Dynamic Optimization of Random Access in Deadline-Constrained Broadcasting

Aoyu Gong, Yijin Zhang, *Senior Member, IEEE*, Lei Deng, Fang Liu, Jun Li, *Senior Member, IEEE*, and Feng Shu, *Member, IEEE*

***Abstract*—This paper considers dynamic optimization of random access in deadline-constrained broadcasting with frame-synchronized traffic. Under the non-retransmission setting, we define a dynamic control scheme that allows each active node to determine the transmission probability based on the local knowledge of current delivery urgency and contention intensity (i.e., the number of active nodes). For an idealized environment where the contention intensity is completely known, we develop a Markov Decision Process (MDP) framework, by which an optimal scheme for maximizing the timely delivery ratio (TDR) can be explicitly obtained. For a realistic environment where the contention intensity is incompletely known, we develop a Partially Observable MDP (POMDP) framework, by which an optimal scheme can only in theory be found. To overcome the infeasibility in obtaining an optimal or near-optimal scheme from the POMDP framework, we investigate the behaviors of the optimal scheme for extreme cases in the MDP framework, and leverage intuition gained from these behaviors together with an approximation on the contention intensity knowledge to propose a heuristic scheme for the realistic environment with TDR close to the maximum TDR in the idealized environment. We further generalize the heuristic scheme to support retransmissions. Numerical results are provided to validate our study.**

***Index Terms*—Distributed algorithms, dynamic optimization, random access, delivery deadline**

I. Introduction

A. Background

BROADCASTING is a fundamental operation in distributed wireless systems. With the explosive growth of ultra-reliable low-latency services for the Internet of things (IoT) [1]–[4], such as detection information sharing in unmanned aerial vehicles (UAV) networks, safety message dissemination in vehicular networks, and industrial control in factory automation, deadline-constrained broadcasting has been becoming a research focus in recent years. For such broadcasting, each packet needs to be transmitted within a strict delivery deadline since its arrival and will be discarded if the deadline expires. Hence, *timely delivery ratio* (TDR), defined as the probability that a broadcast packet is successfully delivered to an arbitrary intended receiver within the given delivery deadline, is considered as a critical metric to evaluate the performance of such broadcasting. Note that such broadcasting is important for agents in these applications to obtain as much timely information of the world as possible for making operations that best achieve their application objectives. For example, in UAV networks for collaborative multitarget tracking [5], due to the limited detection capability, each UAV is only able to detect targets situated within a certain area, and has to use the detection information shared by other UAVs to decide on the optimal path to follow in order to cover as many targets as possible.

A canonical deadline-constrained broadcasting scenario is that, under a given traffic pattern, an uncertain set of nodes with new or backlogged packets attempt to transmit before deadline expiration without centralized scheduling. In this scenario, it is expected that the MAC layer behaves differently from what is commonly believed in conventional deadline-unconstrained protocols, as the contention intensity would be jointly determined by the traffic pattern, delivery deadline, and retransmission setting. So, random access mechanisms tailored for this scenario are needed to support efficient channel sharing under deadline, and careful design of access parameters is needed to maximize the TDR.

B. Related Work and Motivation

Many recent literatures [6]–[9] have been dedicated to this issue by assuming no retransmissions, which is commonly adopted in broadcasting due to the lack of acknowledgments or for the sake of energy efficiency. Under saturated traffic, Bae [6], [7] obtained the optimal slotted-ALOHA for broadcasting single-slot packets and optimal p-persistent CSMA for broadcasting multi-slot packets, respectively. Under a discrete-time Geo/Geo/1 queue model, Bae [8] obtained the optimal slotted-ALOHA for broadcasting single-slot packets. Under frame-synchronized traffic, Campolo *et al.* [9] proposed an analytical model for using IEEE 802.11p CSMA/CA to broadcast multi-slot packets, which can be used to obtain the optimal

This work was supported in part by the National Natural Science Foundation of China under Grants 62071236, U22A2002, 62071234, 61902256, in part by the Major Science and Technology plan of Hainan Province under Grant ZDKJ2021022, in part by the Scientific Research Fund Project of Hainan University under Grant KYQD(ZR)-21008, in part by the Fundamental Research Funds for the Central Universities of China (Nos. 30920021127, 30921013104), in part by Future Network Grant of Provincial Education Board in Jiangsu, and in part by the Open Research Fund of State Key Laboratory of Integrated Services Networks, Xidian University, under Grant ISN22-14. (*Corresponding author: Yijin Zhang*.)

A. Gong, Y. Zhang, and J. Li are with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China. Y. Zhang is also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xian 710071. E-mail: {gongaoyu; yijin.zhang}@gmail.com, jun.li@njust.edu.cn.

L. Deng is with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China. E-mail: ldeng@szu.edu.cn.

F. Liu is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. E-mail: lf015@ie.cuhk.edu.hk.

F. Shu is with the School of Information and Communication Engineering, Hainan University, Haikou 570228, China, and also with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China. E-mail: shufeng0101@163.com.

contention window size. To improve the TDR with the help of retransmissions, Hassan *et al.* [10] investigated the impact of retransmissions on the TDR of IEEE 802.11p CSMA/CA under frame-synchronized traffic, and Bae [11] obtained the optimal slotted-ALOHA with retransmissions under saturated traffic. However, [6]–[8], [11] adopted a static transmission probability and [9], [10] adopted a static contention window size, thus inevitably limiting the maximum achievable TDR.

Other studies on deadline-constrained random access include [12]–[16] for uplink to a common receiver. Under Bernoulli traffic, Bae [12] derived the optimal static transmission probability for maximizing the TDR based on stationary Markov chain modeling. Under frame-synchronized traffic, Deng *et al.* [13] developed an algorithm to recursively analyze the timely throughput for any static transmission probability and characterized the asymptotic behavior of optimal one. However, [12], [13] still restrict their attentions to static access parameters. Using absorbing Markov chain modeling, Bae *et al.* [14] proposed to myopically change the transmission probability when the contention intensity is completely known, which, however, did not account for dynamic programming optimality. Zhao *et al.* [15] proposed to simply double or halve transmission probability based on the channel feedback, which is easily implemented but lacks an explicit optimization goal. Zhang *et al.* [16] proposed to adjust the transmission probability for maximizing the TDR by a joint use of the fixed point iteration and a recursive estimator, but did not utilize all of the observed data. Another type of deadline-constrained access is based on the sequence design [17], [18], where each active node deterministically decides whether to transmit according to the assigned sequence but utilizes no observation to adjust its access behavior.

As such, to enhance the maximum achievable TDR of distributed random access in deadline-constrained broadcasting, it is strongly required to develop a dynamic control scheme that allows each node to adjust its access parameters according to local knowledge of current delivery urgency and contention intensity. Unfortunately, due to random traffic or limited capability on observing the channel status, each node *cannot* obtain a complete knowledge of the current contention intensity in practice, which renders such design a challenging task. So, each node has to estimate the current contention intensity using the information obtained from the observed channel status. A great amount of work has gone into studying such information that can be obtained [16], [19]–[23] under various models and protocols. Our work follows the same direction of [20], [21] to keep an A Posteriori Probability (APP) distribution for the current contention intensity given all past observations and access settings, which is a sufficient statistic for the optimal design [24], but needs to additionally take into account the impact of delivery urgencies. It should be noted that another estimation technique is based on the "Certainty Equivalence" principle [16], [19], [22], [23], which uses simple recursive point estimators to merely estimate the actual value of the current contention intensity, but does not utilize a sufficient statistic for the optimal design. To our best knowledge, this is the first time to study dynamic control for deadline-constrained random access, and the previous estimation approaches [16], [19]–[23] cannot be directly applied here.

Furthermore, it is naturally desirable for this dynamic control to strike a balance between the chance to gain an instantaneous successful transmission and the chance to gain a future successful transmission within the given deadline, which requires reasoning about future sequences of access parameters and observations. So, the dynamic control design under this objective is more challenging than that for maximizing the instantaneous throughput of random access [14], [20]–[23], which is only "one-step look-ahead". By seeing access parameters as actions, in this paper we apply the theories of Markov Decision Process (MDP) and Partially Observable MDP (POMDP) to obtain optimal control schemes for maximizing the TDR. To our best knowledge, this is the first work to apply them in deadline-constrained broadcasting. Although the idea of using MDP and POMDP in the context of random access control is not new [21], [25], [26], our study is different because the delivery urgency plays a nontrivial role in decision making. It not only leads to accounting for time-dependent decision rules, but also leads to a number of new theoretical model properties (see Lemmas 1–3) to answer how the delivery deadline affects optimal policies.

In addition, as solving POMDP is in general computationally prohibitive, it is important to develop a simple control scheme for deadline-constrained broadcasting with little TDR performance loss. However, this design objective is uniquely challenging due to the difficulty in defining a reasonable myopic optimization goal. Note that the instantaneous-throughput-maximization (ITM) goal usually adopted in the literature [14], [20], [21] is no longer a suitable candidate here, because it may significantly degrade the TDR performance especially when the delivery deadline is relatively long and the maximum allowed number of retransmissions is limited. As such, how to utilize the model properties to design a simple control scheme is a major issue that needs to be addressed.

C. Contributions

In this paper, we focus on deadline-constrained broadcasting under frame-synchronized traffic. Such a traffic pattern can capture a number of scenarios in IoT communications [1], [9], [10], [13], [14], [27], [28] where each node has periodic-*i.i.d.* packet arrivals. Our contributions are as follows.

1) For the commonly adopted non-retransmission setting, we generalize slotted-ALOHA to define a dynamic control scheme, i.e., a deterministic Markovian policy, which allows each active node to determine the current transmission probability with certainty based on its current delivery urgency and the knowledge of current contention intensity.
2) For an idealized environment where the contention intensity is completely known, we develop an analytical framework based on the theory of MDP, which leads to an optimal control scheme by applying the backward induction algorithm. We further show it is indeed optimal over all types of policies for this environment.
3) For a realistic environment where the contention intensity is incompletely known, we develop an analytical

TABLE I: Comparison of the proposed schemes and previously known schemes when being applied to deadline-constrained broadcasting under frame-synchronized traffic. The elementary operation is defined as root finding of one univariate polynomial.

Scheme	Observation Requirement	Retrans. Limit	Complexity
The optimal scheme (idealized), $\widehat{\boldsymbol{\pi}}^*$ (Section III)	idealized	no retrans.	$O(ND)$ operations
The optimal scheme (realistic), $\boldsymbol{\pi}^*$ (Section IV)	realistic	no retrans.	$O(\lvert\mathcal{B}_D\rvert D)$ operations
The heuristic scheme (realistic), $\boldsymbol{\pi}^{\text{heu}}$ (Section V)	realistic	no retrans.	$O(1)$ operation (closed-form formula)
The heuristic retrans.-based scheme (realistic), $\boldsymbol{\pi}^{\text{heuR}}$ (Section VI)	realistic	an arbitrary number of retrans.	$O(1)$ operation (closed-form formula)
The ITM scheme ([14], [20], [21])	realistic	no retrans./no limit	$O(\lvert\mathcal{B}_D\rvert)$ operations
The static scheme ([6], [8], [11], [13])	realistic	an arbitrary number of retrans.	$O(1)$ operation

framework based on the theory of POMDP, which can in theory lead to an optimal control scheme by backward induction. We also show it is indeed optimal over all types of policies for this environment.

4) To overcome the infeasibility in obtaining an optimal or near-optimal control scheme from the POMDP framework, we investigate the behaviors of the optimal control scheme for two extreme cases in the idealized environment and use these behaviors as clues to design a simple heuristic control scheme (without need to solve any dynamic programming equations) for the realistic environment with TDR close to the maximum achievable TDR in the idealized environment. In addition, we propose an approximation on the knowledge of contention intensity to further simplify this heuristic scheme.

Note that, although the MDP framework in the idealized environment has limited applicability as the contention intensity *cannot* be completely known in practice, it will serve to provide an upper bound on the maximum achievable TDR in the realistic environment, and serve to provide clues to design a heuristic scheme for the realistic environment.

5) To further improve the TDR in the realistic environment, we generalize the proposed heuristic scheme to support retransmissions.

A comparison of the proposed schemes and previously known schemes when being applied to deadline-constrained broadcasting under frame-synchronized traffic is summarized in Table I. Previously known schemes [6], [8], [11], [13] adopted the static transmission probability and [14], [20], [21] adjusted the transmission probability for ITM merely relying on the knowledge of contention intensity. In contrast, our schemes account for the local knowledge of current delivery urgency and contention intensity simultaneously to adjust the transmission probability, which can yield better performance. Moreover, our schemes can be applied to the realistic environment, can support an arbitrary number of retransmissions, and still have low complexity.

The remainder of this paper is organized as follows. For the non-retransmission setting, the network model, protocol design, and problem formulation are specified in Section II. Optimal schemes for the idealized and realistic environments are studied in Sections III and IV, respectively. The proposed heuristic scheme for the realistic environment is presented in Section V. In Section VI, we generalize this heuristic scheme to support retransmissions. Numerical results with respect to a wide range of configurations are provided in Section VII. Conclusions are given in Section VIII.

II. System Model and Problem Formulation

A. Network Model

Consider a wireless system with global synchronization, where a finite number, $N \geq 2$, of nodes are within the communication range of each other. The global time axis is divided into frames, each of which consists of a finite number, $D \geq 1$, of time slots of equal duration, indexed by $t \in \mathcal{T} \triangleq \{1, 2, \ldots, D\}$. To broadcast the freshest information, at the beginning of each frame, each node independently generates a packet to be transmitted with probability $\lambda \in (0, 1]$. We further assume that every packet has a strict delivery deadline D slots, i.e., a packet generated at the beginning of a frame will be discarded at the end of this frame.

By considering random channel errors due to wireless fading effect, we assume that a packet sent from a node is successfully received by an arbitrary other node with probability $\sigma \in (0, 1]$ if it does not collide with other packets, and otherwise is certainly unsuccessfully received by any other node. Due to the broadcast nature, we assume that every packet is neither acknowledged nor retransmitted. Then, at the beginning of slot t of a frame, a node is called an *active node* if it generated a packet at the beginning of the frame and has not transmitted before slot t, but is called an *inactive node* otherwise. Each active node follows a common control scheme for random access, which will be defined in Section II-B, to generate transmission probabilities at the beginnings of different slots. A slot is said to be in the *idle* channel status if no packet is being transmitted, and in the *busy* status otherwise. At the end of a slot, assume that each node is able to be aware of the channel status of this slot.

The values of N, D and λ are all assumed to be completely known in advance to each node.

In this paper, we will mainly focus on the aforementioned system model. The extension to retransmission-based broadcasting will be discussed in Section VI. It should be noted that, for ease of presentation, although we have assumed an identical arrival rate, results obtained in this paper can be readily extended to general cases.

The main notations used in our analysis are listed in Table II.

TABLE II: Main notations.

Notation	Description
N	Number of nodes
D	Delivery deadline (in time slots)
λ	Packet arrival rate
σ	Packet success rate
n_t	Actual number of other active nodes in the view of an arbitrary node at the beginning of slot t
$\mathbf{b}_t$	Activity belief at the beginning of slot t
p_t	Transmission probability at the beginning of slot t
q_t	Status of the tagged node at the beginning of slot t
β_t	State transition function at slot t
r_t	The reward gained at slot t
$\widehat{\pi}_t$	Transmission function (idealized) at slot t
$\widehat{\boldsymbol{\pi}}$	Policy (idealized)
$R^{\widehat{\boldsymbol{\pi}}}$	Expected total reward (idealized)
$\mathrm{TDR}^{\widehat{\boldsymbol{\pi}}}$	TDR under the policy $\widehat{\boldsymbol{\pi}}$
U_t^*	Value function (idealized) at slot t
o_t	Observation on the channel status (realistic) of slot t
ω_t	Observation function (realistic) at slot t
$\mathbf{h}_\lambda$	Initial activity belief (realistic)
θ_t	Bayesian update function (realistic) at slot t
χ_t	Bayesian update normalizing factor (realistic) at slot t
π_t	Transmission function (realistic) at slot t
$\boldsymbol{\pi}$	Policy (realistic)
$R^{\boldsymbol{\pi}}$	Expected total reward (realistic)
$\mathrm{TDR}^{\boldsymbol{\pi}}$	TDR under the policy $\boldsymbol{\pi}$
V_t^*	Value function (realistic) at slot t
(M_t, α_t)	Parameter vector for approximating $\mathbf{b}_t$ (realistic)

A broadcasting scenario with the above assumptions can be found in Fig. 1. There is a communication link between any two UAVs, i.e., all UAVs are within the communication range of each other, and each UAV adopts the protocol proposed in Section II-B to broadcast its detection information to all other UAVs for collaborative multitarget detection. It can be employed in a traffic mapping application [29] where a team of UAVs keep estimating the locations of ground vehicles, and a surveillance application [30] where a team of UAVs keep tracking several targets with uncertainties caused by occlusion or motion blur.

B. Protocol Description

Due to frame-synchronous traffic, at the beginning of an arbitrary slot with at least one active node, we know each active node has the same delivery urgency. To take into account the joint impact of delivery urgency and the knowledge of contention intensity on determining transmission probabilities, a dynamic control scheme for random access in deadline-constrained broadcasting, which can be seen as a generalization of slotted-ALOHA, is formally defined as follows.

Consider an arbitrary frame. Let the random variable n_t taking values in $\mathcal{N} \triangleq \{0, 1, \ldots, N-1\}$ denote the actual number of other active nodes in the view of an arbitrary node at the beginning of slot $t \in \mathcal{T}$. At the beginning of an arbitrary slot $t \in \mathcal{T}$ with at least one active node, we assume

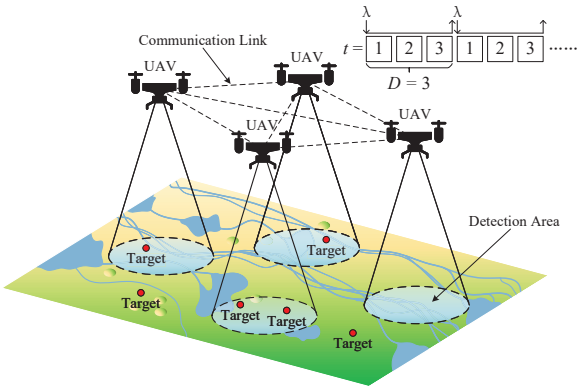

Fig. 1: A deadline-constrained broadcasting scenario for collaborative multitarget detection, which requires each UAV to be equipped with a Ground Moving Target Indicator (GMTI) sensor module and a wireless communication module.

that each active node has the same observation history (for estimating the actual value of n_t) from the environment, and we require each active node to adopt the same transmission probability. Thus, each active node has the same knowledge of the value of n_t based on all past observations from the environment and all past transmission probabilities. Such a knowledge can be summarized by a probability vector $\mathbf{b}_t \triangleq \big(b_t(0), b_t(1), \ldots, b_t(N-1)\big)$, called the *activity belief*, where $b_t(n)$ is the conditional probability (given all past observations from the environment and all past transmission probabilities) that $n_t = n$. Let $\mathcal{B}_t$ denote the set of all possible values of $\mathbf{b}_t$ in $[0,1]^N$ such that $\sum_{n=0}^{N-1} b_t(n) = 1$. Hence, at the beginning of every slot $t \in \mathcal{T}$ with at least one active node, we require each active node to use the values of t and $\mathbf{b}_t$ for determining the value of transmission probability p_t by a transmission function $\pi_t : \mathcal{B}_t \to [0,1]$. An example of the working procedure for $N = 8$, $D = 6$ is illustrated in Fig. 2.

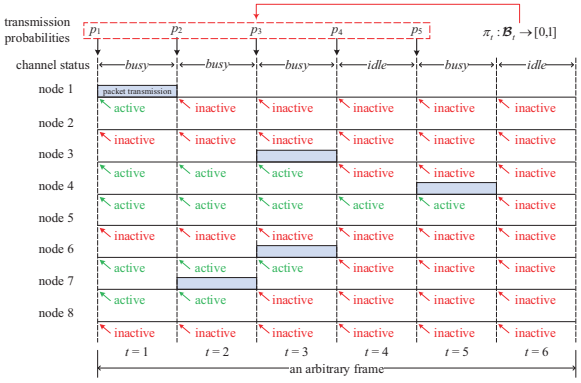

Fig. 2: An example of the working procedure of the dynamic control scheme for $N = 8$, $D = 6$.

We further consider two different environments for active nodes to obtain the value of the activity belief $\mathbf{b}_t$.

1) *Idealized environment:* at the beginning of every slot $t \in \mathcal{T}$ with at least one active node, each active node always has a complete knowledge of the value of n_t, i.e., $\mathbf{b}_t = (0, \ldots, 0, b_t(n) = 1, 0, \ldots, 0)$ if $n_t = n$ actually. Hence,

the transmission function π_t in this environment can be simply written as a function $\widehat{\pi}_t : \mathcal{N} \to [0,1]$, i.e., $\widehat{\pi}_t$ can be chosen from all possible mappings from $\mathcal{N}$ to $[0,1]$. A dynamic control scheme $\widehat{\boldsymbol{\pi}}$ is defined by a sequence of transmission functions for the idealized environment:

$$\widehat{\boldsymbol{\pi}} \triangleq (\widehat{\pi}_1, \widehat{\pi}_2, \ldots, \widehat{\pi}_D), \quad \text{where } \widehat{\pi}_t : \mathcal{N} \to [0,1].$$

Let $\widehat{\boldsymbol{\Pi}}$ denote the set of all possible such schemes.

2) *Realistic environment:* at the beginning of each slot $t \in \mathcal{T}$ with at least one active node, each active node is able to obtain the value of $\mathbf{b}_t$ only based on the characteristic of packet arrivals, all past channel statuses (idle or busy) and all past transmission probabilities, and thus has an incomplete knowledge of the value of n_t. A dynamic control scheme $\boldsymbol{\pi}$ is defined by a sequence of transmission functions for the realistic environment:

$$\boldsymbol{\pi} \triangleq (\pi_1, \pi_2, \ldots, \pi_D), \quad \text{where } \pi_t : \mathcal{B}_t \to [0,1].$$

Let $\boldsymbol{\Pi}$ denote the set of all possible such schemes.

Obviously, the idealized environment is infeasible to implement due to the difficulty in determining the initial actual number of other active nodes and determining the number of nodes involved in each busy slot, whereas the realistic environment can be easily implemented without imposing extra overhead and hardware cost.

C. Problem Formulation

As all active nodes are homogeneous and, accordingly, the performance of them is the same, we can consider an arbitrary node as the tagged node to evaluate the network performance. Let the random variable q_t taking values from $\{0\,(\text{inactive}), 1\,(\text{active})\}$ denote the status of the tagged node at the beginning of slot t.

The optimization problem for the idealized environment can be formulated as

$$(\mathbf{P1}) \quad \max_{\widehat{\boldsymbol{\pi}} \in \widehat{\boldsymbol{\Pi}}} \text{TDR}^{\widehat{\boldsymbol{\pi}}},$$

where $\text{TDR}^{\widehat{\boldsymbol{\pi}}}$ is the TDR under the control scheme $\widehat{\boldsymbol{\pi}}$, i.e.,

$$\text{TDR}^{\widehat{\boldsymbol{\pi}}} \triangleq \sum_{n \in \mathcal{N}} \binom{N-1}{n} \lambda^n (1-\lambda)^{N-1-n} \cdot \mathbb{E}^{\widehat{\boldsymbol{\pi}}} \Big\{ \sum_{t \in \mathcal{T}, q_t = 1} \sigma \widehat{\pi}_t(n_t) \big(1 - \widehat{\pi}_t(n_t)\big)^{n_t} | q_1 = 1, n_1 = n \Big\}.$$

Similarly, the optimization problem for the realistic environment can be formulated as

$$(\mathbf{P2}) \quad \max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \text{TDR}^{\boldsymbol{\pi}},$$

where $\text{TDR}^{\boldsymbol{\pi}}$ is the TDR under the control scheme $\boldsymbol{\pi}$, i.e.,

$$\text{TDR}^{\boldsymbol{\pi}} \triangleq \mathbb{E}^{\boldsymbol{\pi}} \Big\{ \sum_{t \in \mathcal{T}, q_t = 1} \sigma \pi_t(\mathbf{b}_t) \big(1 - \pi_t(\mathbf{b}_t)\big)^{n_t} | q_1 = 1, \mathbf{b}_1 = \mathbf{h}_\lambda \Big\}.$$

The objective of Sections III–IV is to seek optimal control schemes for maximizing the TDR under the idealized and realistic environments, respectively.

III. MDP Framework for the Idealized Environment

In this section, we formulate the random access control problem in the idealized environment as an MDP, use the expected total reward of this MDP to evaluate the TDR, and obtain an optimal control scheme that maximizes the TDR.

A. MDP Formulation

From the dynamic control scheme specified in Section II-B, we see that each node becomes inactive at the beginning of slot $t+1$ with the transmission probability p_t if it is active at the beginning of slot t, and will be always inactive otherwise. This implies that the probability of moving to the next state in the state process $(q_t, n_t)_{t \in \mathcal{T}}$ depends only on the current state. Thus, $(q_t, n_t)_{t \in \mathcal{T}}$ can be viewed as a discrete-time finite-horizon, finite-state Markov chain.

Based on the Markov chain $(q_t, n_t)_{t \in \mathcal{T}}$, we present an MDP formulation by introducing the following definitions.

Actions: At the beginning of each slot $t \in \mathcal{T}$ with $q_t = 1$, the action performed by the tagged node (and the other active nodes) is the chosen transmission probability p_t taking values in the action space $[0,1]$. Note that the tagged node performs no action when $q_t = 0$.

State Transition Function: As the tagged node will never transmit since slot t if $q_t = 0$, we only consider the state transition function when $q_t = 1$. The state transition function $\beta_t\big((q', n'), (1, n), p\big)$ is defined as the transition probability of moving from the state $(q_t, n_t) = (1, n)$ to $(q_{t+1}, n_{t+1}) = (q', n')$ when each active node at the beginning of slot t adopts the transmission probability $p_t = p$. So, we have

$$\begin{aligned} &\beta_t\big((q', n'), (1, n), p\big) \\ &\triangleq \Pr\big((q_{t+1}, n_{t+1}) = (q', n') | (q_t, n_t) = (1, n), p_t = p\big) \\ &= \begin{cases} \binom{n}{n-n'} p^{n-n'+1-q'} (1-p)^{n'+q'}, & \text{if } n' \le n, \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \tag{1}$$

for each $t \in \mathcal{T} \setminus \{D\}$, each $q' \in \{0,1\}$, each $n, n' \in \mathcal{N}$ and each $p \in [0,1]$.

Rewards: The reward gained at slot t is defined as the average number of packets of the tagged node transmitted successfully to an arbitrary other node at slot t. As there is no reward at slot t when $q_t = 0$, we only focus on the cases when $q_t = 1$. Let $r_t\big((1, n), p\big)$ denote the reward at slot t for the state $(q_t, n_t) = (1, n)$ when each active node at the beginning of slot t adopts $p_t = p$. So, we have

$$r_t\big((1, n), p\big) = \sigma p (1-p)^n, \tag{2}$$

for each $t \in \mathcal{T}$, each $n \in \mathcal{N}$ and each $p \in [0,1]$.

Policy: A dynamic control scheme $\widehat{\boldsymbol{\pi}}$ defined in Section II-B can be seen as a deterministic Markovian policy.

Let $R^{\widehat{\boldsymbol{\pi}}}(1, n)$ denote the expected total reward from slot 1 to slot D when $q_1 = 1$, $n_1 = n$ and the policy $\widehat{\boldsymbol{\pi}}$ is used, which can be defined by

$$R^{\widehat{\boldsymbol{\pi}}}(1, n) \triangleq \mathbb{E}^{\widehat{\boldsymbol{\pi}}} \Big\{ \sum_{t=1, q_t = 1}^{D} r_t\big((1, n_t), \widehat{\pi}_t(n_t)\big) | q_1 = 1, n_1 = n \Big\}.$$

Obviously, $\text{TDR}^{\widehat{\boldsymbol{\pi}}} = \sum_{n \in \mathcal{N}} \binom{N-1}{n} \lambda^n (1-\lambda)^{N-1-n} R^{\widehat{\boldsymbol{\pi}}}(1, n)$.

B. MDP solution

Due to the finite horizon, finite state space, compact action space, bounded rewards, continuous rewards with respect to p and continuous state transition function with respect to p in our MDP formulation, [31, Prop. 4.4.3, Ch. 4] indicates that for maximizing $\text{TDR}^{\widehat{\boldsymbol{\pi}}}$, there exists a $\widehat{\boldsymbol{\pi}} \in \widehat{\boldsymbol{\Pi}}$, which is indeed optimal over all random and deterministic, history-dependent and Markovian policies. This property also justifies the design goal for the idealized environment considered in Section II. Hence, we aim to seek

$$\widehat{\boldsymbol{\pi}}^* \in \arg\max_{\widehat{\boldsymbol{\pi}} \in \widehat{\boldsymbol{\Pi}}} R^{\widehat{\boldsymbol{\pi}}}(1,n), \quad \forall n \in \mathcal{N}.$$

Let $U_t^*(1,n)$ denote the value function corresponding to the maximum total expected reward from slot t to slot D when $q_t = 1$ and $n_t = n$. Averaging over all possible next states with $q_{t+1} = 1$, we arrive at the following Bellman's equation:

$$\begin{aligned} U_D^*(1,n) &= \max_{p\in[0,1]} r_D\big((1,n),p\big), \quad \forall n \in \mathcal{N}, \\ U_t^*(1,n) &= \max_{p\in[0,1]} r_t\big((1,n),p\big) \\ &\quad + \sum_{n'\in\mathcal{N}} \beta_t\big((1,n'),(1,n),p\big) U_{t+1}^*(1,n'), \quad \forall n \in \mathcal{N}, \end{aligned} \tag{3}$$

for each $t \in \mathcal{T} \setminus \{D\}$.

Applying the backward induction algorithm to get a solution to Eq. (3) involves finding global maximizers of ND real-coefficient univariate polynomials defined on $[0,1]$, and can formally lead to $\widehat{\boldsymbol{\pi}}^*$. This indicates that even if the contention intensity is completely known, the computation required to obtain an optimal scheme is still demanding in practice.

IV. POMDP Framework for the Realistic Environment

In this section, we formulate the random access control problem in the realistic environment as a POMDP, use the expected total reward of this POMDP to evaluate the TDR, and discuss how to obtain an optimal or near-optimal scheme.

A. POMDP Formulation

Based on the Markov chain $(q_t, n_t)_{t\in\mathcal{T}}$ specified in Section III and the activity belief $\mathbf{b}_t$ for each $t \in \mathcal{T}$ with $q_t = 1$ specified in Section II-B, we present a POMDP formulation by introducing the following definitions.

Actions, State Transition Function, Rewards: The definitions of these elements are the same as in Section III.

Observations and Observation Function: The tagged node at the beginning of slot $t+1$ can obtain an observation on the channel status of slot t, denoted by o_t. When $q_{t+1} = 0$, the tagged node will never transmit since slot $t+1$ and o_t will thus be useless. Hence, we only consider o_t when $q_{t+1} = 1$ taking values from the observation space $\mathcal{O} \triangleq \{0\,(\text{idle}), 1\,(\text{busy})\}$. Further, the observation function $\omega_t\big(o,(1,n),(1,n')\big)$ is defined as the probability that the tagged node at the beginning of slot $t+1$ obtains the observation $o_t = o$ if $(q_t, n_t) = (1,n)$ and $(q_{t+1}, n_{t+1}) = (1,n')$. So, we have

$$\begin{aligned} &\omega_t\big(o,(1,n),(1,n')\big) \\ &\triangleq \Pr\big(o_t = o | (q_t,n_t) = (1,n), (q_{t+1},n_{t+1}) = (1,n')\big) \\ &= \begin{cases} 1, & \text{if } o = 0, n = n', \\ 1, & \text{if } o = 1, n - n' \geq 1, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

for each $t \in \mathcal{T} \setminus \{D\}$, each $o \in \mathcal{O}$ and each $n, n' \in \mathcal{N}$.

Bayesian update of the Activity Belief: It has been shown in [24] that for each $t \in \mathcal{T}$ with $q_t = 1$, the value of the activity belief $\mathbf{b}_t$ is a sufficient statistic for the initial activity belief, all past channel statuses, and all past transmission probabilities. First, by the total number of nodes N and the packet generation probability λ, the tagged node can obtain

$$\mathbf{b}_1 = \mathbf{h}_\lambda \triangleq \big((1-\lambda)^{N-1}, (N-1)\lambda(1-\lambda)^{N-2}, \ldots, \lambda^{N-1}\big). \tag{4}$$

Then, for each $t \in \mathcal{T}\setminus\{D\}$, given the condition $q_t = q_{t+1} = 1$, the activity belief $\mathbf{b}_t = \mathbf{b}$, the observation $o_t = o$, and the transmission probability $p_t = p$ used at slot t, the tagged node at slot $t+1$ can obtain $\mathbf{b}_{t+1}$ via the Bayes' rule:

$$\mathbf{b}_{t+1} \triangleq \theta_t(\mathbf{b}, p, o, 1, 1),$$

$$\begin{aligned} &b_{t+1}(n') \\ &\triangleq \Pr\big(n_{t+1} = n' | \mathbf{b}_t = \mathbf{b}, p_t = p, o_t = o, q_t = q_{t+1} = 1\big) \\ &= \frac{\sum_{n\in\mathcal{N}} b(n)\omega_t\big(o,(1,n),(1,n')\big)\beta_t\big((1,n'),(1,n),p\big)}{\chi_t(o,\mathbf{b},p,1,1)}, \end{aligned}$$

for each $n' \in \mathcal{N}$, where

$$\begin{aligned} &\chi_t(o,\mathbf{b},p,1,1) \triangleq \Pr\big(q_{t+1} = 1, o_t = o | q_t = 1, \mathbf{b}_t = \mathbf{b}, p_t = p\big) \\ &= \sum_{n\in\mathcal{N}} b(n) \sum_{n''\in\mathcal{N}} \omega_t\big(o,(1,n),(1,n'')\big)\beta_t\big((1,n''),(1,n),p\big). \end{aligned}$$

Policy: A dynamic control scheme $\boldsymbol{\pi}$ defined in Section II-B can be seen as a deterministic Markovian policy.

Let $R^{\boldsymbol{\pi}}(1,\mathbf{h}_\lambda)$ denote the expected total reward from slot 1 to slot D when $q_1 = 1$, $\mathbf{b}_1 = \mathbf{h}_\lambda$ and the policy $\boldsymbol{\pi}$ is used, which can be defined by

$$\begin{aligned} &R^{\boldsymbol{\pi}}(1,\mathbf{h}_\lambda) \\ &\triangleq \mathbb{E}^{\boldsymbol{\pi}}\Big\{ \sum_{t=1,q_t=1}^{D} r_t\big((1,n_t),\pi_t(\mathbf{b}_t)\big) | q_1 = 1, \mathbf{b}_1 = \mathbf{h}_\lambda \Big\}. \end{aligned}$$

Obviously, $\text{TDR}^{\boldsymbol{\pi}} = R^{\boldsymbol{\pi}}(1,\mathbf{h}_\lambda)$.

B. POMDP solution

Due to the finite horizon, finite state space, compact action space, bounded rewards, continuous rewards with respect to p and continuous $\chi_t(o,\mathbf{b},p,1,1)$ with respect to p in our POMDP formulation, [31, Prop. 4.4.3, Ch. 4] and [32, Thm. 7.1, Ch. 6] indicate that for maximizing $\text{TDR}^{\boldsymbol{\pi}}$, there exists a $\boldsymbol{\pi} \in \boldsymbol{\Pi}$, which is indeed optimal over all types of policies. This property justifies the design goal for the realistic environment considered in Section II. Hence, we aim to seek

$$\boldsymbol{\pi}^* \in \arg\max_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} R^{\boldsymbol{\pi}}(1,\mathbf{h}_\lambda).$$

Let $V_t^*(1,\mathbf{b})$ denote the value function corresponding to the maximum total expected reward from slot t to slot D when $q_t=1$ and $\mathbf{b}_t=\mathbf{b}$. Averaging over all possible current states with $q_t=1$ and observations with $q_{t+1}=1$, we arrive at the following Bellman's equation:

$$\begin{aligned}
V_D^*(1,\mathbf{b}) &= \max_{p\in[0,1]} \sum_{n\in\mathcal{N}} b(n) r_D\big((1,n),p\big), \quad \forall \mathbf{b}\in\mathcal{B}_D,\\
V_t^*(1,\mathbf{b}) &= \max_{p\in[0,1]} \sum_{n\in\mathcal{N}} b(n) r_t\big((1,n),p\big)\\
&\quad + \sum_{o\in\mathcal{O}} \chi_t(o,\mathbf{b},p,1,1) V_{t+1}^*\big(1,\theta_t(\mathbf{b},p,o,1,1)\big), \quad \forall \mathbf{b}\in\mathcal{B}_t,
\end{aligned} \tag{5}$$

for each $t\in\mathcal{T}\setminus\{D\}$. Solving Eq. (5) formally leads to $\boldsymbol{\pi}^*$.

Unfortunately, getting $\boldsymbol{\pi}^*$ by solving Eq. (5) is computationally intractable, as both the belief state space $\bigcup_{t\in\mathcal{T}}\mathcal{B}_t$ and the action space $[0,1]$ are infinite in our POMDP formulation. As such, an alternative is to consider a discretized action space $\mathcal{A}_d$ that only consists of uniformly distributed samples of the interval $[0,1]$, i.e., $\mathcal{A}_d \triangleq \{0,\Delta p, 2\Delta p,\ldots,1\}$, where Δp denotes the sampling interval. Hence, it is easy to see that $\mathcal{B}_t$ will become finite for each $t\in\mathcal{T}$ due to the finite $\mathcal{A}_d$. Then, theoretically, applying the backward induction algorithm [24] to get a solution to Eq. (5) can lead to a near-optimal policy, whose loss of optimality increases with Δp. However, this approach is still computationally prohibitive due to super-exponential growth in the value function complexity.

V. A Heuristic Scheme for the Realistic Environment

To overcome the infeasibility in obtaining an optimal or near-optimal control scheme for the realistic environment from the POMDP framework, in this section, we propose a simple heuristic control scheme that utilizes the key properties of our problem. It will be shown in Section VII that the heuristic scheme performs quite well in simulations.

A. Heuristics from the idealized environment

We first investigate the behaviors of $\widehat{\boldsymbol{\pi}}^*$ for two extreme cases in the idealized environment, which would serve to provide important clues on approximating $\widehat{\boldsymbol{\pi}}^*$. Let $U_t\big((1,n),p\big)$ denote the total expected reward from slot t to slot D for the state $(q_t,n_t)=(1,n)$ when each active node at the beginning of slot t adopts the transmission probability $p_t=p$ and the optimal decision rules at slots $t+1,t+2,\ldots,D$. So, we have $U_t\big((1,n),\widehat{\pi}_t^*(n)\big)=U_t^*(1,n)$.

Lemma 1. *When $n_1=m\to\infty$, by assuming that each collision involves a finite number of packets, for each $t\in\mathcal{T}$ and each possible $n_t=n$, we have*

$$\lim_{m\to\infty}(n+1)U_t\big((1,n),\frac{1}{n+1}\big)=\frac{(D-t+1)\sigma}{e}, \tag{6}$$

$$\lim_{m\to\infty}(n+1)U_t^*(1,n)=\frac{(D-t+1)\sigma}{e}. \tag{7}$$

Proof. Assume each collision involves at most a finite number, $k\geq 2$, of packets. We begin with the case $t=D$. By Eq. (2), we know $U_D\big((1,n),p\big)=\sigma p(1-p)^n$ and thus $\widehat{\pi}_D^*(n)=\frac{1}{n+1}$. As k and D are both finite, for each $n_D=n\in\{m-k(D-1),m-k(D-1)+1,\ldots,m\}$, we obtain that $m\to\infty$ implies $n\to\infty$ and then

$$\lim_{m\to\infty}(n+1)U_D^*(1,n)=\lim_{m\to\infty}(n+1)U_D\big((1,n),\frac{1}{n+1}\big)=\frac{\sigma}{e}. \tag{8}$$

Consider the case $t=D-1$. By the finite-horizon policy evaluation algorithm [31] and Eqs. (1), (2), for each $n_{D-1}=n\in\{m-k(D-2),m-k(D-2)+1,\ldots,m\}$, we have

$$\begin{aligned}
&(n+1)U_{D-1}\big((1,n),p\big)=(n+1)r_{D-1}\big((1,n),p\big)\\
&\quad+(n+1)\sum_{n'\in\mathcal{N}}\beta_{D-1}\big((1,n'),(1,n),p\big)U_D^*(1,n')\\
&=\sigma(n+1)p(1-p)^n\\
&\quad+\sum_{n'\in\mathcal{N}}(n+1)\frac{n!}{n'!(n-n')!}p^{n-n'}(1-p)^{n'+1}U_D^*(1,n')\\
&=\sigma(n+1)p(1-p)^n\\
&\quad+\sum_{n'\in\mathcal{N}}\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}(n'+1)U_D^*(1,n').
\end{aligned}$$

By assuming each collision involves at most a finite number, $k\geq 2$, of packets, we have

$$\begin{aligned}
&(n+1)U_{D-1}\big((1,n),p\big)=\sigma(n+1)p(1-p)^n\\
&+\sum_{n'=n-k+1}^{n}\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}(n'+1)U_D^*(1,n')\\
&+\Big(1-\sum_{n'=n-k+1}^{n}\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}\Big)\\
&\quad\cdot(n-k+1)U_D^*(1,n-k) \qquad (9)\\
&\leq\sigma(1-\frac{1}{n+1})^n+(n-k+1)U_D^*(1,n-k)\\
&+\sum_{n'=n-k+1}^{n}\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}\\
&\quad\cdot\big((n'+1)U_D^*(1,n')-(n-k+1)U_D^*(1,n-k)\big). \qquad (10)
\end{aligned}$$

For each $n'\in\{n-k+1,n-k+2,\ldots,n\}$, since $0\leq\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}\leq 1$, by applying the squeeze theorem, we obtain from Eq. (8) that

$$\begin{aligned}
&\lim_{m\to\infty}\binom{n+1}{n-n'}p^{n-n'}(1-p)^{n'+1}\\
&\quad\cdot\big((n'+1)U_D^*(1,n')-(n-k+1)U_D^*(1,n-k)\big)=0.
\end{aligned} \tag{11}$$

By Eqs. (8), (11) and inequality (10), as k and D are both finite, we further obtain that $m\to\infty$ implies $n\to\infty$ and

$$\begin{aligned}
&\limsup_{m\to\infty}(n+1)U_{D-1}\big((1,n),p\big)\\
&\leq\limsup_{m\to\infty}\big(\sigma(1-\frac{1}{n+1})^n+(n-k+1)U_D^*(1,n-k)\big)\\
&=\lim_{m\to\infty}\big(\sigma(1-\frac{1}{n+1})^n+(n-k+1)U_D^*(1,n-k)\big)=\frac{2\sigma}{e},
\end{aligned}$$

which implies

$$\limsup_{m\to\infty}(n+1)U_{D-1}^*(1,n)\leq\frac{2\sigma}{e}. \tag{12}$$

By setting $\widehat{\pi}_{D-1}(n) = \frac{1}{n+1}$ for each $n_{D-1} = n \in \{m - k(D-2), m-k(D-2)+1, \ldots, m\}$, as k and D are both finite, we obtain that $m \to \infty$ implies $n \to \infty$, and then obtain from Eqs. (8), (9) and (11) that

$$\lim_{m\to\infty} (n+1) U_{D-1}\big((1,n), \frac{1}{n+1}\big) = \lim_{m\to\infty} \big(\sigma(1-\frac{1}{n+1})^n + (n-k+1)U_D^*(1,n-k)\big) = \frac{2\sigma}{e}.$$

Since $U_{D-1}^*(1,n) \geq U_{D-1}\big((1,n), \frac{1}{n+1}\big)$, we have

$$\begin{aligned} &\liminf_{m\to\infty} (n+1) U_{D-1}^*(1,n) \\ &\geq \liminf_{m\to\infty} (n+1) U_{D-1}\big((1,n), \frac{1}{n+1}\big) = \frac{2\sigma}{e}. \end{aligned} \tag{13}$$

Combining inequalities (12) and (13), we have $\lim_{m\to\infty}(n+1)U_{D-1}^*(1,n) = \frac{2\sigma}{e}$.

For the cases of $t = D-2, D-3, \ldots, 1$, iteratively repeating the above argument can lead to Eqs. (6) and (7) for each possible $n_t = n$. □

Lemma 1 motivates us to conjecture that, if n_1 takes a value sufficiently larger than D, the realizations of $(n_t+1)\widehat{\pi}_t^*(n_t)$ would always approach 1 for each $t \in \mathcal{T}$. Fig. 3 shows 1000 such realizations when $D = 10$ for $n_1 = 30, 50, 100$, respectively, which confirm our conjecture.

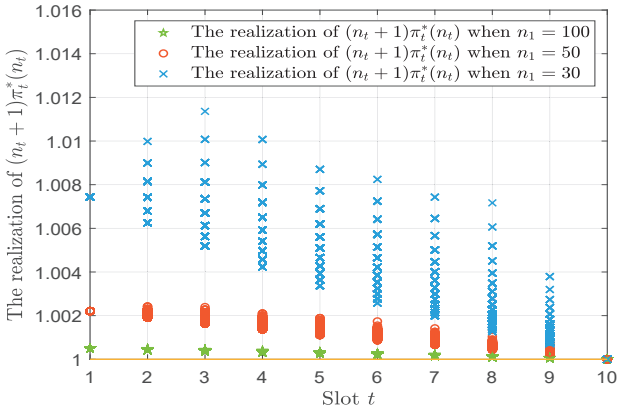

Fig. 3: Realizations of $(n_t+1)\widehat{\pi}_t^*(n_t)$ when $n_1 = 30, 50, 100$ and $D = 10$.

We further investigate the behaviors of $\widehat{\boldsymbol{\pi}}^*$ for the extreme case that n_1 takes a value sufficiently smaller than D.

Lemma 2. *For each $t \in \mathcal{T}$, we have*

$$U_t^*(1,1) = \frac{3D-3t+1}{3D-3t+4}\sigma, \tag{14}$$

and for each $t \in \mathcal{T} \setminus \{D\}$, we have

$$\widehat{\pi}_t^*(1) = \frac{3}{3D-3t+4}. \tag{15}$$

Proof. As $U_t^*(1,0) = \sigma$ for each $t \in \mathcal{T}$, we have

$$U_t\big((1,1),p\big) = 2\sigma p(1-p) + (1-p)^2 U_{t+1}^*(1,1), \tag{16}$$

for each $t \in \mathcal{T} \setminus \{D\}$. Taking the derivative of $U_t\big((1,1),p\big)$ with respect to p derives that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}p} U_t\big((1,1),p\big) \\ &= \big(2\sigma - 2U_{t+1}^*(1,1)\big) - \big(4\sigma - 2U_{t+1}^*(1,1)\big)p. \end{aligned}$$

As $\sigma > 0$ and $U_t^*(1,1) \leq \sigma$ for each $t \in \mathcal{T} \setminus \{D\}$, we have

$$\widehat{\pi}_t^*(1) = \frac{\sigma - U_{t+1}^*(1,1)}{2\sigma - U_{t+1}^*(1,1)}, \tag{17}$$

for each $t \in \mathcal{T} \setminus \{D\}$. In particular, as $U_D^*(1,1) = \sigma/4$, we obtain $\widehat{\pi}_{D-1}^*(1) = 3/7$, which satisfies Eq. (15).

Then, we aim to investigate the relation between $\widehat{\pi}_t^*(1)$ and $\widehat{\pi}_{D-1}^*(1)$ for each $t \in \mathcal{T} \setminus \{D-1, D\}$. By setting $p = \widehat{\pi}_t^*(1)$ in Eq. (16), we obtain

$$\begin{aligned} U_t^*(1,1) &= 2\sigma\widehat{\pi}_t^*(1)\big(1-\widehat{\pi}_t^*(1)\big) + \big(1-\widehat{\pi}_t^*(1)\big)^2 U_{t+1}^*(1,1) \\ &= \frac{\sigma^2}{2\sigma - U_{t+1}^*(1,1)}. \end{aligned} \tag{18}$$

Using Eq. (17) to express $U_{t+1}^*(1,1)$ and $U_t^*(1,1)$ in Eq. (18) in terms of $\widehat{\pi}_t^*(1)$ and $\widehat{\pi}_{t-1}^*(1)$, respectively, we have

$$\widehat{\pi}_t^*(1) = \frac{\widehat{\pi}_{t+1}^*(1)}{1+\widehat{\pi}_{t+1}^*(1)}, \tag{19}$$

for each $t \in \mathcal{T} \setminus \{D-1, D\}$. Furthermore, recursively using Eq. (19) yields

$$\widehat{\pi}_t^*(1) = \frac{\widehat{\pi}_{D-1}^*(1)}{1+(D-t-1)\widehat{\pi}_{D-1}^*(1)}, \tag{20}$$

and thus implies Eq. (15) by $\widehat{\pi}_{D-1}^*(1) = 3/7$.

Finally, combining Eqs. (15) and (17) yields

$$U_t^*(1,1) = \frac{1-2\widehat{\pi}_{t-1}^*(1)}{1-\widehat{\pi}_{t-1}^*(1)}\sigma = \frac{3D-3t+1}{3D-3t+4}\sigma, \tag{21}$$

for each $t \in \mathcal{T} \setminus \{1\}$, and substituting Eq. (21) into Eq. (18) yields $U_1^*(1,1) = \frac{3D-2}{3D+1}\sigma$. Hence we complete the proof for Eq. (14). □

Inspired by Eq. (15), we consider a simple control scheme $\widehat{\boldsymbol{\pi}}^{\text{eve}} \triangleq [\widehat{\pi}_1^{\text{eve}}, \widehat{\pi}_2^{\text{eve}}, \ldots, \widehat{\pi}_D^{\text{eve}}] \in \widehat{\boldsymbol{\Pi}}$ where

$$\widehat{\pi}_t^{\text{eve}}(n) = \frac{1}{D-t+1}, \tag{22}$$

for each $t \in \mathcal{T}$ and each $n \in \mathcal{N}$.

Let $U_t^{\text{eve}}(1,n)$ denote the expected total reward from slot t to slot D for the state $(q_t, n_t) = (1,n)$ when each active node adopts the decision rules $\widehat{\pi}_t^{\text{eve}}$ at slots $t, t+1, \ldots, D$. So, using the finite-horizon policy evaluation algorithm [31], we have

$$\begin{aligned} &U_D^{\text{eve}}(1,n) = r_D\big((1,n), \widehat{\pi}_D^{\text{eve}}(n)\big), \quad \forall n \in \mathcal{N}, \\ &U_t^{\text{eve}}(1,n) = r_t\big((1,n), \widehat{\pi}_t^{\text{eve}}(n)\big) \\ &\quad + \sum_{n' \in \mathcal{N}} \beta_t\big((1,n'),(1,n), \widehat{\pi}_t^{\text{eve}}(n)\big) U_{t+1}^{\text{eve}}(1,n'), \ \forall n \in \mathcal{N}. \end{aligned} \tag{23}$$

for each $t \in \mathcal{T} \setminus \{D\}$.

Lemma 3. *For each $t \in \mathcal{T} \setminus \{D\}$ and each $n \in \mathcal{N}$, we have*

$$U_t^{eve}(1,n) = \sigma\big(1 - \frac{1}{D-t+1}\big)^n. \tag{24}$$

Proof. We shall prove $U_t^{\text{eve}}(1,n)=\sigma\big(1-\frac{1}{D-t+1}\big)^n$ for each $n\in\mathcal{N}$ by induction from $t=D-1$ down to 1.

First, when $t=D-1$, by Eqs. (1), (2) and (23), we have

$$\begin{aligned}U_{D-1}^{\text{eve}}(1,n)&=r_{D-1}\big((1,n),\widehat{\pi}_{D-1}^{\text{eve}}(n)\big)\\&\quad+\sum_{n'\in\mathcal{N}}\beta_{D-1}\big((1,n'),(1,n),\widehat{\pi}_{D-1}^{\text{eve}}(n)\big)U_D^{\text{eve}}(1,n')\\&=\sigma\frac{1}{2}\big(1-\frac{1}{2}\big)^n+(1-\frac{1}{2})^{n+1}U_D^{\text{eve}}(1,0)=\sigma(1-\frac{1}{2})^n,\end{aligned}$$

for each $n\in\mathcal{N}$, thereby establishing the induction basis.

When $t\in\mathcal{T}\setminus\{D-1,D\}$, assume $U_{t+1}^{\text{eve}}(1,n)=\sigma\big(1-\frac{1}{D-t}\big)^n$ for each $n\in\mathcal{N}$. By Eqs. (1), (2) and (23), we have

$$\begin{aligned}U_t^{\text{eve}}(1,n)&=r_t\big((1,n),\widehat{\pi}_t^{\text{eve}}(n)\big)\\&\quad+\sum_{n'\in\mathcal{N}}\beta_t\big((1,n'),(1,n),\widehat{\pi}_t^{\text{eve}}(n)\big)U_{t+1}^{\text{eve}}(1,n')\\&=\sigma\frac{1}{D-t+1}\big(1-\frac{1}{D-t+1}\big)^n\\&\quad+\sum_{n'\in\mathcal{N}}\binom{n}{n-n'}\big(\frac{1}{D-t+1}\big)^{n-n'}\\&\qquad\cdot\big(1-\frac{1}{D-t+1}\big)^{n'+1}\sigma\big(1-\frac{1}{D-t}\big)^{n'}\\&=\sigma\big(1-\frac{1}{D-t+1}\big)^n\frac{1}{D-t+1}\\&\quad+\sigma\big(1-\frac{1}{D-t+1}\big)^n\frac{D-t}{D-t+1}\\&\qquad\cdot\sum_{n'\in\mathcal{N}}\binom{n}{n-n'}\big(\frac{1}{D-t}\big)^{n-n'}\big(1-\frac{1}{D-t}\big)^{n'}\\&=\sigma\big(1-\frac{1}{D-t+1}\big)^n,\end{aligned}$$

for each $n\in\mathcal{N}$. So, the inductive step is established.

Since both the base case and the inductive step have been proved as true, we have $U_t^{\text{eve}}(1,n)=\sigma\big(1-\frac{1}{D-t+1}\big)^n$ for each $t\in\mathcal{T}\setminus\{D\}$ and each $n\in\mathcal{N}$. □

For each $t\in\mathcal{T}\setminus\{D\}$ and each $n\in\mathcal{N}$, based on the fact $U_t^*(1,n)\le\sigma$, we have

$$\frac{U_t^{\text{eve}}(1,n)}{U_t^*(1,n)}\ge\big(1-\frac{1}{D-t+1}\big)^n. \tag{25}$$

We can observe from Eq. (25) that, if n is sufficiently smaller than $D-t+1$, the value of $U_t^*(1,n)$ is close to the value of $U_t^{\text{eve}}(1,n)$. Then, Eq. (25) motivates us to conjecture that, if n_t takes a value sufficiently smaller than $D-t+1$, $\widehat{\pi}_t^{\text{eve}}$ may behave like $\widehat{\pi}_t^*$, i.e., the realizations of $(D-t+1)\widehat{\pi}_t^*(n_t)$ would always approach $(D-t+1)\widehat{\pi}_t^{\text{eve}}(n_t)=1$ for each $n_t=n\in\mathcal{N}$. Fig. 4 shows 1000 such realizations when $n_1=10$ for $D=30,50,100$, respectively, which confirm our conjecture.

Naturally, we obtain the following heuristics from Lemma 1 and Eq. (25), respectively.

1) *From Lemma 1:* When the number of active nodes is sufficiently large compared with the value of remaining slots, it is desirable for the active nodes to adopt the transmission probability that maximizes the instantaneous throughput. This implies that, when the remaining slots are not enough, the active nodes should utilize the channel as much as possible, which is time-independent.

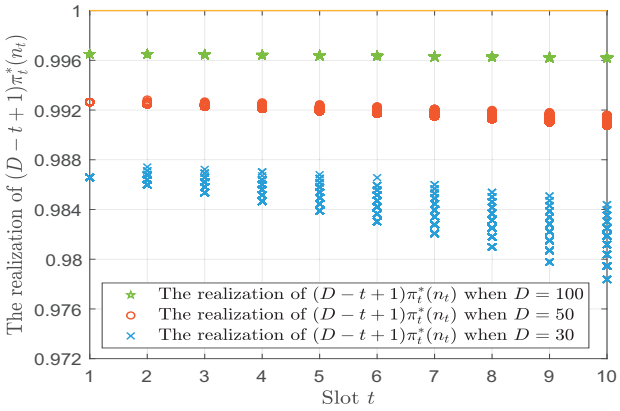

Fig. 4: Realizations of $(D-t+1)\widehat{\pi}_t^*(n_t)$ when $n_1=10$ and $D=30,50,100$.

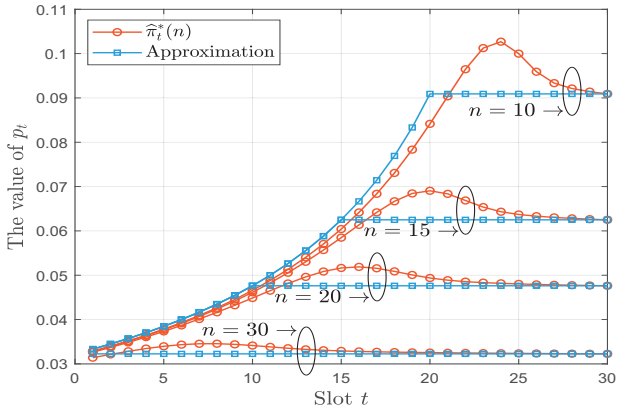

Fig. 5: $\widehat{\pi}_t^*(n)$ and its approximation for typical choices of parameters when $D=30$.

2) *From Eq. (25):* When the number of active nodes is sufficiently small compared with the value of remaining slots, it is desirable for the active nodes to adopt the transmission probability to ensure that all the backlogged packets would be almost evenly transmitted in the remaining slots. This implies that, when the remaining slots are enough, the active nodes should cherish the transmission chance in order to utilize the time as much as possible, which is time-dependent.

Based on these heuristics and the obvious fact $\widehat{\pi}_D^*(n)=\frac{1}{n+1}$ for each $n\in\mathcal{N}$, we propose a simple approximation on $\widehat{\boldsymbol{\pi}}^*$.

Approximation on $\widehat{\boldsymbol{\pi}}^*$: For each slot $t\in\mathcal{T}$ and each $n_t=n\in\mathcal{N}$, if the number of active nodes $n+1$ is larger than the value of remaining slots $D-t+1$ or $t=D$, $\widehat{\pi}_t^*(n)$ can be estimated by $\frac{1}{n+1}$, otherwise $\widehat{\pi}_t^*(n)$ can be estimated by $\frac{1}{D-t+1}$.

Fig. 5 compares $\widehat{\pi}_t^*(n)$ and its approximation for typical choices of parameters when $D=30$. It is shown that the approximation error is very small when the difference between n and $D-t$ is large. This is because $\widehat{\pi}_t^*(n)$ is dominated by n if n is much larger than $D-t$ but is dominated by t if n is much smaller than $D-t$, which is consistent with the heuristics. However, it is shown that the approximation error is noticeable

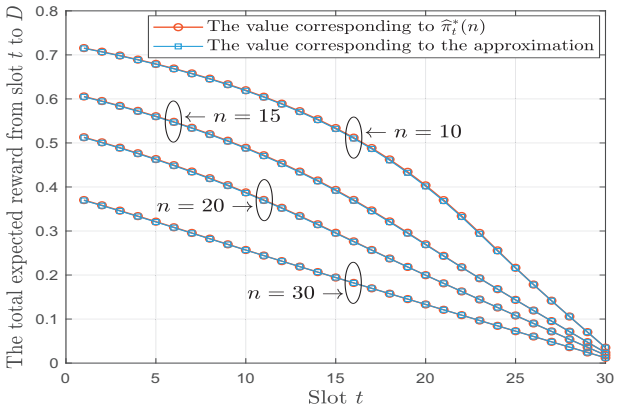

Fig. 6: The total expected rewards from slot t to D corresponding to $\widehat{\pi}_t^*(n)$ and its approximation for typical choices of parameters when $D = 30$.

when the difference between n and $D - t$ is small. This is because both n and t have notable impacts on $\widehat{\pi}_t^*(n)$ for this case, which is not considered in the heuristics. The results also show that the ratio of the cases with the approximation error larger than 8% is 6.67% and the largest approximation error is 11.47%, thus justifying our approximation. Furthermore, Fig. 6 compares the total expected rewards from slot t to D corresponding to $\widehat{\pi}_t^*(n)$ and its approximation for typical choices of parameters when $D = 30$. The results show that the approximation leads to at most 0.66% reward loss (much smaller than the approximation error), and thus verify that our approximation can be used to obtain TDR quite close to the maximum achievable TDR in the idealized environment.

B. A simple approximation on the activity belief of the realistic environment

To apply the approximation on $\widehat{\boldsymbol{\pi}}^*$ to the realistic environment, it is necessary for each active node to perform a runtime updating of the activity belief $\mathbf{b}_t$. However, as shown in Section IV-A, the full Bayesian updating of $\mathbf{b}_t$ is a bit computationally demanding to implement. So, we shall propose a simple approximation on $\mathbf{b}_t$, denoted by $\mathbf{b}_t^{\text{bd}} \triangleq \big(b_t^{\text{bd}}(0), b_t^{\text{bd}}(1), \ldots, b_t^{\text{bd}}(N-1)\big)$, relying on a binomial distribution with a changeable parameter vector (M_t, α_t). More specifically, if $(M_t, \alpha_t) = (M, \alpha)$, we have

$$b_t^{\text{bd}}(n) = \begin{cases} \binom{M}{n}\alpha^n(1-\alpha)^{M-n}, & \text{if } 0 \le n \le M, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

As such, in this manner, each active node will only keep the parameter vector (M_t, α_t) rather than the activity belief $\mathbf{b}_t$.

Obviously, by Eq. (4), we can set $(M_1, \alpha_1) = (N-1, \lambda)$ to achieve $\mathbf{b}_t = \mathbf{b}_t^{\text{bd}}$. Then, for each $t \in \mathcal{T} \setminus \{D\}$, we will show that we can use the Bayes' rule exactly to set the value of (M_{t+1}, α_{t+1}) when the observation $o_t = 0$, but must introduce an approximation assumption when $o_t = 1$.

For each $t \in \mathcal{T} \setminus \{D\}$, given $(M_t, \alpha_t) = (M, \alpha)$, $\mathbf{b}_t^{\text{bd}} = \mathbf{b}^{\text{bd}}$ and $p_t = p$, the following procedure first uses the Bayes' rule to compute $\mathbf{b}_{t+1}^{\text{med}} \triangleq \big(b_{t+1}^{\text{med}}(0), b_{t+1}^{\text{med}}(1), \ldots, b_{t+1}^{\text{med}}(N-1)\big)$, and then computes (M_{t+1}, α_{t+1}) based on the value of $\mathbf{b}_{t+1}^{\text{med}}$.

Case 1: if $o_t = 0$, the Bayesian update yields

$$\begin{aligned} &b_{t+1}^{\text{med}}(n') \\ &= \frac{\sum_{n\in\mathcal{N}} b^{\text{bd}}(n)\omega_t\big(0,(1,n),(1,n')\big)\beta_t\big((1,n'),(1,n),p\big)}{\chi_t(0,\mathbf{b}^{\text{bd}},p,1,1)} \\ &= \begin{cases} \binom{M}{n'}\left(\frac{\alpha-\alpha p}{1-\alpha p}\right)^{n'}\left(1-\frac{\alpha-\alpha p}{1-\alpha p}\right)^{M-n'}, & \text{if } 0 \le n' \le M, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

We require $\mathbf{b}_{t+1}^{\text{bd}}$ to directly take the value of $\mathbf{b}_{t+1}^{\text{med}}$, and set

$$(M_{t+1}, \alpha_{t+1}) = (M, \frac{\alpha - \alpha p}{1 - \alpha p}).$$

Case 2: if $o_t = 1$, the Bayesian update yields

$$\begin{aligned} &b_{t+1}^{\text{med}}(n') \\ &= \frac{\sum_{n\in\mathcal{N}} b^{\text{bd}}(n)\omega_t\big(1,(1,n),(1,n')\big)\beta_t\big((1,n'),(1,n),p\big)}{\chi_t(1,\mathbf{b}^{\text{bd}},p,1,1)} \\ &= \begin{cases} \frac{1}{1-(1-\alpha p)^M} \\ \cdot\Big[\binom{M}{n'}\big(\alpha(1-p)\big)^{n'}\big(1-\alpha(1-p)\big)^{M-n'} \\ \quad -(1-\alpha p)^M\binom{M}{n'}\left(\frac{\alpha-\alpha p}{1-\alpha p}\right)^{n'}\left(1-\frac{\alpha-\alpha p}{1-\alpha p}\right)^{M-n'}\Big], \\ \qquad\qquad \text{if } 0 \le n' \le M-1, \\ 0, \qquad\qquad \text{otherwise.} \end{cases} \end{aligned}$$

Such a Bayesian update does not yield a distribution in the form (26). However, we modify the value of $\mathbf{b}_{t+1}^{\text{med}}$ to a distribution in the form (26) by keeping the mean of the distribution unchanged and considering that the number of active nodes will be reduced by at least one due to a busy slot. So, when $M > 1$, we set

$$(M_{t+1}, \alpha_{t+1}) = \left(M-1, \frac{M(\alpha-\alpha p)\big(1-(1-\alpha p)^{M-1}\big)}{(M-1)\big(1-(1-\alpha p)^M\big)}\right),$$

and when $M = 1$, we adopt the convention that

$$(M_{t+1}, \alpha_{t+1}) = (M-1, 1).$$

The accuracy of this approximation will be examined via numerical results at the end of this section.

C. A heuristic scheme

With the investigations in Sections V-A and V-B together, we are ready to propose a heuristic but very simple control scheme for the realistic environment, $\boldsymbol{\pi}^{\text{heu}}$.

At the beginning of each slot $t \in \mathcal{T}$, given the parameter vector of belief approximation $(M_t, \alpha_t) = (M, \alpha)$, each active node uses $M\alpha$ to estimate the mean of n_t, and further uses the following simple rule $\pi_t^{\text{heu}}(M, \alpha)$ to determine the value of transmission probability p_t.

1) If $M\alpha + 1 > D - t + 1$ or $t = D$, set $\pi_t^{\text{heu}}(M, \alpha)$ to maximize the expected instantaneous throughput, i.e.,

$$\begin{aligned} \pi_t^{\text{heu}}(M,\alpha) &\in \arg\max_{p\in[0,1]} \sum_{n\in\mathcal{N}} b^{\text{bd}}(n) r_t\big((1,n),p\big) \\ &= \min\left(\frac{1}{M\alpha+\alpha}, 1\right). \end{aligned} \tag{27}$$

TABLE III: A comparison between the realizations of $\mathbf{b}_t$ and its approximation $\mathbf{b}_t^{\text{bd}}$ when each active node adopts $\boldsymbol{\pi}^{\text{heu}}$ for $N = 10$, $\lambda = 0.8$, $D = 8$.

		$b_t(0)$	$b_t(1)$	$b_t(2)$	$b_t(3)$	$b_t(4)$	$b_t(5)$	$b_t(6)$	$b_t(7)$	$b_t(8)$	$b_t(9)$
$t=1$	$\mathbf{b}_t$	0.000001	0.000018	0.000295	0.002753	0.016515	0.066060	0.176161	0.301990	0.301990	0.134218
$o_1=0$	Approx.	0.000001	0.000018	0.000295	0.002753	0.016515	0.066060	0.176161	0.301990	0.301990	0.134218
$t=2$	$\mathbf{b}_t$	0.000001	0.000042	0.000583	0.004760	0.024988	0.087458	0.204068	0.306102	0.267839	0.104160
$o_2=1$	Approx.	0.000001	0.000042	0.000583	0.004760	0.024988	0.087458	0.204068	0.306102	0.267839	0.104160
$t=3$	$\mathbf{b}_t$	0.000059	0.001098	0.009014	0.042646	0.127254	0.245406	0.298859	0.210235	0.065430	0
$o_3=1$	Approx.	0.000052	0.001004	0.008559	0.041692	0.126924	0.247294	0.301138	0.209545	0.063792	0
$t=4$	$\mathbf{b}_t$	0.001086	0.012248	0.059916	0.164987	0.276437	0.282086	0.162465	0.040774	0	0
$o_4=1$	Approx.	0.000974	0.011537	0.058598	0.165343	0.279925	0.284347	0.160466	0.038810	0	0
$t=5$	$\mathbf{b}_t$	0.010921	0.072058	0.201100	0.304173	0.263268	0.123764	0.024716	0	0	0
$o_5=1$	Approx.	0.010329	0.070827	0.202359	0.308353	0.264299	0.120821	0.023013	0	0	0
$t=6$	$\mathbf{b}_t$	0.068102	0.238724	0.340491	0.247285	0.091556	0.013842	0	0	0	0
$o_6=0$	Approx.	0.067210	0.240606	0.344541	0.246686	0.088312	0.012646	0	0	0	0
$t=7$	$\mathbf{b}_t$	0.169904	0.357679	0.306377	0.133629	0.029713	0.002698	0	0	0	0
$o_7=0$	Approx.	0.167239	0.359554	0.309208	0.132956	0.028585	0.002458	0	0	0	0
$t=8$	$\mathbf{b}_t$	0.421334	0.395352	0.150943	0.029344	0.002908	0.000118	0	0	0	0
$o_8=1$	Approx.	0.416144	0.398784	0.152859	0.029297	0.002807	0.000108	0	0	0	0

The proof of Eq. (27) is provided in the supplemental material.

2) Otherwise, set

$$\pi_t^{\text{heu}}(M,\alpha) = \frac{1}{D-t+1}.$$

The heuristic scheme depends on not only n_t but also t, so it is a time-dependent dynamic policy.

Table III compares the realizations of $\mathbf{b}_t$ and its approximation $\mathbf{b}_t^{\text{bd}}$ when each active node adopts $\boldsymbol{\pi}^{\text{heu}}$ for $N = 10$, $\lambda = 0.8$, $D = 8$, and verifies that the proposed approximation is reasonable.

Let $R^{\boldsymbol{\pi}^{\text{heu}}}\big(1,(N-1,\lambda)\big)$ denote the expected total reward from slot 1 to slot D when $q_1 = 1$, $(M_1,\alpha_1) = (N-1,\lambda)$ and the policy $\boldsymbol{\pi}^{\text{heu}}$ is used, which can be defined by

$$\begin{aligned} &R^{\boldsymbol{\pi}^{\text{heu}}}\big(1,(N-1,\lambda)\big) \\ &\triangleq \mathbb{E}^{\boldsymbol{\pi}^{\text{heu}}}\Big\{ \sum_{t=1,q_t=1}^{D} r_t\big((1,n_t),\pi_t^{\text{heu}}(M_t,\alpha_t)\big) \\ &\qquad\qquad |q_1=1,(M_1,\alpha_1)=(N-1,\lambda)\Big\}. \end{aligned}$$

Obviously, we have $\text{TDR}^{\boldsymbol{\pi}^{\text{heu}}} = R^{\boldsymbol{\pi}^{\text{heu}}}\big(1,(N-1,\lambda)\big)$.

Remark: The proposed heuristic scheme for the realistic environment is based on the model-based MDP and POMDP formulations, which require each node to know other nodes' traffic parameters. Owing to the known model parameters, the heuristic scheme enjoys threefold advantages of being implemented without imposing extra overhead and hardware cost, of being implemented with very low computational complexity, and of achieving TDR close to the maximum achievable TDR in the idealized environment. However, in practice, the model parameters may not be available to formulate the optimization and obtain the heuristic scheme. As such, reinforcement learning algorithms [33] can be used to learn the model parameters by directly interacting with the environment, but at the cost of extra training phase with slow convergence speed and high computational complexity, which may not be feasible for low-cost nodes.

VI. Extension to Retransmission-based Broadcasting

In our study so far, we have assumed that every packet is never retransmitted, which is a common setting in broadcasting due to the lack of acknowledgements, and a desirable setting for energy saving in many energy-constrained applications. This assumption also allows us to provide a nice presentation in Sections III–V for designing optimal and heuristic schemes. In this section, to further improve the TDR, we relax this assumption to allow at most $K \geq 1$ copies of every packet to be transmitted within the deadline D slots. However, the updating of the activity belief for the case $K \geq 2$ would lead to much more complicated modeling than that for the case $K = 1$. Instead, we directly generalize $\boldsymbol{\pi}^{\text{heu}}$ to propose a heuristic scheme for the realistic environment with $K \geq 1$, $\boldsymbol{\pi}^{\text{heuR}}$, which makes good use of the results in Sections III–V.

The key idea behind $\boldsymbol{\pi}^{\text{heuR}}$ is based on the heuristic that each copy of every packet has "average deadline" D/K slots. So, we first divide a frame into K consecutive subframes, indexed by $k \in \{1,2,\ldots,K\}$, and require the k-th subframe to occupy D_k slots by the following rule: if $r = 0$, $D_k = D/K$ for $k = 1,2,\ldots,K$; otherwise, $D_k = \lfloor D/K \rfloor$ for $k = 1,2,\ldots,K-r$, and $D_k = \lceil D/K \rceil$ for $k = K+1-r,\ldots,K$, where $r = D \mod K$. Then, we require all the nodes, which are active at the beginning of a frame, to only broadcast the k-th copy at most once during the k-th subframe, for $k = 1,2,\ldots,K$. In this manner, it is easy to see that the updating of the activity belief proposed in Sections IV–V is still applicable to each subframe. This property motivates us to use $\boldsymbol{\pi}^{\text{heu}}$ with the same initial belief to broadcast the k-th copy during the k-th subframe for $k = 1,2,\ldots,K$, i.e., repeatedly use $\boldsymbol{\pi}^{\text{heu}}$ for K times. Clearly, the study in Section V can be seen as a particular case here (i.e., $K = 1$, $D_1 = D$). An example of the working procedure of $\boldsymbol{\pi}^{\text{heuR}}$ for $D = 6, K = 2$ can be found in Fig. 7.

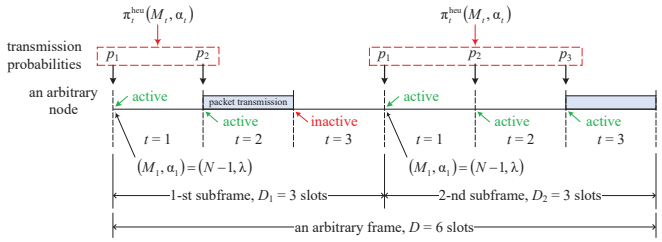

Fig. 7: An example of the working procedure of $\boldsymbol{\pi}^{\text{heuR}}$ for $D = 6, K = 2$.

Let $R_k^{\boldsymbol{\pi}^{\text{heu}}}\big(1, n, (N-1, \lambda)\big)$ denote the expected total reward of the k-th subframe from slot 1 to slot D_k when $q_1 = 1$, $n_1 = n$, $(M_1, \alpha_1) = (N-1, \lambda)$ and the policy $\boldsymbol{\pi}^{\text{heu}}$ is used, which can be defined by

$$\begin{aligned} &R_k^{\boldsymbol{\pi}^{\text{heu}}}\big(1, n, (N-1, \lambda)\big) \\ &\triangleq \mathbb{E}^{\boldsymbol{\pi}^{\text{heu}}}\Bigg\{ \sum_{t=1, q_t=1}^{D_k} r_t\big((1, n_t), \pi_t^{\text{heu}}(M_t, \alpha_t)\big) \\ &\qquad\qquad |q_1 = 1, n_1 = n, (M_1, \alpha_1) = (N-1, \lambda)\Bigg\}. \end{aligned}$$

Then, the TDR under the policy $\boldsymbol{\pi}^{\text{heuR}}$ can be computed by

$$\begin{aligned} \text{TDR}^{\boldsymbol{\pi}^{\text{heuR}}} = 1 - \sum_{n \in \mathcal{N}} \binom{N-1}{n} \lambda^n (1-\lambda)^{N-1-n} \\ \cdot \prod_{k=1}^{K} \Big(1 - R_k^{\boldsymbol{\pi}^{\text{heu}}}\big(1, n, (N-1, \lambda)\big)\Big). \end{aligned}$$

The performance improvement owning to the use of $\boldsymbol{\pi}^{\text{heuR}}$ will be investigated in Section VII.

VII. Numerical Evaluation

This section includes two subsections. To validate the studies in Sections III–V without considering retransmissions, the first subsection compares the TDR performance of the optimal scheme for the idealized environment $\widehat{\boldsymbol{\pi}}^*$, the proposed heuristic scheme for the realistic environment $\boldsymbol{\pi}^{\text{heu}}$, an optimal static scheme for the realistic environment $\boldsymbol{\pi}^{\text{sta}}$ [6], [8], [11], [13], and the ITM scheme for the realistic environment (also can be seen as the myopic scheme in our POMDP model) $\boldsymbol{\pi}^{\text{myo}}$ [14], [20], [21]. Here, $\boldsymbol{\pi}^{\text{sta}}$ requires each active node to always adopt a static and identical transmission probability, and an optimal transmission probability can be obtained by solving $\max_{p \in [0,1]} R^{\boldsymbol{\pi}}(1, \mathbf{h}_\lambda)$ s.t. $\pi_t(\mathbf{b}) = p$, $\forall t \in \mathcal{T}$, $\forall \mathbf{b} \in \mathcal{B}_t$. Such comparisons are not only helpful to demonstrate the performance loss due to the incomplete knowledge of the value of n_t, but also helpful to demonstrate the performance advantage benefitting from dynamically adjusting the transmission probability. In the second subsection, for the scenario where each packet is allowed to be transmitted at most K times, we compare the TDR performance of the proposed heuristic scheme for the realistic environment $\boldsymbol{\pi}^{\text{heuR}}$ with $K = K^{\text{heuR}*}$, and an optimal static scheme for the realistic environment $\boldsymbol{\pi}^{\text{staR}}$ with $K = K^{\text{staR}*}$. Here, $K^{\text{heuR}*}$ and $K^{\text{staR}*}$ are optimal values of K that maximize the TDR of $\boldsymbol{\pi}^{\text{heuR}}$ and $\boldsymbol{\pi}^{\text{staR}}$, respectively.

It is worth mentioning that, to our best knowledge, designing the ITM scheme with $1 < K < D$ for the realistic environment requires a new modeling to update the activity belief due to the impact of the retransmission limit on the node status, which is very different from the modelings in this paper and previous studies [19]–[23]. Meanwhile, the ITM scheme with $K = D$ for the realistic environment can be seen as a particular static scheme. So, based on these two considerations, the ITM scheme with retransmissions is not presented in Subsection VII-B for comparisons.

The scenarios considered in the numerical experiments for the first subsection are in accordance with the system model specified in Section II-A, while the scenarios for the second subsection additionally allows retransmissions. We shall vary the system configuration over a wide range to investigate the impact of control scheme design on the TDR performance. Each numerical result is obtained from 10^7 independent numerical experiments.

A. Comparisons Without Retransmissions

Fig. 8 shows the TDR performance without retransmissions as a function of the packet arrival rate λ. The TDR of each scheme except $\boldsymbol{\pi}^{\text{myo}}$ always decreases when λ increases, due to the increase of contention intensity. But λ has a more complicated impact on $\boldsymbol{\pi}^{\text{myo}}$, because $\boldsymbol{\pi}^{\text{myo}}$ behaves more optimally when λ increases as indicated by Lemma 1, which compensates the negative impact of the increase of contention intensity. We observe that $\boldsymbol{\pi}^{\text{heu}}$ performs close to $\widehat{\boldsymbol{\pi}}^*$: 2.87%–8.18% loss when $D = 10$ and 0.56%–4.04% loss when $D = 20$. This indicates that the design of $\boldsymbol{\pi}^{\text{heu}}$ is reasonable, and the incomplete knowledge of the value of n_t has a minor impact on the TDR performance. We further observe that $\boldsymbol{\pi}^{\text{heu}}$ significantly outperforms $\boldsymbol{\pi}^{\text{sta}}$: 2.03%–17.06% improvement when $D = 10$ and 11.09%–19.59% improvement when $D = 20$. The reason is obvious that $\boldsymbol{\pi}^{\text{sta}}$ does not adjust the transmission probability according to the current delivery urgency and contention intensity. It is interesting to note that $\boldsymbol{\pi}^{\text{sta}}$ performs closer to other schemes as λ increases. This is because the optimal transmission probabilities for different values of t and n_t become closer with the value of n_1, as indicated by Lemma 1. Moreover, we also observe that $\boldsymbol{\pi}^{\text{heu}}$ significantly outperforms $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda < D$: 14.43%–58.24% improvement when $D = 10$ and 12.37%–106.57% improvement when $D = 20$; $\boldsymbol{\pi}^{\text{heu}}$ achieves almost the same TDR as $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda \geq D$. This observation confirms again the heuristic from Lemma 1 in Section V-A and further indicates that the ITM goal [14], [20], [21] is unsuitable here.

Fig. 9 shows the TDR performance without retransmissions as a function of the delivery deadline D. We observe that the TDR of each scheme except $\boldsymbol{\pi}^{\text{myo}}$ always increases with D, due to the decrease of delivery urgency. But the TDR of $\boldsymbol{\pi}^{\text{myo}}$ first increases and then remains the same with D. This phenomenon is due to the fact that, when $N\lambda < D$, $\boldsymbol{\pi}^{\text{myo}}$ would transmit all backlogged packets in the first $\lceil N\lambda \rceil$ slots at average and waste the remaining slots. We observe that $\boldsymbol{\pi}^{\text{heu}}$ performs significantly better than $\boldsymbol{\pi}^{\text{sta}}$: 6.61%–16.74% improvement when $\sigma = 0.8$ and 6.56%–16.89% improvement

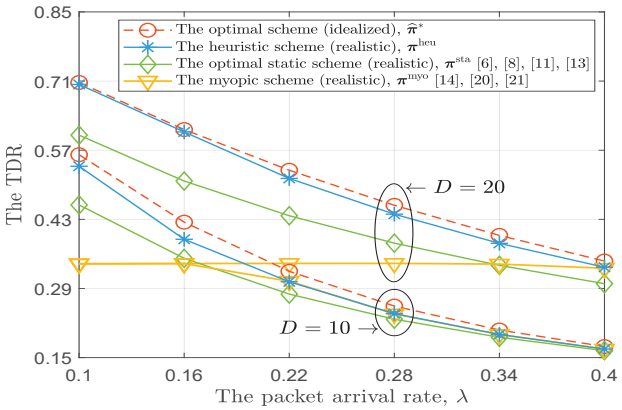

Fig. 8: The TDR as a function of the packet arrival rate λ for $N = 50$, $D = 10, 20$, $\sigma = 0.9$.

when $\sigma = 1$; $\boldsymbol{\pi}^{\text{heu}}$ performs close to $\widehat{\boldsymbol{\pi}}^*$: 3.30%–6.56% loss when $\sigma = 0.8$ and 3.17%–6.60% loss when $\sigma = 1$. It is shown that the relative gap between $\widehat{\boldsymbol{\pi}}^*$ and $\boldsymbol{\pi}^{\text{heu}}$ is small when D is much larger than $N\lambda$. The rationale is that, for this case in the realistic environment, n_t probably takes a value n much smaller than $D-t$ and it is easy for the active nodes to judge correctly whether $n < D-t$, which is crucial for $\boldsymbol{\pi}^{\text{heu}}$. We also observe that $\boldsymbol{\pi}^{\text{heu}}$ significantly outperforms $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda < D$: 6.15%–39.84% improvement when $\sigma = 0.8$ and 6.58%–39.72% improvement when $\sigma = 1$, and achieves almost the same TDR performance as $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda \geq D$. This confirms again that the inefficient time utilization of $\boldsymbol{\pi}^{\text{myo}}$, which is opposite to the insight behind the heuristic from Eq. (25), yields the poor TDR performance of $\boldsymbol{\pi}^{\text{myo}}$.

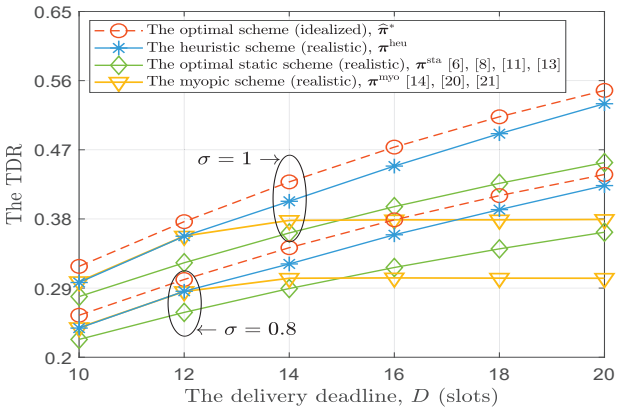

Fig. 9: The TDR as a function of the delivery deadline D for $N = 50$, $\lambda = 0.25$, $\sigma = 0.8, 1$.

Fig. 10 shows the TDR performance without retransmissions as a function of the packet success rate σ. The TDR of each scheme always increases with σ, due to the increase of channel quality. We observe that $\boldsymbol{\pi}^{\text{heu}}$ performs significantly better than $\boldsymbol{\pi}^{\text{sta}}$: 18.59%–19.03% improvement when $\lambda = 0.1$ and 5.51%–5.77% improvement when $\lambda = 0.4$, and performs close to $\widehat{\boldsymbol{\pi}}^*$: 0.70%–1.06% loss when $\lambda = 0.1$ and 4.09%–4.33% loss when $\lambda = 0.4$. We also observe that $\boldsymbol{\pi}^{\text{heu}}$ significantly outperforms $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda < D$: 89.67%–91.00% improvement when $\lambda = 0.1$; $\boldsymbol{\pi}^{\text{heu}}$ achieves almost the same TDR as $\boldsymbol{\pi}^{\text{myo}}$ when $N\lambda \geq D$. This confirms again the benefits of determining the transmission probability simultaneously based on the current delivery urgency and contention intensity. It is also shown that $\boldsymbol{\pi}^{\text{sta}}$ performs close to other schemes as $\frac{N\lambda}{D}$ becomes larger. This is because, n_t probably takes a value n much larger than $D-t$ and the knowledge of n has a minor impact on TDR due to large n.

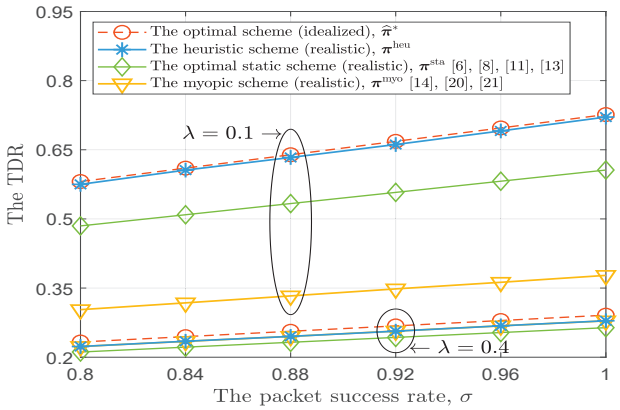

Fig. 10: The TDR as a function of the packet success rate σ for $N = 50$, $\lambda = 0.1, 0.4$, $D = 15$.

B. Comparisons With Retransmissions

Figs. 11–13 show the TDR performance with retransmissions as functions of λ, D and σ, respectively. We observe from Fig. 11 that $\boldsymbol{\pi}^{\text{heuR}}$ with $K = K^{\text{heuR}*}$ outperforms $\boldsymbol{\pi}^{\text{staR}}$ with $K = K^{\text{staR}*}$: 2.06%–9.29% improvement when $D = 10$ and 3.42%–11.04% improvement when $D = 20$. We observe from Fig. 12 that $\boldsymbol{\pi}^{\text{heuR}}$ with $K = K^{\text{heuR}*}$ enjoys 5.64%–10.46% improvement when $\sigma = 0.8$ and 6.58%–11.25% improvement when $\sigma = 1$. We observe from Fig. 13 that $\boldsymbol{\pi}^{\text{heuR}}$ with $K = K^{\text{heuR}*}$ enjoys 1.80%–3.74% improvement when $\lambda = 0.1$ and 5.61%–5.95% improvement when $\lambda = 0.4$. These results indicate that the idea of generalizing $\boldsymbol{\pi}^{\text{heu}}$ to $\boldsymbol{\pi}^{\text{heuR}}$ is effective in improving the TDR for a wide range of configurations when retransmissions are allowed. Next, we can see that, in general, $K^{\text{heuR}*}$ increases when λ decreases, D increases, or σ decreases. This is because appropriately introducing more retransmissions would be useful to improve the time utilization for smaller λ and larger D, and to resist the risk of transmission failures for smaller σ. Moreover, it is shown that $K^{\text{heuR}*} \leq K^{\text{staR}*}$ in all the cases, which is desirable for the sake of energy efficiency. This is because dynamic control is helpful to improve the delivery reliability of a single transmission and alleviate the need of retransmissions. It also implies that although the performance gap in Figs. 11–13 is less notable than that when $K = 1$, such a phenomenon is caused by allowing the static scheme to unfairly perform more retransmissions.

VIII. Conclusion

In this paper, under the idealized and realistic environments without retransmissions, optimal dynamic control schemes for random access in deadline-constrained broadcasting with

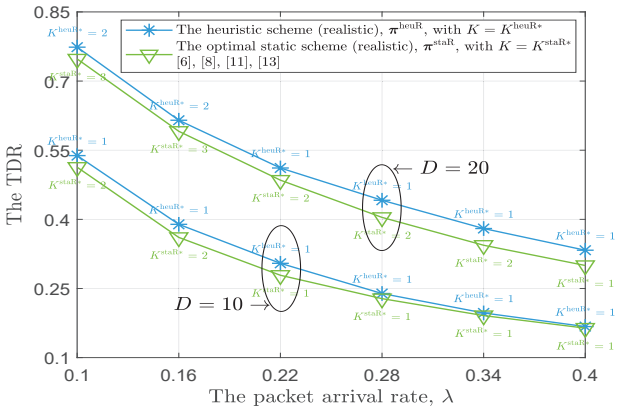

Fig. 11: The TDR as a function of the packet arrival rate λ for $N = 50$, $D = 10, 20$, $\sigma = 0.9$.

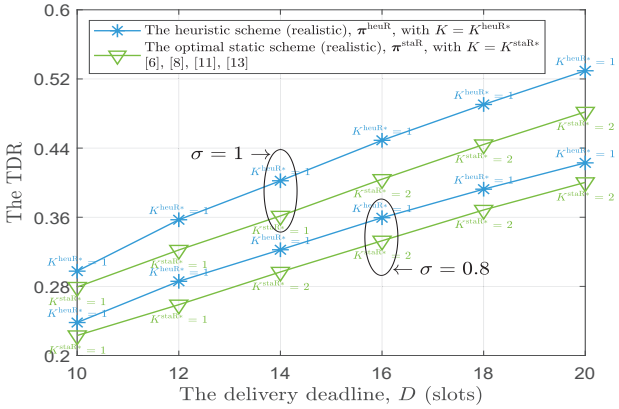

Fig. 12: The TDR as a function of the delivery deadline D for $N = 50$, $\lambda = 0.25$, $\sigma = 0.8, 1$.

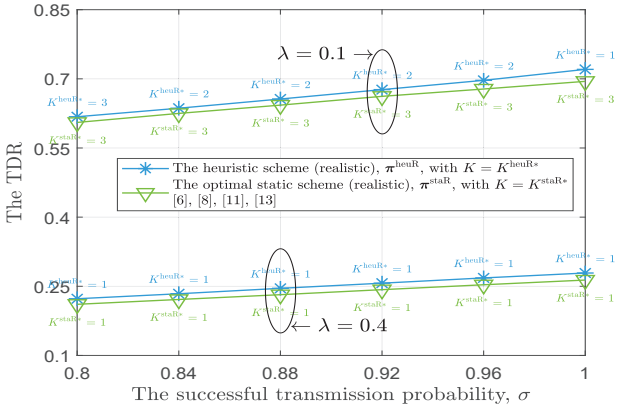

Fig. 13: The TDR as a function of the packet success rate σ for $N = 50$, $\lambda = 0.1, 0.4$, $D = 15$.

frame-synchronized traffic have been investigated based on the theories of MDP and POMDP, respectively. The proposed heuristic scheme for the realistic environment is able to achieve the threefold goal of being implemented without imposing extra overhead and hardware cost, of being implemented with very low computational complexity, and of achieving TDR close to the maximum achievable TDR in the idealized environment. Moreover, it has been shown that the proposed heuristic scheme can be easily extended to incorporate retransmissions for further improving the TDR.

An interesting and important future research direction is to optimize deadline-constrained broadcasting under general traffic patterns. To handle with such an asymmetric scenario, a standard way is to develop a dynamic control scheme based on the theory of decentralized MDP (Dec-MDP). However, solving a Dec-MDP is in general NEXP-complete [34]. Hence, an appropriate practical scheme needs to be further studied, which is our ongoing work.

ACKNOWLEDGEMENT

The authors would like to thank Dr. He Chen for helpful suggestions and discussions.

REFERENCES

[1] D. Feng, C. She, K. Ying, L. Lai, Z. Hou, T. Q. S. Quek, Y. Li, and B. Vucetic, "Toward ultrareliable low-latency communications: Typical scenarios, possible solutions, and open issues," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 94–102, 2019.

[2] J. Gao, M. Li, L. Zhao, and X. Shen, "Contention intensity based distributed coordination for V2V safety message broadcast," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12288–12301, 2018.

[3] M. Luvisotto, Z. Pang, and D. Dzung, "High-performance wireless networks for industrial control applications: New targets and feasibility," *Proc. IEEE*, vol. 107, no. 6, pp. 1074–1093, 2019.

[4] C. Chen, H. Li, H. Li, R. Fu, Y. Liu, and S. Wan, "Efficiency and fairness oriented dynamic task offloading in Internet of Vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1481–1493, 2022.

[5] Y. Zhao, X. Wang, C. Wang, Y. Cong, and L. Shen, "Systemic design of distributed multi-UAV cooperative decision-making for multi-target tracking," *Auton. Agents Multi-Agent Syst.*, vol. 33, no. 1, pp. 132–158, 2019.

[6] Y. H. Bae, "Analysis of optimal random access for broadcasting with deadline in cognitive radio networks," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 573–575, 2013.

[7] Y. H. Bae, "Random access scheme to improve broadcast reliability," *IEEE Commun. Lett.*, vol. 17, no. 7, pp. 1467–1470, 2013.

[8] Y. H. Bae, "Queueing analysis of deadline-constrained broadcasting in wireless networks," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1782–1785, 2015.

[9] C. Campolo, A. Vinel, A. Molinaro, and Y. Koucheryavy, "Modeling broadcasting in IEEE 802.11p/WAVE vehicular networks," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 199–201, 2011.

[10] M. I. Hassan, H. L. Vu, and T. Sakurai, "Performance analysis of the IEEE 802.11 MAC protocol for DSRC with and without retransmissions," in *Proc. IEEE WoWMoM*, 2010, pp. 1–8.

[11] Y. H. Bae, "Optimal retransmission-based broadcasting under delivery deadline constraint," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 1041–1044, 2015.

[12] Y. H. Bae, "Modeling timely-delivery ratio of slotted aloha with energy harvesting," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1823–1826, 2017.

[13] L. Deng, J. Deng, P. Chen, and Y. S. Han, "On the asymptotic performance of delay-constrained slotted ALOHA," in *Proc. IEEE ICCCN*, 2018, pp. 1–8.

[14] Y. H. Bae and J. W. Baek, "Age of information and throughput in random access-based IoT systems with periodic updating," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 821–825, 2022.

[15] L. Zhao, X. Chi, L. Qian, and W. Chen, "Analysis on latency-bounded reliability for adaptive grant-free access with multipackets reception (MPR) in URLLCs," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 892–895, 2019.

[16] Y. Zhang, Y. Lo, F. Shu, and J. Li, "Achieving maximum reliability in deadline-constrained random access with multiple-packet reception," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5997–6008, 2019.

[17] C. Boyd, R. Vehkalahti, O. Tirkkonen, and A. Laaksonen, "Code design principles for ultra-reliable random access with preassigned patterns," in *Proc. IEEE ISIT*, 2019, pp. 2604–2608.

[18] Y.-H. Lo, K. W. Shum, W. S. Wong, and Y. Zhang, "Multichannel conflict-avoiding codes of weights three and four," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3557–3568, 2020.

[19] B. Hajek and T. Van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Automat. Contr.*, vol. 27, no. 3, pp. 559–569, 1982.

[20] R. Rivest, "Network control by Bayesian broadcast," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 3, pp. 323–328, 1987.

[21] G. del Angel and T. L. Fine, "Optimal power and retransmission control policies for random access systems," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1156–1166, 2004.

[22] L. Bononi, M. Conti, and E. Gregori, "Runtime optimization of IEEE 802.11 wireless LANs performance," *IEEE Trans. Parallel Distrib. Syst.*, vol. 15, no. 1, pp. 66–80, 2004.

[23] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1904–1917, 2013.

[24] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Oper. Res.*, vol. 21, no. 5, pp. 1071–1088, 1973.

[25] Y. Zhang, A. Gong, Y. Lo, J. Li, F. Shu, and W. S. Wong, "Generalized p-persistent CSMA for asynchronous multiple-packet reception," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6966–6979, 2019.

[26] A. Biason, S. Dey, and M. Zorzi, "A decentralized optimization framework for energy harvesting devices," *IEEE Trans. Mob. Comput.*, vol. 17, no. 11, pp. 2483–2496, 2018.

[27] L. Deng, F. Liu, Y. Zhang, and W. S. Wong, "Delay-constrained topology-transparent distributed scheduling for MANETs," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1083–1088, 2021.

[28] *A 5G traffic model for industrial use cases*. White Paper, 5G Alliance for Connected Industries and Automation, 2019.

[29] E. Adamey and U. Ozguner, "Cooperative multitarget tracking and surveillance with mobile sensing agents: A decentralized approach," in *Proc. IEEE ITSC*, 2011, pp. 1916–1922.

[30] J. Capitan, L. Merino, and A. Ollero, "Cooperative decision-making under uncertainties for multi-target surveillance with multiples UAVs," *J. Intell. Robot Syst.*, vol. 84, no. 1, pp. 371–386, 2016.

[31] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[32] P. R. Kumar and P. Varaiya, *Stochastic systems: Estimation, identification, and adaptive control*. SIAM, 2015.

[33] D. Wu, L. Deng, Z. Liu, Y. Zhang, and Y. S. Han, "Reinforcement learning random access for delay-constrained heterogeneous wireless networks: A two-user case," in *Proc. IEEE Globecom Workshops*, 2021, pp. 1–7.

[34] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," *Math. Oper. Res.*, vol. 27, no. 4, pp. 819–840, 2002.

Supplemental Material for the paper "Dynamic Optimization of Random Access in Deadline-Constrained Broadcasting"

Aoyu Gong, Yijin Zhang, Lei Deng, Fang Liu, Jun Li, and Feng Shu

PROOF OF EQ. (27)

Letting $f\big((M,\alpha),p\big) \triangleq \frac{(M+1)\alpha}{\sigma}\sum_{n\in\mathcal{N}} b^{\mathrm{bd}}(n) r_t\big((1,n),p\big)$ for $p\in[0,1]$ and $c_i = \binom{M+1}{i}\alpha^i(1-\alpha)^{(M+1-i)}$ for $1\le i\le M+1$, we have

$$f\big((M,\alpha),p\big) = \sum_{i=1}^{M+1} i c_i p(1-p)^{i-1}.$$

The derivative of $f\big((M,\alpha),p\big)$ with respect to p is given by

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}p} f\big((M,\alpha),p\big)\\
&= \sum_{i=1}^{M+1} i c_i (1-p)^{i-1} - \sum_{i=2}^{M+1} i(i-1)c_i p(1-p)^{i-2}\\
&= (M+1)\alpha + (M+1)^2\alpha^{M+1}(-p)^M + \sum_{j=1}^{M-1}\beta_j p^j, \quad (28)
\end{aligned}$$

where β_j, $1\le j\le M-1$, is derived as follows:

$$\begin{aligned}
\beta_j &= (-1)^j \sum_{k=1}^{M+1-j}\binom{M+1}{j+k}\alpha^{j+k}(1-\alpha)^{M+1-j-k}\\
&\quad\cdot (j+k)\left[\binom{j+k-1}{j} + (j+k-1)\binom{j+k-2}{j-1}\right]\\
&= (-1)^j (j+1)^2\alpha^{j+1}\\
&\quad\cdot \sum_{k=1}^{M+1-j}\binom{j+k}{k-1}\binom{M+1}{j+k}\alpha^{k-1}(1-\alpha)^{M+1-j-k}\\
&= (-1)^j (j+1)^2\alpha^{j+1}\binom{M+1}{j+1}\\
&\quad\cdot \sum_{k=1}^{M+1-j}\binom{M-j}{k-1}\alpha^{k-1}(1-\alpha)^{M-j-k+1}\\
&= (-1)^j\alpha^{j+1}\\
&\quad\cdot \big(j(M+1)^2 + (M+1)(M-j)\big)\frac{(M-1)!}{(M-j)!j!}\\
&= (-1)^j\alpha^{j+1}\\
&\quad\cdot \left[(M+1)^2\binom{M-1}{j-1} + (M+1)\binom{M-1}{j}\right]. \quad (29)
\end{aligned}$$

Combining Eqs. (28) and (29), we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}p} f\big((M,\alpha),p\big)\\
&= (M+1)\alpha + (M+1)^2\alpha^{M+1}(-p)^M\\
&\quad + \sum_{j=1}^{M-1}\left[(M+1)^2\binom{M-1}{j-1} + (M+1)\binom{M-1}{j}\right]\\
&\qquad\cdot \alpha^{j+1}(-p)^j\\
&= (M+1)\alpha\big(1-(M+1)\alpha p\big)\\
&\quad + (M+1)\alpha\big(1-(M+1)\alpha p\big)(-\alpha p)^{M-1}\\
&\quad + \sum_{j=1}^{M-2}\binom{M-1}{j}(M+1)\alpha\big(1-(M+1)\alpha p\big)(-\alpha p)^j\\
&= (M+1)\alpha\big(1-(M+1)\alpha p\big)\big(1-\alpha p\big)^{M-1}. \quad (30)
\end{aligned}$$

From Eq. (30), for $p\in[0,1]$, we obtain that $f\big((M,\alpha),p\big) \le f\big((M,\alpha),\frac{1}{M\alpha+\alpha}\big)$ when $\frac{1}{M\alpha+\alpha}\le 1$, and $f\big((M,\alpha),p\big) \le f\big((M,\alpha),1\big)$ when $\frac{1}{M\alpha+\alpha} > 1$.

Hence we complete the proof for Eq. (27).