max_i\{D_i\}}$  in problem (28), it is still difficult to solve, and the direct application of backward induction approach in Section IV cannot help. Here we follow a random pattern to alternate and traverse the target path ( $\hat{i}_t$ ) to price over time, and we still need to adapt pricing online to a huge AoI evolution set  $A(t)$  with an exponentially increasing state space in  $T$ .

### C. Backward-Clustered Computation for Low-Complexity Approximation Algorithm

In this subsection, we propose a new innovative simplification approach called backward-clustered computation method to solve problem (28). Different from the repeated long-term cost functions under the same sampled paths we analyzed in Fig. 3 (a), the huge number of cost functions across paths are non-repeated themselves. To reduce the exponentially increasing state space, at current time  $t$ , we cluster those states with the same number of future sampled paths from time  $t$  to be one approximated state. By applying this approximation for clustered functions, we no longer count and compute all the long-term cost functions, and we do not need to consider  $\hat{i}_t^*$  any more.

More specifically, at current time  $t$ , we propose to use  $A^{(0)}(t)$  and  $C_t^{(0)}(A^{(0)}(t), s(t-1))$  to approximate foreseen

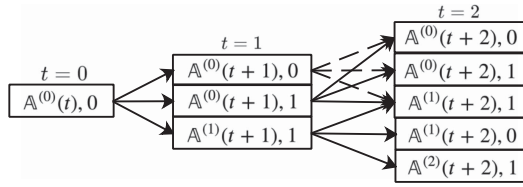


Fig. 5. An illustrative example of approximated backward-clustered process with three decision slots ( $t = 0, 1, 2$  under  $T - \max_i\{D_i\} = 2$ ) in the multi-path scenario.

AoI evolution set  $\mathbb{A}(t)$  and its cost function  $C_t^*(S_t)$  in (28), respectively, where the superscript 0 means the number of the sampled paths from time  $t$  is still 0 at time  $t$ . Note that the initial approximated AoI set  $\mathbb{A}^{(0)}(t)$  equals  $\mathbb{A}(t)$ . Then the three possible states at time  $t+1$  in (22) can be approximated to  $(\mathbb{A}^{(0)}(t+1), 0)$ ,  $(\mathbb{A}^{(0)}(t+1), 1)$  and  $(\mathbb{A}^{(1)}(t+1), 1)$ , respectively. Here only  $\mathbb{A}^{(1)}(t+1)$  with superscript 1 tells that one path is sampled from time  $t$  to  $t+1$ . Take the clustered approximation of  $C_t^*(S_t)$  in (28) with three decision slots ( $t = 0, 1, 2$  under  $T - \max_i\{D_i\} = 2$ ) as an illustrative example, and Fig. 5 shows the approximated branching network from  $t = 0$  to 2 as explained below.

- At  $t = 2$ , there are only 3 approximate AoI sets (i.e.,  $\mathbb{A}^{(0)}(t+2)$ ,  $\mathbb{A}^{(1)}(t+2)$  and  $\mathbb{A}^{(2)}(t+2)$ ) in Fig. 5, because there are at most two paths being sampled since  $t = 0$ , and we only have 5 approximated states instead of  $3^{T-D} = 9$ .
- At  $t = 1$ , our two approximated states with superscript 0 (i.e.,  $(\mathbb{A}^{(0)}(1), 1)$  with last arrival and  $(\mathbb{A}^{(0)}(1), 0)$  without last arrival) have the same branching structure to  $(\mathbb{A}^{(0)}(2), 1)$ ,  $(\mathbb{A}^{(0)}(2), 0)$  and  $(\mathbb{A}^{(1)}(2), 1)$  at  $t = 2$ . Thus, it suffices to cluster  $(\mathbb{A}^{(0)}(1), 1)$  and  $(\mathbb{A}^{(0)}(1), 0)$  by only expanding one of them to  $t = 2$  in Fig. 5.

Note that the original states  $S_{t+1}^b$  and  $S_{t+1}^i$  in (29) cannot branch to the same state at time  $t+2$ , because the path  $\hat{i}_t^*$  is sampled during different time slots, i.e.,  $t = 0$  and  $t = 1$ , respectively. In our approximation of backward-clustered computation, the actual AoI of a sampled path is mildly enlarged to the approximated AoI. Yet this error is always bounded by the elapsed time since current time  $t$  and will not take effect until this path is priced to sampled again. By clustering the approximated cost functions with the same number of sampled paths from time  $t$  to the whole branching network, we successfully reduce the number of long-term cost functions to compute from  $3^{T-\max_i\{D_i\}}$  to around  $(T - \max_i\{D_i\})^2$ , which can be solved using backward induction.

Regarding the pricing solution, similar to (12), the approximation online pricing  $p_{\hat{i}_t^{(0)}, t}(\mathbb{A}^{(0)}(t))$  here satisfies the first-order condition of (28), yet the cost functions are replaced by their approximation and  $\hat{i}_t^{(0)} = \max_i\{A_i^{(0)}(t + D_i)\}$ . Now we present Algorithm 3 to efficiently return the approximation online pricing to (28). We apply our backward-clustered computation to calculate approximation cost functions from  $t = T - \max_i D_i$  to  $t = 0$ :

- From step 3 to step 12, we backward compute all the possible long-term cost functions from time slot

**Algorithm 3:** Polynomial-time backward-clustered computation of approximation online pricing to (28).

- 1: Observe current AoI and predict maximum foreseen AoI set  $\mathbb{A}^{(0)}(t)$ ;
- 2: Initialize  $\hat{i}_t^{(0)} = \arg \max_i\{A_i^{(0)}(t + D_i)\}$  and  $k = T - \max_i\{D_i\} - t$ ;
- 3: **while**  $k \geq 1$  **do**
- 4:     Initialize  $A_i^{(0)}(t + D_i + k) = A_i^{(0)}(t + D_i) + k$  for all  $i \in \{1, \dots, N\}$ ;
- 5:     **for**  $0 \leq j \leq k$  **do**
- 6:         Obtain  $\hat{i}_{t+k}^{(j)} = \arg \max_i\{A_i^{(j)}(t + D_i + k)\}$ ;
- 7:         Compute  $p_{\hat{i}_{t+k}^{(j)}, t+k}(\mathbb{A}^{(j)}(t + k))$  using binary search with error  $\epsilon$ ;
- 8:         Compute  $C_{t+k}^{(j)}(\mathbb{A}^{(j)}(t + k), 1)$  and  $C_{t+k}^{(j)}(\mathbb{A}^{(j)}(t + k), 0)$  according to (28) from last decision slot;
- 9:         Update  $A_{\hat{i}_{t+k}^{(j)}}^{(j+1)}(t + D_{\hat{i}_{t+k}^{(j)}} + k) = D_{\hat{i}_{t+k}^{(j)}} + \frac{k-j}{2}$ ;
- 10:     **end for**
- 11:      $k = k - 1$
- 12: **end while**
- 13: Apply binary search of  $p_{\hat{i}_t^{(0)}, t}(\mathbb{A}^{(0)}(t))$  with error  $\epsilon$ ;
- 14: **return**  $p_{\hat{i}_t^{(0)}, t}(\mathbb{A}^{(0)}(t))$  for current time  $t$

$T - \max_i D_i$  to  $t + 1$ , and finally apply binary search to derive  $p_{\hat{i}_t^{(0)}, t}(\mathbb{A}^{(0)}(t))$  in step 13.

- At each future time slot  $t + k$ , we first calculate the approximated foreseen AoI set  $\mathbb{A}^{(j)}(t + k)$  for all  $j \in \{0, \dots, k\}$ . At  $j = 0$ , all the paths are not sampled, such that  $\mathbb{A}^{(0)}(t + k)$  is initialized at step 4. Then we choose to update the AoI of path  $\hat{i}_{t+k}^{(j)}$  to increase  $j$  to  $j + 1$  at  $j$ -th loop in step 9.
- In steps 7 and 8, we compute the corresponding approximation price and cost functions of each AoI set  $\mathbb{A}^{(j)}(t + k)$ .

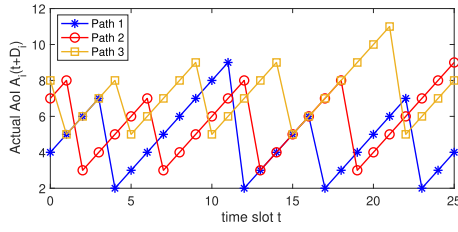
Note that in step 9, we update the AoI of path  $\hat{i}_{t+k}^{(j)}$  to  $D_{\hat{i}_{t+k}^{(j)}} + \frac{k-j}{2}$  because this  $j$ -th updated path can be sampled at any time from  $t + j$  to  $t + k$ , such that we take the average of all the possible AoI based on our linear model in (23).

As a counter-part of (28), we denote  $\hat{C}_t(\mathbb{A}(t), s(t-1))$  as the approximated cost function under our approximation pricing, and we next prove that it has bounded performance gap to the optimum in the following theorem.

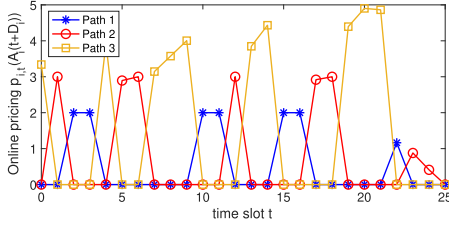
**Theorem 2:** At any time  $t$ , our approximation Algorithm 3 takes only polynomial time complexity  $O((T - \max_i\{D_i\})^2 \log_2(\frac{\max_i\{D_i\}}{\epsilon}))$  to return the approximation online pricing  $p_{\hat{i}_t^{(0)}, t}(\mathbb{A}^{(0)}(t))$ , and the complexity does not depend on the path number  $N$ . As compared to the cost optimal objective  $C_t^*(\mathbb{A}(t), s(t-1))$  in (28), Algorithm 3's cost objective  $\hat{C}_t(\mathbb{A}(t), s(t-1))$  achieves the following approximation error in the worst case:

$$\begin{aligned} & |\hat{C}_t(\mathbb{A}(t), s(t-1)) - C_t^*(\mathbb{A}(t), s(t-1))| \\ & < \frac{\rho^{g(N)+1} - \rho^{T-\max_i\{D_i\}-t+1}}{(1-\rho)^2}, \end{aligned} \quad (30)$$

and this error increases with  $T$  but decreases with  $N$ .



(a) Evolution of foreseen AoI in all the three paths under the approximation online pricing  $p_{i_t^{(0)},t}(\hat{A}^{(0)}(t))$ ,  $t \in \{0, \dots, T - \max_i D_i = 25\}$ .



(b) Online approximation pricing policy returned by Algorithm 3.

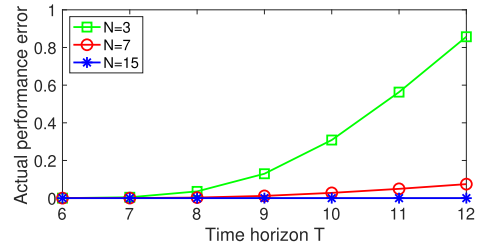
Fig. 6. The evolution of foreseen AoI set  $\hat{A}(t)$  and online approximation price  $p_{i_t^{(0)},t}(\hat{A}^{(0)}(t))$  returned by Algorithm 3 from  $t = 0$  to 25. Here we set  $\hat{A}(0) = \{2, 4, 3\}$ ,  $D = \{2, 3, 5\}$  with  $N = 3$ ,  $T = 30$ ,  $\rho = 0.85$ , probabilities  $\alpha = 0.8$ ,  $\beta = 0.6$  for the Markov chain in Fig. 2, and the cost CDF  $F(x)$  satisfies a truncated normal distribution with mean 0.6 and variance 0.7. (a) Evolution of foreseen AoI in all the three paths under the approximation online pricing  $p_{i_t^{(0)},t}(\hat{A}^{(0)}(t))$ ,  $t \in \{0, \dots, T - \max_i D_i = 25\}$  (b) Online approximation pricing policy returned by Algorithm 3.

The proof of Theorem 2 is given in [46]. The error bound (30) is derived based on  $g(N)$ , which is an increasing function of  $N$  to return the updating time cycle to sample path  $\hat{i}_t^*$  again. The complexity of our Algorithm 3 does not depend on path number  $N$  and can thus fit large-scale road networks. Actually, the approximation error on the right hand side of (30) is small most of the time. Perhaps surprisingly, it decreases with a larger path number  $N$ . This is because individual paths become similar to each other in a larger path choice pool, and our backward-clustered computation in Algorithm 3 only counts the number of sampled paths (without checking which paths) for approximating cost functions.

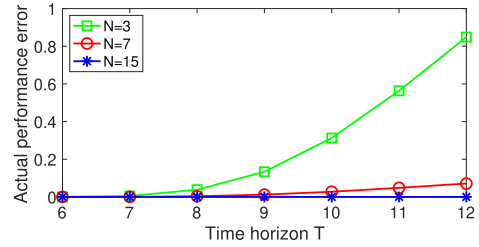
We create a typical instance but with  $N = 3$  paths in Fig. 6 to run the online approximation pricing  $p_{i_t^{(0)},t}(\hat{A}^{(0)}(t))$  and corresponding foreseen AoI set  $\hat{A}(t)$ ,  $\forall t \in \{0, \dots, T = 25\}$  returned by Algorithm 3. Here we set the travel delay set  $D = \{2, 3, 5\}$  for the three paths, whose initial AoI  $\hat{A}(0) = \{2, 4, 3\}$ , and the cost distributions of each driver to follow a truncated normal distribution with mean 0.6 and variance 0.7.

Fig. 6(a) shows the evolution of the foreseen AoI of all the three paths, and the maximum AoI among all the paths are well controlled thanks to our online pricing solution in Fig. 6 (b). We can also see that the platform will select the path with the maximum AoI to price, which is consistent with Proposition 3. In this case, the online pricing follows a  $D_i$ -delayed pattern of the actual AoI over time.

Finally, unlike (30)'s error for the worst-case, we run simulations to examine Algorithm 3's actual performance loss as compared to the optimum, versus time horizon  $T$  and path



(a) Actual performance error of truncated normal distribution with mean 0.6 and variance 0.7.



(b) Actual performance error of truncated logistic distribution with location parameter 0.6 and scale 1.

Fig. 7. The actual performance loss of our Algorithm 3 versus path number  $N$  and time horizon  $T$  under truncated normal and truncated logistic distributions. Here we vary the path numbers  $N = 3, 7$  and 15, time horizon  $T = \{6, \dots, 12\}$ , and consider the time  $t = 0$ 's online decision making and set maximum travel delay among all paths as  $\max_i \{D_i\} = 5$ . (a) Actual performance error of truncated normal distribution with mean 0.6 and variance 0.7 (b) Actual performance error of truncated logistic distribution with location parameter 0.6 and scale 1.

number  $N$  in Fig. 7. Here we generally consider initial  $t = 0$ 's online decision making, then the cost objective  $\hat{C}_0(\hat{A}(0)), 0$  uses the approximation pricing returned by Algorithm 3. While the optimal cost function  $C_0^*(\hat{A}(0)), 0$  in (28) uses the optimal pricing returned by traditional iterative computation methods, whose complexity is high and only allows us to keep  $T \leq 12$  for simulation. In Fig. 7 with almost the same Fig. 7 (a) and Fig. 7(b), we find that effect of  $F(x)$  to the performance error is limited, which shows our scalable Algorithm 3's robustness to different cost distributions. The performance error is mild from the optimum under our algorithm. This result is also consistent with Theorem 2 to show the approximation error increase in time horizon  $T$  and decrease in path number  $N$ .

## VI. CONCLUSION

In this paper, we have proposed online pricing for a content provider to economically reward drivers for diverse routing and control the actual AoI dynamics over time and spatial path domains. This online pricing optimization problem should be solved without knowing drivers' costs and even arrivals, and is intractable due to the curse of dimensionality in both time and space. If there is only one non-shortest path, we leverage the Markov decision process (MDP) techniques to analyze the problem. Accordingly, we design a linear-time algorithm for returning optimal online pricing, where a higher pricing reward is needed for a larger AoI. We also show that our algorithm is applicable to infinite time horizon. If there

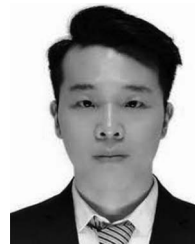
are a number of non-shortest paths, we prove that pricing one path at a time is optimal, yet it is not optimal to choose the path with the largest current AoI. Then we propose a new backward-clustered computation method and develop an approximation algorithm to alternate different paths to price over time. This algorithm has only polynomial-time complexity, and its complexity does not depend on the number of paths. Perhaps surprisingly, our analysis of approximation ratio suggests that our algorithm's performance approaches closer to optimum if more paths are involved to sample.

We can consider some possible future works directions to extend this work. Recall that this paper focuses on the min-max foreseen AoI optimization in problem (25) if there are a number of non-shortest paths. We can extend to study the min-sum AoI optimization problem to minimize the total AoI across all the paths. Our backward-clustered computation method should still work and we should still choose to pricing one path at a time. Yet we may not choose the path with the largest foreseen AoI in Proposition 3. Moreover, we can also extend our analysis to deal with the random travel delay instead of deterministic delay in each path, where we need to further take expectation in the online optimization. The evaluation about the performance of our algorithms with real data is another point in our next step.

## REFERENCES

- [1] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'hamed, "Sparse mobile crowdsensing: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 161–167, Jul. 2016.
- [2] D. Asghar, M. Zubair, and D. Ahmad, "A review of location technologies for wireless mobile location-based services," *J. Amer. Sci.*, vol. 10, no. 7, pp. 110–118, 2014.
- [3] R. Bhoraskar, N. Vankadhara, B. Raman, and P. Kulkarni, "Wolverine: Traffic and road condition estimation using smartphone sensors," in *Proc. IEEE 4th Int. Conf. Commun. Syst. Netw.*, 2012, pp. 1–6.
- [4] Y. Li, C. Courcoubetis, and L. Duan, "Recommending paths: Follow or not follow?," in *Proc. IEEE INFOCOM -IEEE Conf. Comput. Commun.*, 2019, pp. 928–936.
- [5] D. R. Figueiredo, M. Garetto, and D. Towsley, "Exploiting mobility in ad-hoc wireless networks with incentives," Tech. Rep., 2004. [Online]. Available: <https://web.cs.umass.edu/publication/docs/2004/UM-CS-2004-066.pdf>
- [6] L. Xiao, T. Chen, C. Xie, H. Dai, and H. V. Poor, "Mobile crowdsensing games in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1535–1545, Feb. 2018.
- [7] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2542–2550.
- [8] S. Xu, X. Chen, X. Pi, C. Joe-Wong, P. Zhang, and H. Y. Noh, "iLOCuS: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing," *IEEE Trans. Mobile Comput.*, vol. 19, no. 8, pp. 1831–1847, Aug. 2020.
- [9] Y. Lai, Y. Xu, D. Mai, Y. Fan, and F. Yang, "Optimized large-scale road sensing through crowdsourced vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3878–3889, Apr. 2022.
- [10] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE INFOCOM*, 2012, pp. 2731–2735.
- [11] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 1583–1587.
- [12] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations Trends Netw.*, vol. 12, no. 3, pp. 162–259, 2017.
- [13] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Trans. Mobile Comput.*, vol. 19, no. 12, pp. 2903–2915, Dec. 2020.
- [14] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *Proc. IEEE INFOCOM -IEEE Conf. Comput. Commun. Workshops*, 2018, pp. 136–141.
- [15] 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 Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.
- [16] C. Li, S. Li, Y. Chen, T. Hou, and W. Lou, "Minimizing age of information under general models for IoT data collection," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2256–2270, Oct.–Dec. 2020.
- [17] Q. He, D. Yuan, and A. Ephremides, "On optimal link scheduling with min-max peak age of information in wireless systems," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–7.
- [18] C. Zhou et al., "Deep RL-based trajectory planning for AoI minimization in UAV-assisted IoT," in *Proc. IEEE 11th Int. Conf. Wireless Commun. Signal Process.*, 2019, pp. 1–6.
- [19] Z. Jia, X. Qin, Z. Wang, and B. Liu, "Age-based path planning and data acquisition in UAV-assisted IoT networks," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2019, pp. 1–6.
- [20] W. Li, L. Wang, and A. Fei, "Minimizing packet expiration loss with path planning in UAV-assisted data sensing," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1520–1523, Dec. 2019.
- [21] E. Altman, R. El-Azouzi, D. S. Menasche, and Y. Xu, "Forever young: Aging control for hybrid networks," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2019, pp. 91–100.
- [22] M. Zhang, A. Arafat, J. Huang, and H. V. Poor, "Pricing fresh data," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1211–1225, May 2021.
- [23] J. He, Q. Ma, M. Zhang, and J. Huang, "Optimal fresh data sampling and trading," in *Proc. IEEE 19th Int. Symp. Model. Optim. Mobile, Ad hoc, Wireless Netw.*, 2021, pp. 1–8.
- [24] N. Modina, R. El Azouzi, F. De Pellegrini, D. S. Menasche, and R. Figueiredo, "Joint traffic offloading and aging control in 5 G IoT networks," *IEEE Trans. Mobile Comput.*, to be published, doi: [10.1109/TMC.2022.3154089](https://doi.org/10.1109/TMC.2022.3154089).
- [25] X. Wang and L. Duan, "Dynamic pricing for controlling age of information," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 962–966.
- [26] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst., Man, Cybern., Part B*, vol. 38, no. 4, pp. 943–949, Aug. 2008.
- [27] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proc. Amer. Control Conf.*, 2010, pp. 1568–1573.
- [28] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2014.
- [29] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the complexity of solving Markov decision problems," in *Proc. 11th Conf. Uncertainty Artif. Intell.*, 1995, pp. 394–402.
- [30] D. Zhou, M. Sheng, J. Luo, R. Liu, J. Li, and Z. Han, "Collaborative data scheduling with joint forward and backward induction in small satellite networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3443–3456, May 2019.
- [31] Z. Xiong, J. Zhao, D. Niyato, R. Deng, and J. Zhang, "Reward optimization for content providers with mobile data subsidization: A hierarchical game approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2363–2377, Oct.–Dec. 2020.
- [32] Q. Liu, H. Zeng, and M. Chen, "Minimizing age-of-information with throughput requirements in multi-path network communication," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2019, pp. 41–50.
- [33] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2019, pp. 121–130.
- [34] C. Fang, H. Lu, Y. Hong, S. Liu, and J. Chang, "Dynamic pricing for electric vehicle extreme fast charging," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 531–541, Jan. 2021.
- [35] T. Ramkha and S. D. Boyles, "Dynamic pricing in discrete time stochastic day-to-day route choice models," *Transp. Res. Part B: Methodological*, vol. 92, pp. 104–118, 2016.
- [36] A. A. Baktayan, I. A. Al-Baltah, and A. A. Abd Ghani, "Intelligent pricing model for task offloading in unmanned aerial vehicle mounted mobile edge computing for vehicular network," *J. Commun. Softw. Syst.*, vol. 18, no. 2, pp. 111–123, 2022.
- [37] G. Yu, J. Hu, C. Zhang, L. Zhuang, and J. Song, "Short-term traffic flow forecasting based on Markov chain model," in *Proc. IEEE IV Intell. Veh. Symp. Proc.*, 2003, pp. 208–212.

- [38] E. Meijer and J. Rouwendal, "Measuring welfare effects in models with random coefficients," *J. Appl. Econometrics*, vol. 21, no. 2, pp. 227–244, 2006.
- [39] B.-D. Kim, R. C. Blattberg, and P. E. Rossi, "Modeling the distribution of price sensitivity and implications for optimal retail pricing," in *Perspectives On Promotion And Database Marketing: The Collected Works of Robert C Blattberg*. Singapore: World Sci. Publishing Co Pte Ltd, 2010, pp. 283–295.
- [40] C. Ewerhart, "Regular type distributions in mechanism design and  $\rho$ -concavity," *Econ. Theory*, vol. 53, no. 3, pp. 591–603, 2013.
- [41] Y.-P. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 561–565.
- [42] D. E. Knuth, "Optimum binary search trees," *Acta Informatica*, vol. 1, no. 1, pp. 14–25, 1971.
- [43] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. Hoboken, NJ, USA: Wiley, 2012.
- [44] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.
- [45] P. Kakumanu, "Continuously discounted Markov decision model with countable state and action space," *Ann. Math. Statist.*, vol. 42, no. 3, pp. 919–926, 1971.
- [46] H. Li and L. Duan, "Online pricing incentive to sample fresh information," Tech. Rep., 2022. [Online]. Available: <https://arxiv.org/abs/2209.08711>
- [47] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.
- [48] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 2, pp. 267–278, Mar. 2010.



**Hongbo Li** received the B.S. degree in electronics and electric engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019. He is currently working toward the Ph.D. degree with the Pillar of Engineering Systems and Design, Singapore University of Technology and Design, Singapore. His research interests include network economics, game theory, and mechanism design.



**Lingjie Duan** (Senior Member, IEEE) received the Ph.D. degree from The Chinese University of Hong Kong, Hong Kong, in 2012. He is currently an Associate Professor of engineering systems and design with the Singapore University of Technology and Design, Singapore. In 2011, he was a Visiting Scholar with the University of California at Berkeley, Berkeley, CA, USA. His research interests include network economics and game theory, cognitive and green networks, and energy harvesting wireless communications. He is the Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

He was the Editor of IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He is also the Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on Human-in-the-Loop Mobile Networks, and *IEEE Wireless Communications Magazine*. He was the recipient of the SUTD Excellence in Research Award in 2016 and 10th IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2015.