Dynamic Uploading Scheduling in mmWave-Based Sensor Networks via Mobile Blocker Detection

Yifei Sun[†‡], Bojie Lv[†], Rui Wang[†], Haisheng Tan[§], and Francis C. M. Lau[‡]
[†]Southern University of Science and Technology, Shenzhen, China
[‡]The University of Hong Kong, Hong Kong, China
[§]University of Science and Technology of China, Hefei, China

Abstract—The freshness of information, measured as Age of Information (AoI), is critical for many applications in next-generation wireless sensor networks (WSNs). Due to its high bandwidth, millimeter wave (mmWave) communication is seen to be frequently exploited in WSNs to facilitate the deployment of bandwidth-demanding applications. However, the vulnerability of mmWave to user mobility typically results in link blockage and thus postponed real-time communications. In this paper, joint sampling and uploading scheduling in an AoI-oriented WSN working in mmWave band is considered, where a single human blocker is moving randomly and signal propagation paths may be blocked. The locations of signal reflectors and the real-time position of the blocker can be detected via wireless sensing technologies. With the knowledge of blocker motion pattern, the statistics of future wireless channels can be predicted. As a result, the AoI degradation arising from link blockage can be forecast and mitigated. Specifically, we formulate the long-term sampling, uplink transmission time and power allocation as an infinite-horizon Markov decision process (MDP) with discounted cost. Due to the curse of dimensionality, the optimal solution is infeasible. A novel low-complexity solution framework with guaranteed performance in the worst case is proposed where the forecast of link blockage is exploited in a value function approximation. Simulations show that compared with several heuristic benchmarks, our proposed policy, benefiting from the awareness of link blockage, can reduce average cost up to 49.6%.

Index Terms—Markov decision process, millimeter wave, age of information, wireless sensor networks

I. INTRODUCTION

In wireless sensor networks (WSNs), the timeliness of status information sampled at the sensor locations (e.g., camera snapshots or broadband radar signals) is critical for effective monitoring in many real-time and data-intensive applications. With the ever-increasing dense sensor deployment and bandwidth demand, communication in millimeter wave (mmWave) band has great potential to be integrated in WSNs due to its wide spectrum. However, because of the sparse propagation paths and the pencil-shaped beams for overcoming the extremely high path loss, mmWave communications may suffer from the blockage problem in dynamic environments. This will cause severe channel fluctuation and link outage, making the delivered status information outdated [1].

To quantify the freshness of the status information, the concept of *Age of Information* (AoI) has been adopted as a metric [2]. AoI is defined as the time elapsed since the status update is sampled at the sensor. From the joint sensing and transmission scheduling perspective, the AoI of samples delivered from the sensors should be as fresh as possible in order to guarantee accurate monitor and synchronous control. There have been a number of research works on AoI scheduling design in various WSNs [3]–[9]. In [3], the authors minimized the long-term average AoI of a single sensor with an energy harvesting battery. The scheduling design was formulated as a continuous time stochastic control problem. In [4], the problem of status update control in an EH-enabled source via an mmWave link was formulated as a Markov decision process (MDP). In addition to the single sensor scenario, the works in [5], [6] extended the joint sampling and transmission design to the multi-sensor scenario. [5] minimized the summation of average AoIs of multiple sensors under the sampling and transmission energy constraints via an MDP formulation. This work was further extended in [6] by considering transmission failures and non-uniform sample size. However, the algorithms proposed in many of the existing works, such as the Q-learning algorithm used in [4] and the policy iteration algorithm used in [9], are prohibitive in computation complexity which grows exponentially with respect to the number of sensors. Although, the linear approximation of value functions adopted in [5], [6] can reduce the computation complexity, it is difficult to analyze their performance analytically. As a result, how to design approximate MDP (AMDP) for AoI-oriented scheduling design with a guarantee on worst-case performance and low computational complexity is still unanswered.

Because of the small wavelength, the link blockage has become one of the open issues in mmWave communications, especially in indoor scenarios where swarms of mobile users may roam around constantly. Fortunately, with the technique of wireless sensing, the room layout can be reconstructed in advance [10]–[12], and the mobility of the human body can be detected in real time [13], such that potential link degradation due to blockage can be predicted. To our best knowledge, there exists no previous work on the AoI-oriented scheduling design problem exploiting the environment and human motion detection. In fact, all the existing works, e.g., [4]–[6], [8], [9], adopted the stochastic channel models, which are oblivious to

the signal propagation environment [14]. Hence, link blockage due to human mobility can hardly be captured.

In order to exploit wireless sensing in AoI-oriented scheduling design, we consider joint sampling and uploading scheduling in an mmWave-based sensor network with awareness of the communication environment which includes the static signal reflectors and one dynamic human blocker. Specifically, in the mmWave-based WSN, the sensors detect their targets, save the sensing samples in their buffers, and deliver the data to a server via an mmWave uplink. A human blocker is moving randomly, which blocks some signal paths, causing a significant degradation in the corresponding uplink channel. Sampling at the sensor and uplink transmission are jointly scheduled to minimize the AoI of the data samples at the monitor and energy consumption for sampling and uplink transmission. The joint scheduling is formulated as an infinite-horizon MDP with discounted cost, so that knowledge of the mmWave propagation channel and human dynamics can be exploited. The main contributions of this work are summarized as follows:

- A predictive scheduling framework is provided for environment-aware transmission scheduling. In the proposed MDP formulation, the future AoI degradation due to potential link blockage is naturally considered in the current scheduling according to the Bellman's equations.

- We propose a low-complexity AMDP framework with a guarantee of the worst-case performance. Specifically, we first introduce a decoupling principle to design heuristic scheduling polices as the reference policy, whose average cost (i.e., value function) can be derived analytically. With the expression of the value function, the above policy iteration can be formulated analytically, whose optimization efficiency is significantly better than the conventional numerical search. Simulations show that compared with several heuristic benchmarks, our proposed policy, benefiting from the awareness of the link blockage, can reduce the average cost with a high performance gain.

The remainder of this paper is structured as follows. The system model is introduced in Section II. In Section III, the problem of dynamic sampling and uploading scheduling is formulated as an infinite-horizon MDP. In Section IV, a low-complexity suboptimal scheduling framework is proposed with a guarantee on the worst-case performance. Finally, the performance of the proposed low-complexity scheduling scheme is verified by comparing with benchmarks in Section V, and the conclusion is drawn in Section VI.

II. SYSTEM MODEL

In this section, we first give an overview of the mmWave-based WSN including the blocker mobility model, and then define the channel model as well as the AoI model. We use the following notation throughout this paper. Bold lowercase $\mathbf{a}$ denotes a column vector, bold uppercase $\mathbf{A}$ denotes a matrix, non-bold letters a and A denote scalar values, and the letter $\mathcal{A}$ denotes a set. $(a)^+$ denotes $\max(0, a)$. $[\mathbf{A}]_{i,j}$, $\mathbf{A}^{\mathsf{T}}$, and
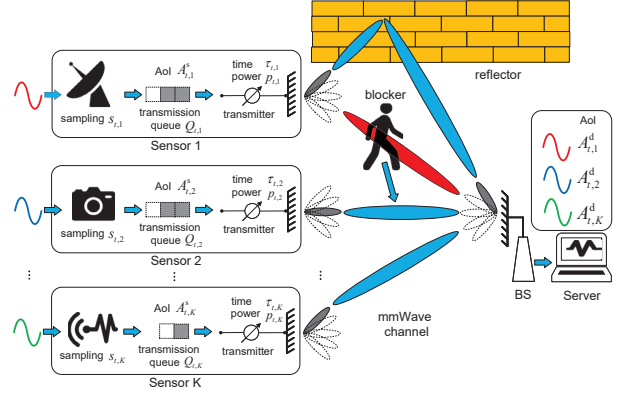

Fig. 1. Network model of the considered mmWave-based WSN.

Symbol	Description
$K/\mathcal{K}$	Number/Set of sensors
$\mathbf{H}_{t,k}$	Channel matrix of k-th sensor in t-th frame
l_t^{B}	Location index of the mobile blocker in t-th frame
$Y_{t,k}$	Baseband channel power gain of k-th sensor in t-th frame
$Q_{t,k}$	Queue length of k-th sensor in t-th frame
$A_{t,k}^{\mathrm{s}}$	AoI at k-th sensor in t-th frame
$A_{t,k}^{\mathrm{d}}$	AoI for k-th sensor at the server in t-th frame
$s_{t,k}$	Sampling action for k-th sensor in t-th frame
$\tau_{t,k}$	Transmission time allocated to k-th sensor in t-th frame
$p_{t,k}$	Transmission power of k-th sensor in t-th frame

TABLE I. Main notations.

$\mathbf{A}^{\mathsf{H}}$ denote the (i, j)-th element, transpose, and conjugate transpose, respectively. $\mathcal{CN}(m, R)$ denotes complex Gaussian distribution with mean m and variance R. $\mathcal{U}[a, b]$ denotes a uniform distribution over the interval $[a, b]$. $\mathbb{E}[.]$ denotes an expectation operator. $\mathbb{I}[.]$ denotes an indicator function. The main notations used are listed in Table I.

A. Network Model

As illustrated in Fig. 1, we consider an mmWave-based wireless monitoring system in a two-dimensional indoor space consisting of one server connected with the base station (BS) and K sensors, where the sensors sense the targets (e.g., taking photos or detecting motions via radar waves), collect sensing samples, and deliver the sensing samples to the server via the BS. The set of sensors is denoted as $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. The sensing sample collection, namely sampling, and sample uploading are scheduled to maintain the timeliness of sensing data at the server. Both line-of-sight (LoS) and non-line-of-sight (NLoS) paths (e.g., reflection paths from the walls) exist between the sensors and the BS. There is one person moving in the space, who may block the propagation paths from the sensors to the BS. In this work, we consider the case of only one mobile blocker in the ambient environment which represents most practical scenarios while leave multi-blocker scenarios to future work. With wireless sensing techniques [15], [16], it is assumed that the location of the human blocker can be detected. The analog MIMO architecture with single radio frequency (RF) chain and a half-wavelength uniform linear phased array (ULA) is adopted at both the BS and the sensors, so that both transmission and receiving beams can

be aligned to the available paths. The linear phased arrays at the BS and sensors have N_R and N_T antenna elements, respectively.

The uplink transmission time is organized by physical-layer frames with duration T_F, where the channel state information (CSI) is assumed to be quasi-static in one frame. The locations of the indoor space are quantized into grids with indexes. Let $\mathbb{L} \triangleq \{1, 2, \ldots, |\mathbb{L}|\}$ be the set of location indexes. The location indexes of the BS and the k-th sensors are denoted as l^{BS} and l_k^s, respectively. The location index of the mobile human blocker in the t-th frame is denoted as l_t^B. It is assumed that the mobility of the blocker follows a time-invariant Markov chain, with the following transition probabilities,

$$\Pr\left[l_{t+1}^B = \ell' | l_t^B = \ell\right] = \left[\mathbf{P}^B\right]_{\ell, \ell'}, \ \forall t, \ \forall \ell, \ell' \in \mathbb{L}, \quad (1)$$

where $\mathbf{P}^B \in \mathbb{R}^{|\mathbb{L}| \times |\mathbb{L}|}$ denotes the transition matrix of the blocker's mobility.

B. Channel Model

In order to capture the impact of human blockage, the geometric channel model [17] is adopted in this paper. There are at most M NLoS paths and one LoS path from one sensor to the BS. Specifically, denote $\mathcal{M} = \{0, 1, 2, \ldots, M\}$ as the index set of propagation paths, where the index of the LoS path is 0. For the convenience of exposition, the i-th path of the k-th sensor is denoted as the (k, i)-th path. Hence, the channel matrix $\mathbf{H}_{t,k} \in \mathbb{C}^{N_R \times N_T}$ from the k-th sensor to the BS in the t-th frame can be written as

$$\mathbf{H}_{t,k} = \sum_{i \in \mathcal{M}} B_{t,k,i}(l_t^B) \alpha_{t,k,i} \mathbf{a}_R(\phi_{k,i}) \mathbf{a}_T^H(\theta_{k,i}), \quad (2)$$

where $B_{t,k,i}(l_t^B) \in \{0, 1\}$ is the blockage indicator. Thus, $B_{t,k,i}(l_t^B) = 0$ when (k, i)-th path is blocked, otherwise $B_{t,k,i}(l_t^B) = 1$. Moreover, $\phi_{k,i}$ and $\theta_{k,i}$ denote the angle of arrival (AoA) and angle of departure (AoD) of the (k, i)-th path, respectively. $\alpha_{t,k,i}$ denotes the complex gain of the (k, i)-th path in the t-th frame obeying a complex Gaussian distribution with zero mean and variance $\rho_{k,i}^{-1}$, i.e., $\alpha_{t,k,i} \sim \mathcal{CN}(0, \rho_{k,i}^{-1})$. The path loss $\rho_{k,i}$ depends on the path length R and the reflection loss (for NLoS path). Moreover, $\mathbf{a}_R(\phi_{k,i})$ and $\mathbf{a}_T(\theta_{k,i})$ represent the normalized array response vectors of the ULAs at the sensors and the BS, respectively. Thus,

$$\mathbf{a}_R(\phi_{k,i}) = \frac{1}{\sqrt{N_R}} \left[1, e^{-j\pi \sin(\phi_{k,i})}, \ldots, e^{-j\pi(N_R-1)\sin(\phi_{k,i})}\right]^T,$$

$$\mathbf{a}_T(\theta_{k,i}) = \frac{1}{\sqrt{N_T}} \left[1, e^{-j\pi \sin(\theta_{k,i})}, \ldots, e^{-j\pi(N_T-1)\sin(\theta_{k,i})}\right]^T.$$

The human blocker is modeled as a disk with radius r_B, and hence, the indicator $B_{t,k,i}$ can be determined by comparing the blocker radius with the shortest distances between the blocker's centroid and the i-th path. Let $\mathcal{P}_{k,i}$ be the set of line segments in the (k, i)-th path, and $\mathbb{D}(\ell, \mathcal{P}_{k,i})$ be the minimum distance from the disk centroid to the line segments of the (k, i)-th path, then $B_{t,k,i}(l_t^B) = \mathbb{I}[\mathbb{D}(l_t^B, \mathcal{P}_{k,i}) \geq r_B]$. The path loss of the LoS path is usually much smaller than that of the

NLoS paths. Therefore, the uplink transmission suffers from significant degradation when the LoS path is blocked.

Let $\mathbf{w}_{t,k} \in \mathbb{C}^{N_R \times 1}$ and $\mathbf{f}_{t,k} \in \mathbb{C}^{N_T \times 1}$ be the analog combiner and precoder of the BS and the k-th sensor when the k-th sensor is transmitting in the t-th frame, respectively. The uplink capacity of the k-th sensor in the t-th frame can be expressed as

$$R_{t,k} \triangleq W \log_2\left(1 + p_{t,k} \underbrace{\frac{\left|\mathbf{w}_{t,k}^H \mathbf{H}_{t,k} \mathbf{f}_{t,k}\right|^2}{\|\mathbf{w}_{t,k}\|^2 N_0 W}}_{\text{Baseband gain } Y_{t,k}}\right), \quad (3)$$

where $Y_{t,k}(l_t^B) \triangleq \frac{\left|\mathbf{w}_{t,k}^H \mathbf{H}_{t,k} \mathbf{f}_{t,k}\right|^2}{\|\mathbf{w}_{t,k}\|^2 N_0 W}$ denotes the baseband channel power gain, $p_{t,k}$ denotes the transmission power of the k-th sensor, N_0 is the noise power spectral density, W is the bandwidth. Due to the hardware constraint, the maximum transmission power satisfies

$$p_{t,k} \leq P_{\max}, \ \forall t, k \in \mathcal{K}. \quad (4)$$

The analog precoders and combiners are chosen from a pre-defined codebook composed of a finite number of beam directions, as follows,

$$\mathbf{w}_{t,k} \in \mathcal{W} \triangleq \{\mathbf{a}_R(\phi_q), q = 1, 2, \ldots, N_R\}, \quad (5)$$

$$\mathbf{f}_{t,k} \in \mathcal{F} \triangleq \{\mathbf{a}_T(\phi_p), p = 1, 2, \ldots, N_T\}, \quad (6)$$

where $\phi_q = \arcsin\left(\frac{2(q-1)}{N_R} - 1\right)$, and $\theta_p = \arcsin\left(\frac{2(p-1)}{N_T} - 1\right)$. In order to avoid the costly channel matrix estimation in every frame, the transmission and receiving beams are adapted to maximize the average signal-to-noise ratio (SNR) instead of instantaneous SNR, as follows,

$$(\mathbf{w}_{t,k}, \mathbf{f}_{t,k}) = \underset{\substack{\mathbf{a}_R(\phi_q) \in \mathcal{W} \\ \mathbf{a}_T(\phi_p) \in \mathcal{F}}}{\arg\max} \mathbb{E}_{\mathbf{H}_{t,k}} \left|\mathbf{a}_R^H(\phi_q)\mathbf{H}_{t,k}\mathbf{a}_T(\theta_p)\right|^2. \quad (7)$$

Since the AoAs and AoDs of the paths in (2) are static, they can be estimated before the scheduling, as in [10], [11]. The variances of path gains $\{\alpha_{k,i}, \forall k, i\}$, can also be measured before the transmission [11]. Given the location of the human blocker by real-time localization [12], the distributions of channel matrices $\{\mathbf{H}_{t,k}, \forall t, k\}$ can be predicted. Hence, the optimization in (7) is feasible even if $\mathbf{H}_{t,k}$ is unknown. Intuitively, (7) would align the transmission and receiving beams along the path with minimum path loss.

Time-Division Multiple Access (TDMA) is adopted in each frame. Let $\tau_{t,k}$ be the transmission time allocated to the k-th sensor in the t-th frame; the following constraints should be satisfied,

$$\sum_{k \in \mathcal{K}} \tau_{t,k} = T_F, \ \forall t, \quad (8)$$

$$0 \leq \tau_{t,k} \leq T_F, \ \forall t, k \in \mathcal{K}. \quad (9)$$

Moreover, the throughput of the k-th sensor in the t-th frame is $\tau_{t,k} R_{t,k}$.

C. AoI Model for Sample Uploading

If the server delivers a sensing decision to the k-th sensor at the beginning of the t-th frame, a sensing sample representing

the latest target status is generated by the sensor. Since the sensing decision is usually short, we ignore the downlink signalling overhead. It is assumed that the data volume of each sample generated by the k-th sensor consists of L_k packets, each with $N_{\rm b}$ information bits. Let $Q_{t,k}$ be the length of the uplink transmission queue (number of uplink packets) of the k-th sensor in the t-th frame, where the sensing sample is buffered. When new data sample is generated at one sensor, the existing packets from the previous sampling in its uplink queue will be dropped, and new packets will be added. The sampling action will induce a constant sampling energy cost which is denoted by $C^{\rm s}$. Let $s_{t,k} \in \{0,1\}$ be the sampling action for the k-th sensor in the t-th frame, where $s_{t,k} = 0$ indicates that the k-th sensor will continue transmitting data packets of its current sample in the t-th frame, and $s_{t,k} = 1$ indicates that the k-th sensor will start transmitting the packets of the new sample. Thus,

$$Q_{t+1,k} = \begin{cases} (Q_{t,k} - D_{t,k})^+, & s_{t,k} = 0 \\ L_k - D_{t,k}, & s_{t,k} = 1 \end{cases} \quad (10)$$

where $D_{t,k}$ denotes the departure packet number of the k-th sensor in the t-th frame. Thus, $D_{t,k} = \left\lfloor \frac{R_{t,k}\tau_{t,k}}{N_{\rm b}} \right\rfloor$.

To characterize the freshness of samples, the AoI at the k-th sensor, denoted by $A^{\rm s}_{t,k}$, is defined as the number of frames since the generation time of the latest sample. Moreover, since samples with extremely large AoIs will bring no benefit to the server due to its outdatedness, we set an AoI threshold $A_{\max}$ to characterize outdated samples. The AoI dynamics at the k-th sensor is given by

$$A^{\rm s}_{t+1,k} = \begin{cases} \min\{A^{\rm s}_{t,k} + 1, A_{\max}\}, & s_{t,k} = 0 \\ 1, & s_{t,k} = 1 \end{cases} \quad (11)$$

Besides the AoIs at the sensors, the server records AoIs of the latest uploaded samples of all sensors. Once the transmission of all packets at the k-th sensor is accomplished, the AoI at the server for the k-th sensor will be updated by the AoI at the k-th sensor in next frame. Hence, let $A^{\rm d}_{t,k}$ be the AoI for the k-th sensor at the server in the t-th frame, we have

$$A^{\rm d}_{t+1,k} = \begin{cases} \min\{A^{\rm s}_{t,k} + 1, A_{\max}\}, & (Q_{t,k} - D_{t,k})^+ = 0 \\ \min\{A^{\rm d}_{t,k} + 1, A_{\max}\}, & \text{otherwise} \end{cases} \quad (12)$$

III. PROBLEM FORMULATION

The average AoI at the server depends on the sampling action, time and uplink power allocations of all the frames. Clearly, due to the random motion of the human blocker and random channel matrices, it is impossible for the BS to determine the sampling action, time and uplink power allocations of all the frames in a deterministic manner. Instead, we shall formulate their optimization as an infinite-horizon MDP. The system state, scheduling policy, and system cost are first defined below.

Definition 1 (System State). *At the beginning of the t-th frame, the global system state is uniquely specified by a tuple $\mathcal{S}_t \triangleq (l^{\rm B}_t, \mathcal{Y}_t, \mathcal{Q}_t, \mathcal{A}^{\rm s}_t, \mathcal{A}^{\rm d}_t)$, where $\mathcal{Y}_t \triangleq \{Y_{t,k}\}_{k \in \mathcal{K}}$ is the set of*

baseband channel power gains of all sensors, $\mathcal{Q}_t \triangleq \{Q_{t,k}\}_{k \in \mathcal{K}}$ is the set of remaining numbers of packets at the uplink queues, $\mathcal{A}^{\rm s}_t \triangleq \{A^{\rm s}_{t,k}\}_{k \in \mathcal{K}}$ is the set of AoIs at the sensors, and $\mathcal{A}^{\rm d}_t \triangleq \{A^{\rm d}_{t,k}\}_{k \in \mathcal{K}}$ is the set of AoIs at the server. Moreover, the local system state of the k-th sensor in the t-th frame is defined by $\mathcal{S}_{t,k} \triangleq (l^{\rm B}_t, Y_{t,k}, Q_{t,k}, A^{\rm s}_{t,k}, A^{\rm d}_{t,k})$.

Definition 2 (Action and Policy). *The local scheduling action of the k-th sensor is defined as $\mathbf{a}_{t,k} \triangleq (s_{t,k}, \tau_{t,k}, p_{t,k})$, including the sampling decision, and uplink transmission time and power. The global scheduling action is defined as the aggregation of the local actions of all the sensors; thus, $\mathbf{a}_t \triangleq \{\mathbf{a}_{t,k}\}_{k \in \mathcal{K}}$. Hence, the scheduling policy, denoted as Ω, is a mapping from the system state $\mathcal{S}_t$ to the scheduling actions, i.e., $\Omega(\mathcal{S}_t) = \mathbf{a}_t$.*

Given the scheduling policy Ω, the state transition probability can be written as in (13).

In the t-th frame, the per-frame cost is defined as the weighted sum of AoIs at the server, energy consumption for sampling and uplink transmission, and outdated AoI penalties at the server for all sensors. That is,

$$g(\mathcal{S}_t, \Omega(\mathcal{S}_t)) = \sum\nolimits_{k \in \mathcal{K}} \big[A^{\rm d}_{t,k} + w_{\rm P}(s_{t,k}C^{\rm s} + \tau_{t,k}p_{t,k}) \\ + w_{\rm Q}\mathbb{I}[A^{\rm d}_{t,k} = A_{\max}] \big], \quad (14)$$

where $w_{\rm P}$ and $w_{\rm Q}$ denote the weights for energy consumption and AoI outdatedness penalty, respectively. Note that the first and third terms of (14) imply a nonlinear cost function for AoI [18].

Hence, the overall cost from the 1-st frame is defined as

$$\overline{G}(\mathcal{S}_1, \Omega) \triangleq \lim_{T \to \infty} \left[\mathbb{E}_{\mathcal{Y}, \mathcal{L}^{\rm B}} \sum_{t=1}^{T} \gamma^{t-1} g_t(\mathcal{S}_t, \Omega(\mathcal{S}_t)) \Big| \mathcal{S}_1 \right], \quad (15)$$

where $\gamma \in (0, 1)$ is the discount factor, and the expectation is taken on $\mathcal{Y} \triangleq \{\mathcal{Y}_1, \mathcal{Y}_2 \ldots, \mathcal{Y}_T\}$ and $\mathcal{L}^{\rm B} \triangleq \{l^{\rm B}_1, l^{\rm B}_2, \ldots, l^{\rm B}_T\}$. As a result, the joint sampling and uploading optimization can be formulated by the following infinite-horizon MDP with discounted cost.

$$\text{P1}: \ \Omega^\star = \arg\min_\Omega \overline{G}(\mathcal{S}_1, \Omega)$$
$$\text{s.t. Constraints in (4), (8), (9).}$$

The Bellman's equations of the above MDP are

$$W(\mathcal{S}) = \min_{\Omega(\mathcal{S})} \big[g(\mathcal{S}, \Omega(\mathcal{S})) \\ + \gamma \sum\nolimits_{\mathcal{S}'} W(\mathcal{S}') \Pr[\mathcal{S}'|\mathcal{S}, \Omega(\mathcal{S})] \big], \forall \mathcal{S}, \quad (16)$$

where $W(\cdot)$ is the value function of the optimal scheduling policy (i.e., the optimal value function), and $\mathcal{S}'$ is the system state in the next frame given system state $\mathcal{S}$ and scheduling action $\Omega(\mathcal{S})$. Moreover, the policy minimizing the right-hand-side (RHS) of the above Bellman's equations is proved to be the optimal one [19].

Note that the baseband gain in the system state is continuously and independently distributed in all frames, which can be eliminated from the value function to reduce the complexity. Hence, we first define the local and global *abstract*

$$\Pr\left[\mathcal{S}_{t+1}|\mathcal{S}_t,\Omega(\mathcal{S}_t)\right] = \Pr\left[l_{t+1}^{\mathrm{B}}|l_t^{\mathrm{B}}\right]\Pr\left[\mathcal{Y}_{t+1}|l_{t+1}^{\mathrm{B}}\right]\times\prod_{k\in\mathcal{K}}\left\{(1-s_{t,k})\mathbb{I}\left[A_{t+1,k}^{\mathrm{s}}=\min\{A_{t,k}^{\mathrm{s}}+1,A_{\max}\}\right]+s_{t,k}\mathbb{I}\left[A_{t+1,k}=1\right]\right\}$$

$$\times\prod_{k\in\mathcal{K}}\left\{(1-s_{t,k})\mathbb{I}\left[Q_{t+1,k}=(Q_{t,k}-D_{t,k})^+\big|Y_{t,k},\tau_{t,k},p_{t,k}\right]+s_{t,k}\mathbb{I}\left[Q_{t+1,k}=L_k-D_{t,k}\big|Y_{t,k},\tau_{t,k},p_{t,k}\right]\right\}$$

$$\times\prod_{k\in\mathcal{K}}\left\{\mathbb{I}\left[(Q_{t,k}-D_{t,k})^+=0\right]\mathbb{I}\left[A_{t+1,k}^{\mathrm{d}}=\min\{A_{t,k}^{\mathrm{s}}+1,A_{\max}\}\right]+\mathbb{I}\left[(Q_{t,k}-D_{t,k})^+\neq0\right]\mathbb{I}\left[A_{t+1,k}^{\mathrm{d}}=\min\{A_{t,k}^{\mathrm{d}}+1,A_{\max}\}\right]\right\}. \quad (13)$$

---

state [20] with the baseband gain eliminated, i.e., $\widetilde{\mathcal{S}}_{t,k} \triangleq (l_t^{\mathrm{B}}, Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}})$ and $\widetilde{\mathcal{S}}_t \triangleq (l_t^{\mathrm{B}}, \mathcal{Q}_t, \mathcal{A}_t^{\mathrm{s}}, \mathcal{A}_t^{\mathrm{d}})$. By taking expectation on both sides of (16), the Bellman's equations with respect to the abstract state can be simplified as

$$W(\widetilde{\mathcal{S}}) = \mathbb{E}_{\mathcal{Y}}\min_{\Omega(\mathcal{S})}\left[g\big(\mathcal{S},\Omega(\mathcal{S})\big)\right.$$
$$\left.+\gamma\sum_{\widetilde{\mathcal{S}}'}W(\widetilde{\mathcal{S}}')\Pr[\widetilde{\mathcal{S}}'|\mathcal{S},\Omega(\mathcal{S})]\right], \forall\widetilde{\mathcal{S}}, \quad (17)$$

where $\widetilde{\mathcal{S}}'$ is the abstract state in the next frame. The optimal value function with respect to the abstract state can be represented as follows,

$$W(\widetilde{\mathcal{S}}) = \mathbb{E}_{\mathcal{Y}}W(\mathcal{S}) = \min_{\Omega}\mathbb{E}_{\mathcal{Y},\mathcal{L}^{\mathrm{B}}}^{\Omega}$$
$$\lim_{T\to\infty}\sum_{t=1}^{T}\gamma^{t-1}\left[g\big(\widetilde{\mathcal{S}}_t,\Omega(\widetilde{\mathcal{S}}_t)\big)\big|\widetilde{\mathcal{S}}_1=\widetilde{\mathcal{S}}\right], \forall\widetilde{\mathcal{S}}. \quad (18)$$

With the value function of the optimal policy for an arbitrary abstract state $\{W(\widetilde{\mathcal{S}})|\forall\widetilde{\mathcal{S}}\}$ and the current system state $\mathcal{S}_t$, the optimal action in the t-th frame is given by

$$\Omega^{\star}(\mathcal{S}_t) = \underset{\Omega(\mathcal{S}_t)}{\arg\min}\left[g_t\big(\mathcal{S}_t,\Omega(\mathcal{S}_t)\big)\right.$$
$$\left.+\gamma\sum_{\widetilde{\mathcal{S}}_{t+1}}W(\widetilde{\mathcal{S}}_{t+1})\Pr[\widetilde{\mathcal{S}}_{t+1}|\mathcal{S}_t,\Omega(\mathcal{S}_t)]\right], \forall\mathcal{S}_t. \quad (19)$$

Although conventional approaches such as policy and value iteration can be used to find the optimal scheduling policy [19], they suffer from the *curse of dimensionality*: due to the huge system state space, the evaluation of $\{W(\widetilde{\mathcal{S}}_t)|\forall\widetilde{\mathcal{S}}_t\}$ is prohibitive. In the following section, we shall propose a low-complexity scheduling scheme to address this issue.

IV. LOW-COMPLEXITY SCHEDULING

In this section, a low-complexity scheduling scheme with analytical performance bound is proposed. Here is a sketch of the proposed scheme:

- **Section IV-A:** We first design a heuristic scheduling policy as the reference policy, whose value function (average discounted cost) can be analytically expressed with mobile blocker detection.
- **Section IV-B:** Approximating the optimal value function via the above derived value function, the scheduling action of each frame can be obtained by solving the optimization problem in (19).
- **Section IV-C:** We decouple the problem and then propose an alternative optimization algorithm to derive the suboptimal solution.
- **Section IV-D:** We analyze the performance bound and time complexity of the proposed suboptimal solution.

A. Decoupled Reference Policy

A heuristic reference policy is proposed to provide an expression for an achievable average discounted cost. In fact, given the policy, the system evolves as a Markov chain, whose cost can be derived via the transition matrix. However, the dimension of the transition matrix grows exponentially with respect to the number of sensors, which makes the above approach infeasible. In order to address this issue, in the reference policy, the transmission time allocation of the K sensors is fixed. Hence, the state transition of the sensors can be decoupled into K Markov chains with significantly smaller state space. We adopt the following decoupled reference policy (which we refer to as reference policy); however, the design of reference policy is not unique.

Policy 1 (Decoupled Reference Policy Π). *The reference policy, denoted as $\Pi\triangleq(s_{t,k}^{\Pi}, \tau_{t,k}^{\Pi}, p_{t,k}^{\Pi})$, is elaborated below.*
- *The sampling decision is made when the uplink queue is empty, i.e., $s_{t,k}^{\Pi}=\mathbb{I}[Q_{t,k}=0]$.*
- *The transmission time allocation is proportional to the data volume of corresponding sampling, i.e., $\tau_{t,k}^{\Pi} = T_{\mathrm{F}}L_k/\sum_{k'\in\mathcal{K}}L_{k'}$.*
- *The transmission power is constant, i.e., $p_{t,k}^{\Pi}=P^{\Pi}$.*

To derive the average discounted cost for each sensor, we first express the transition matrices of the local abstract state of the sensors. Specifically, let $\mathbf{s}_{t,k} \in \mathcal{R}^{|\mathbb{L}|(L_k+1)A_{\max}^2\times1}$ be the vector representing the probabilities of all local abstract states of the k-th sensor, where the $\kappa(l_t^{\mathrm{B}},\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}}))$-th entry denotes the probability of the local abstract state $\widetilde{\mathcal{S}}_{t,k} = (l_t^{\mathrm{B}}, Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}})$, and $\kappa(l_t^{\mathrm{B}},\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}})) \triangleq (l_t^{\mathrm{B}}-1)(L_k+1)A_{\max}^2+\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}})$ and $\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}})\triangleq Q_{t,k}A_{\max}^2 + (A_{t,k}^{\mathrm{s}}-1)A_{\max}+A_{t,k}^{\mathrm{d}}$ are indexes. Let $\mathbf{P}_k \in \mathbb{R}^{|\mathbb{L}|(L_k+1)A_{\max}^2\times|\mathbb{L}|(L_k+1)A_{\max}^2}$ be the transition probability matrix of the local abstract state of the k-th sensor, we have $\mathbf{s}_{t+1,k} = \mathbf{P}_k^{\mathsf{T}}\mathbf{s}_{t,k}$. In order to derive the expression of $\mathbf{P}_k$, we first introduce the following lemma on the distribution of departure packet number.

Lemma 1. *With sufficiently large N_{R} and N_{T}, given the reference policy Π, the precoder and combiner in (7), and the blocker's location l_t^{B}, the probability mass function (PMF) of departure packet number can be written by (20), where $i^{\star}=\arg\max_i B_{t,k,i}(l_t^{\mathrm{B}})\rho_{k,i}^{-1}$.*

Proof. Please refer to Appendix A. $\square$

As a result, we have the following lemma on the transition matrix $\mathbf{P}_k$.

Lemma 2. *With sufficiently large N_{R} and N_{T}, given the reference policy Π, the precoder and combiner in (7), and the blocker's location l_t^{B}, the transition probability matrix of local abstract state of the k-th sensor is given by (21), where $\mathbf{M}_k^{(\ell)} \in \mathbb{R}^{(L_k+1)A_{\max}^2\times(L_k+1)A_{\max}^2}$ is given by Table II, and $\mathbf{P}^{\mathrm{B}}$ is defined in (1).*

$$\Pr\left[D_{t,k}^{\Pi}=d\big|l_t^{\mathrm{B}}\right]=\exp\left(-\frac{\rho_{k,i^\star}N_0W}{P^{\Pi}}\left(2^{\frac{dN_{\mathrm{b}}}{W\tau_{t,k}^{\Pi}}}-1\right)\right)-\exp\left(-\frac{\rho_{k,i^\star}N_0W}{P^{\Pi}}\left(2^{\frac{(d+1)N_{\mathrm{b}}}{W\tau_{t,k}^{\Pi}}}-1\right)\right). \tag{20}$$

$$\mathbf{P}_k=\begin{pmatrix}[\mathbf{P}^{\mathrm{B}}]_{1,1}\mathbf{M}_k^{(1)} & [\mathbf{P}^{\mathrm{B}}]_{1,2}\mathbf{M}_k^{(1)} & \cdots & [\mathbf{P}^{\mathrm{B}}]_{1,|\mathbb{L}|}\mathbf{M}_k^{(1)}\\[2pt] [\mathbf{P}^{\mathrm{B}}]_{2,1}\mathbf{M}_k^{(2)} & [\mathbf{P}^{\mathrm{B}}]_{2,2}\mathbf{M}_k^{(2)} & \cdots & [\mathbf{P}^{\mathrm{B}}]_{2,|\mathbb{L}|}\mathbf{M}_k^{(2)}\\[2pt] \vdots & \vdots & \ddots & \vdots\\[2pt] [\mathbf{P}^{\mathrm{B}}]_{|\mathbb{L}|,1}\mathbf{M}_k^{(|\mathbb{L}|)} & [\mathbf{P}^{\mathrm{B}}]_{|\mathbb{L}|,2}\mathbf{M}_k^{(|\mathbb{L}|)} & \cdots & [\mathbf{P}^{\mathrm{B}}]_{|\mathbb{L}|,|\mathbb{L}|}\mathbf{M}_k^{(|\mathbb{L}|)}\end{pmatrix}, \tag{21}$$

$Q_{t,k}$	$A_{t,k}^{\mathrm{s}}$	$A_{t,k}^{\mathrm{d}}$	$Q_{t+1,k}$	$A_{t+1,k}^{\mathrm{s}}$	$A_{t+1,k}^{\mathrm{d}}$	$\left[\mathbf{M}_k^{(\ell)}\right]_{\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}}),\,\epsilon(Q_{t+1,k},A_{t+1,k}^{\mathrm{s}},A_{t+1,k}^{\mathrm{d}})}$
0	$1,\dots,A_{\max}$	$1,\dots,A_{\max}$	0	1	1	$\Pr\left[D_{t,k}^{\Pi}\geq L_k\big|l_t^{\mathrm{B}}=\ell\right]$
0	$1,\dots,A_{\max}$	$1,\dots,A_{\max}$	$1,\dots,L_k$	1	$\min\{A_{t,k}^{\mathrm{d}}+1,A_{\max}\}$	$\Pr\left[D_{t,k}^{\Pi}=L_k-Q_{t+1,k}\big|l_t^{\mathrm{B}}=\ell\right]$
$1,\dots,L_k$	$1,\dots,A_{\max}$	$1,\dots,A_{\max}$	0	$\min\{A_{t,k}^{\mathrm{s}}+1,A_{\max}\}$	$\min\{A_{t,k}^{\mathrm{d}}+1,A_{\max}\}$	$\Pr\left[D_{t,k}^{\Pi}\geq Q_{t,k}\big|l_t^{\mathrm{B}}=\ell\right]$
$1,\dots,L_k$	$1,\dots,A_{\max}$	$1,\dots,A_{\max}$	$1,\dots,Q_{t,k}$	$\min\{A_{t,k}^{\mathrm{s}}+1,A_{\max}\}$	$\min\{A_{t,k}^{\mathrm{d}}+1,A_{\max}\}$	$\Pr\left[D_{t,k}^{\Pi}=Q_{t,k}-Q_{t+1,k}\big|l_t^{\mathrm{B}}=\ell\right]$

TABLE II. Non-zero entries of matrix $\mathbf{M}_k^{(\ell)}$.

Proof. Please refer to Appendix B. □

Finally, the value function of the reference policy, namely, the approximate value function, is given by the following theorem.

Theorem 1 (Value Function of Reference Policy Π)**.** *With the reference policy Π, the value function is given by*

$$W^{\Pi}(\widetilde{\mathcal{S}}_t)=\sum_{k\in\mathcal{K}}\underbrace{(\mathbf{e}_{\kappa(l_t^{\mathrm{B}},\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}}))}^{|\mathbb{L}|(L_k+1)A_{\max}^2})^{\mathsf{T}}[\mathbf{I}-\gamma\mathbf{P}_k]^{-1}\mathbf{g}_k}_{\triangleq W_k^{\Pi}(\widetilde{\mathcal{S}}_{t,k})}$$
$$+\frac{1}{1-\gamma}w_{\mathrm{P}}T_{\mathrm{F}}P^{\Pi}, \tag{22}$$

where $\mathbf{e}_n^N$ denotes an $N\times 1$ column vector whose n-th element is 1 and otherwise 0, and $\mathbf{g}_k\in\mathbb{R}^{|\mathbb{L}|(L_k+1)A_{\max}^2\times 1}$ is the cost vector for all abstract local state at the k-th sensor. Specifically, $[\mathbf{g}_k]_{\kappa(l_t^{\mathrm{B}},\epsilon(Q_{t,k},A_{t,k}^{\mathrm{s}},A_{t,k}^{\mathrm{d}}))}=A_{t,k}^{\mathrm{d}}+\mathbb{I}[Q_{t,k}=0]w_{\mathrm{P}}C^{\mathrm{s}}+\mathbb{I}[A_{t,k}^{\mathrm{d}}=A_{\max}]w_{\mathrm{Q}}$.

Proof. Please refer to Appendix C. □

B. Scheduling with Approximate Value Function

Substituting the optimal value function $W(\widetilde{\mathcal{S}}_{t+1})$ of the problem in (19) with the approximate value function $W^{\Pi}(\widetilde{\mathcal{S}}_{t+1})$, the proposed scheduling policy in one frame (say the t-th frame) given the global system state $\mathcal{S}_t$ can be obtained from the following optimization problem.

$$\mathsf{P2}:\ \left\{(s_{t,k}^{\star},\tau_{t,k}^{\star},p_{t,k}^{\star})\right\}_{k\in\mathcal{K}}=$$
$$\underset{(s_{t,k},\tau_{t,k},p_{t,k})_{k\in\mathcal{K}}}{\arg\min}\sum_{k\in\mathcal{K}}w_{\mathrm{P}}(\tau_{t,k}p_{t,k}+s_{t,k}C^{\mathrm{s}})$$
$$+\gamma\sum_{\widetilde{\mathcal{S}}_{t+1}}W^{\Pi}(\widetilde{\mathcal{S}}_{t+1})\Pr\left[\widetilde{\mathcal{S}}_{t+1}\Big|\mathcal{S}_t,(s_{t,k},\tau_{t,k},p_{t,k})_{k\in\mathcal{K}}\right],$$
$$\text{s.t.}\quad\text{Constraints in (4), (8), (9).}$$

Remark 1 (Predictive Scheduling)**.** *Note that the LoS path is usually much better than the NLoS paths. If the LoS path of one sensor (say the k-th sensor) is likely blocked soon,*

the penalty of AoI outdatedness will start to incur. Thus, its local value function $W_k^{\Pi}(\widetilde{\mathcal{S}}_{t,k})$ could be large for large values of AoI $A_{t,k}^{\mathrm{d}}$. Hence, the problem $\mathsf{P2}$ tends to schedule more transmission resources to this sensor, so that its AoI can be maintained at a low level.

C. Alternative Optimization Algorithm

Problem $\mathsf{P2}$ is a mixed continuous and discrete optimization problem with coupled variables $\{s_{t,k}\}_{k\in\mathcal{K}}$, $\{\tau_{t,k}\}_{k\in\mathcal{K}}$ and $\{p_{t,k}\}_{k\in\mathcal{K}}$. An alternative optimization algorithm is proposed in Algorithm 1 to obtain the suboptimal solution.

Algorithm 1: Alternative optimization algorithm.

Input :
$\mathcal{S}_t$: System state in the t-th frame
$\{W_k^{\Pi}(\widetilde{\mathcal{S}}_{t,k})\}_{\forall k,\widetilde{\mathcal{S}}_{t,k}}$: Approximate local value functions via Theorem 1
Output:
$\Psi^{\infty}(\mathcal{S}_t)$: Converged action in the t-th frame

1 $n\leftarrow 0,s_{t,k}^{(0)}\leftarrow s_{t,k}^{\Pi},\tau_{t,k}^{(0)}\leftarrow\tau_{t,k}^{\Pi},d_{t,k}^{(0)}\leftarrow d_{t,k}^{\Pi}$
2 **while** *not converge* **do**
3    $n\leftarrow n+1$
4   **for** $k\in\mathcal{K}$ **do in parallel**
5      $\lfloor$ *Solve $s_{t,k}^{(n)}$ in* $\mathsf{P2.1}(n,k)$ *via comparison.*
6   **for** $k\in\mathcal{K}$ **do in parallel**
7      $\lfloor$ *Solve $d_{t,k}^{(n)}$ in* $\mathsf{P2.2}(n,k)$ *via exhaustive search.*
8   *Solve $\{\tau_{t,k}^{(n)}\}_{k\in\mathcal{K}}$ in* $\mathsf{P2.3}(n)$ *via Lemma 3.*
9    $p_{t,k}^{(n)}=\frac{1}{Y_{t,k}}\left[2^{\wedge}\left(\frac{d_{t,k}^{(n)}N_{\mathrm{b}}}{\tau_{t,k}^{(n)}W}\right)-1\right],\forall k$
10 **return** $\Psi^{\infty}(\mathcal{S}_t)\leftarrow\left\{(s_{t,k}^{\infty},\tau_{t,k}^{\infty},p_{t,k}^{\infty})\right\}_{k\in\mathcal{K}}$

Instead of optimizing $\{s_{t,k}\}_{k\in\mathcal{K}}$, $\{\tau_{t,k}\}_{k\in\mathcal{K}}$ and $\{p_{t,k}\}_{k\in\mathcal{K}}$, we optimize $\{s_{t,k}\}_{k\in\mathcal{K}}$, $\{\tau_{t,k}\}_{k\in\mathcal{K}}$ and $\{d_{t,k}\}_{k\in\mathcal{K}}$ alternatively, where $d_{t,k}$ denotes the number of transmission packets from the k-th sensor in the t-th frame. Let $\{s_{t,k}^{(n)}\}_{k\in\mathcal{K}}$,

$$\text{P2.1}(n,k): \ s_{t,k}^{(n)} = \underset{s_{t,k}\in\{0,1\}}{\arg\min} \ w_{\mathrm{P}}s_{t,k}C^{\mathrm{s}} + \gamma\sum_{\widetilde{\mathcal{S}}_{t+1,k}} W_k^{\Pi}(\widetilde{\mathcal{S}}_{t+1,k})\Pr\left[\widetilde{\mathcal{S}}_{t+1,k}\Big|\mathcal{S}_{t,k},(s_{t,k},\tau_{t,k}^{(n-1)},p_{t,k}^{(n-1)})\right] \tag{23}$$

$$\text{P2.2}(n,k): \ d_{t,k}^{(n)} = \underset{d_{t,k}:\ p_{t,k}(d_{t,k})\le P_{\max}}{\arg\min} \ w_{\mathrm{P}}\tau_{t,k}^{(n-1)}p_{t,k}(d_{t,k}) + \gamma\sum_{\widetilde{\mathcal{S}}_{t+1,k}} W_k^{\Pi}(\widetilde{\mathcal{S}}_{t+1,k})\Pr\left[\widetilde{\mathcal{S}}_{t+1,k}\Big|\mathcal{S}_{t,k},\left(s_{t,k}^{(n)},\tau_{t,k}^{(n-1)},p_{t,k}(d_{t,k})\right)\right] \tag{24}$$

$$\text{P2.3}(n): \ \left\{\tau_{t,k}^{(n)}\right\}_{k\in\mathcal{K}} = \underset{\{\tau_{t,k}\}_{k\in\mathcal{K}}}{\arg\min} \sum_{k\in\mathcal{K}}\tau_{t,k}p_{t,k}(\tau_{t,k}) \quad \text{s.t.} \ \sum_{k\in\mathcal{K}}\tau_{t,k}=T_{\mathrm{F}}, \ 0\le\tau_{t,k}\le T_{\mathrm{F}}, \ p_{t,k}(\tau_{t,k})\le P_{\max}, \ \forall k\in\mathcal{K}, \tag{25}$$

$\{\tau_{t,k}^{(n)}\}_{k\in\mathcal{K}}$ and $\{d_{t,k}^{(n)}\}_{k\in\mathcal{K}}$ be the correspondingly optimized variables in the n-th iteration, respectively. Initializing the actions with the reference policy Π (Line 1), i.e., $(s_{t,k}^{(0)},\tau_{t,k}^{(0)},d_{t,k}^{(0)}) = (s_{t,k}^{\Pi},\tau_{t,k}^{\Pi},d_{t,k}^{\Pi})$, the entire procedure of solving P2 consists of a number of iterations, and the n-th iteration includes the sub-problems $\{\text{P2.1}(n,k)|\forall k\}$, $\{\text{P2.2}(n,k)|\forall k\}$, and $\text{P2.3}(n)$ in (23), (24) and (25), respectively (Line 4–8), where

$$p_{t,k}(d_{t,k}) = \frac{1}{Y_{t,k}}\left[2^{\frac{d_{t,k}N_{\mathrm{b}}}{\tau_{t,k}^{(n-1)}W}}-1\right], \ p_{t,k}(\tau_{t,k}) = \frac{1}{Y_{t,k}}\left[2^{\frac{d_{t,k}^{(n)}N_{\mathrm{b}}}{\tau_{t,k}W}}-1\right].$$

Notice that the originally coupled sampling decision $s_{t,k}^{(n)}$ and transmission packet number allocation $d_{t,k}^{(n)}$ of all sensors are decoupled given the transmission time allocation $\{\tau_{t,k}^{(n)}\}_{k\in\mathcal{K}}$, thus the complexity is significantly reduced. The optimal solution of $\{\text{P2.1}(n,k)|\forall k\}$ can be derived by evaluating the binary local sampling action for $s_{t,k}=0$ and $s_{t,k}=1$ and choosing the one with the smaller value of the objective function (Line 5). Similarly, the optimal solution of $\{\text{P2.2}(n,k)|\forall k\}$ can be solved by one-dimensional search over $d_{t,k}\in\{0,1,\ldots,Q_{t,k}\}$ (Line 7).

Moreover, $\text{P2.3}(n)$ is a convex optimization problem and the optimal solution can be derived by the following lemma (Line 8).

Lemma 3. *The optimal solution of* $\text{P2.3}(n)$ *is given by*

$$\tau_{t,k}^{(n)} = \frac{d_{t,k}N_{\mathrm{b}}}{W}\max\left\{\frac{\ln 2}{1+\mathbb{W}_0\left(\frac{Y_{t,k}\nu^\star-1}{e}\right)},\frac{1}{\log_2(1+P_{\max}Y_{t,k})}\right\},$$

where $\mathbb{W}_0(\cdot)$ *denotes the principal branch of the Lambert W function, and* $\nu^\star$ *denotes the optimal Lagrangian multiplier for equality constraint* (8) *which can be solved by*

$$\sum_{k\in\mathcal{K}}\max\left\{\frac{d_{t,k}N_{\mathrm{b}}\ln 2}{W\left[1+\mathbb{W}_0\left(\frac{Y_{t,k}\nu^\star-1}{e}\right)\right]},\right.$$
$$\left.\frac{d_{t,k}N_{\mathrm{b}}}{W\log_2(1+P_{\max}Y_{t,k})}\right\}-T_{\mathrm{F}}=0. \tag{26}$$

Proof. Please refer to the Theorem 1 in [21] for the proof. $\square$

Because the left-hand-side (LHS) of (26) is non-increasing, the bisection method can be used to solve $\nu^\star$. Given $d_{t,k}^{(n)}$ and $\tau_{t,k}^{(n)}$, the transmission power allocation in the n-th iteration, denoted by $p_{t,k}^{(n)}$, can be derived analytically (Line 9).

Since the iterative optimization of the sub-problems $\{\text{P2.1}(n,k)|\forall k\}$, $\{\text{P2.2}(n,k)|\forall k\}$, and $\text{P2.3}(n)$ would lead

to a smaller or equal value of the objective function in P2, the above iteration will always converge (Line 10).

D. Performance Bound and Time Complexity

Denote $\Psi^{(n)}: \widetilde{\mathcal{S}}_t \to \{s_{t,k}^{(n)},\tau_{t,k}^{(n)},p_{t,k}^{(n)}\}_{k\in\mathcal{K}}$ as the scheduling policy obtained after n iterations, $\widetilde{W}^{\Psi^{(n)}}(\widetilde{\mathcal{S}}_t)$ as the corresponding value functions, and $\Psi^\infty: \widetilde{\mathcal{S}}_t \to \{s_{t,k}^\infty,\tau_{t,k}^\infty,p_{t,k}^\infty\}_{k\in\mathcal{K}}$ as the scheduling policy after convergence. The performance of Ψ^∞ and $\Psi^{(n)}$ can be guaranteed by the following lemma.

Lemma 4 (Performance Bound). *The average discounted cost of policies* $\Psi^{(n)}$ *and* Ψ^∞ *can be bounded by*

$$W(\widetilde{\mathcal{S}}_t)\le W^{\Psi^\infty}(\widetilde{\mathcal{S}}_t)\le\ldots\le\widetilde{W}^{\Psi^{(n)}}(\widetilde{\mathcal{S}}_t)\le\ldots\le\widetilde{W}^{\Psi^{(1)}}(\widetilde{\mathcal{S}}_t)\le W^{\Pi}(\widetilde{\mathcal{S}}_t).$$

Proof. Since $W(\widetilde{\mathcal{S}}_t)$ is the optimal value function, it is the lower bound of value function of an arbitrary policy. The proof of $W^{\Psi^\infty}(\widetilde{\mathcal{S}}_t)\le\ldots\le\widetilde{W}^{\Psi^{(n)}}(\widetilde{\mathcal{S}}_t)\le\ldots\le\widetilde{W}^{\Psi^{(1)}}(\widetilde{\mathcal{S}}_t)$ and $\widetilde{W}^{\Psi^{(1)}}(\widetilde{\mathcal{S}}_t)\le W^{\Pi}(\widetilde{\mathcal{S}}_t)$ resembles the proof of the policy improvement property in Chapter II of [19]. $\square$

For conventional value iteration method, the transmission time and power shall be first discretized into τ_{D} and p_{D} levels, respectively. Then the cardinalities of the abstract state space and action space can be denoted by $|\widetilde{\mathbb{S}}|\triangleq|\mathbb{L}|A_{\max}^{2K}\prod_{k\in\mathcal{K}}L_k$ and $|\mathbb{A}|\triangleq(2\tau_{\mathrm{D}}p_{\mathrm{D}})^K$, respectively. The time complexity of the conventional value iteration is $O(|\widetilde{\mathbb{S}}|^2|\mathbb{A}|)$ for each iteration. On the other hand, the proposed scheme consists of two stages, namely the evaluation of the approximate value function and the alternative optimization algorithm. The time complexity of the approximate value function is $O(|\widetilde{\mathbb{S}}|[|\mathbb{L}|A_{\max}^2\sum_{k\in\mathcal{K}}L_k]^3)$ and the time complexity of the alternative optimization algorithm is $O(|\widetilde{\mathbb{S}}|\sum_{k\in\mathcal{K}}L_k)$ for each iteration. Benefiting from the analytical expressed approximate value function, a number of iterations from the initial value function to the value function of a roughly-good policy is prevented, and thus the time complexity is essentially reduced.

V. SIMULATIONS AND DISCUSSIONS

In this section, the performance of the proposed algorithm is demonstrated via simulations, where a number of benchmarks are used in the comparison. We summarize the key findings of our simulations as follows:

- Our proposed scheme can converge after only a few iterations and reduce the average per-frame cost by 13.5%–49.6% compared with the benchmarks.
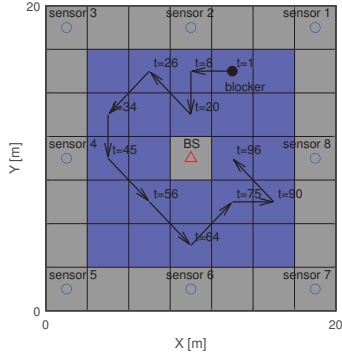
Fig. 2. Simulation scenario and an illustrative trajectory of the blocker.

- Our proposed scheme can proactively keep the AoIs of sensors with future channel degradation at a low level in example traces, which intuitively verifies the benefits of exploiting mobile blocker detection in AoI-oriented scheduling design.

- Our proposed scheme shows the better performance compared to the benchmarks with robustness against the number of sensors.

As illustrated in Fig. 2, we consider a 20m×20m square room with walls serving as the reflectors of the NLoS paths ($M = 4$), where the BS is deployed at the center block of the room and the locations of sensors are uniformly distributed near the walls. The mobility of the human blocker with radius $r_B = 0.3$m follows a modified random walk. The probability of staying in the same grid in next frame is 0.90, while the probabilities of moving to one of the feasible neighboring blocks in the blue region are equal. An illustrative blocker trajectory is shown by arrows in Fig. 2. Other parameters are summarized in Table III. We evaluate the proposed algorithms with different numbers of iterations in solving P2, namely, $\Psi^{(1)}$, $\Psi^{(2)}$, $\Psi^{(3)}$, Ψ^∞, and compare them with the following three benchmark policies, which are referred to as BM1, BM2 and BM3, respectively. All the benchmarks adopt the same sampling policy and transmission power allocation as the reference policy, i.e., $s_{t,k} = \mathbb{I}[Q_{t,k} = 0]$ and $p_{t,k} = P^\Pi$, while the transmission time allocation is elaborated as follows.

BM 1 (Reference Policy). *The transmission time allocation of each sensor is proportional to the corresponding data volume of sample at each sensor, i.e., $\tau_{t,k} = T_F L_k / \sum_{k' \in \mathcal{K}} L_{k'}$.*

BM 2 (Largest-AoI First). *The sensor with the largest AoI at the server, i.e., $\arg\max_k A_{t,k}^d$, is scheduled for transmission sequentially, until transmission time of the frame is used up.*

BM 3 (Dynamic Backpressure [23]). *The sensor with the largest product of buffer length and uplink capacity, i.e., $\arg\max_k Q_{t,k} R_{t,k}$, is scheduled for transmission sequentially, until transmission time of the frame is used up.*

BM1 corresponds to the reference policy we adopt in the proposed scheme with fair allocation based on uploading

workload. BM2 selects UE greedily to mitigate the outdatedness AoI penalty. BM3 adopts a UE selection strategy based on both queue length and uplink channel capacity.

Fig. 3(a) displays the cumulative distribution functions (CDFs) of per-frame cost of the proposed scheduling scheme as well as those of the three benchmarks, when the number of sensors is $K = 8$. The step-shape curves are caused by the outdated AoI penalty (whose weight is $w_Q = 100$ in the simulation). It can be observed that the proposed policy with $n = 1$, denoted as $\Psi^{(1)}$, has a significantly better curve than the reference policy (BM1). Moreover, the curves of $\Psi^{(3)}$ and Ψ^∞ are close, demonstrating that only a small number of iterations are needed in the optimization of P2. BM2 allocates all the transmission time to the sensors with the largest AoIs at the server, which leads to the highest probability of per-frame costs below 150 (one sensor with outdated AoI penalty in average). However, the largest-AoI-first scheduling in BM2 may be stuck to the sensor whose LoS path is blocked, leading to more outdated AoI penalty. It can be observed that compared with BM3, BM2 has much lower probability when the per-frame cost is below 250. The average per-frame costs are 129.3, 256.3, 203.5 and 149.5 for the proposed scheme Ψ^∞ and the benchmarks respectively, indicating that the average performance of the proposed scheme (cost reduction by 49.6%, 36.5% and 13.5%, respectively) is the best.

Fig. 3(b) sketches the bar chart of the four components of per-frame cost. BM1 has the highest outdated AoI penalty, because its scheduling is independent of the system state. BM2 has the lowest AoI penalty, but consumes significantly higher transmission energy than the proposed scheme. BM3 has the highest sampling energy consumption and thus highest sampling frequency, because it tends to select the sensor with high throughput. Hence, the proposed scheme achieves the best balance between AoI and energy consumption.

The insights on blockage-predictive scheduling of the proposed policy can be obtained in Fig. 4, which shows the

Parameter	Symbol	Value
Carrier frequency	f_c	60 GHz
# of sensors	K	4,5,6,7,8
# of receive/transmit antenna elements	N_R, N_T	64, 128
Path loss	$\rho_{k,i}$	LoS: $32.5 + 20\log(f_c)$ $+ 20\log(R)$ [22] NLoS: $32.5 + 20\log(f_c)$ $+ 20\log(R) + 15$
# of possible locations of the blocker	$\mathbb{L}$	24
Bandwidth	W	400 MHz
Frame duration	T_F	10 ms
Packet size	N_b	200 KB
Noise power spectral density	N_0	-174 dBm/Hz
Data volume	L_k	$\mathcal{U}(3,5)$
Threshold for outdated AoI	A_{max}	10
Maximum transmission power	P_{max}	100 mW
Discount factor	γ	0.98
Weights for energy consumption and outdated AoI penalty	w_P, w_Q	10000, 100
Sampling energy consumption	C^s	10^{-4} J
Transmission power of reference policy	P^Π	50 mW

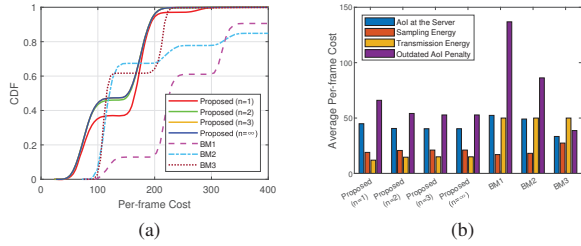TABLE III. Configuration parameters of the simulation.

Fig. 3. (a) CDF of the per-frame cost. (b) Average performance of the AoI at the server, sampling energy, transmission energy and outdated AoI penalty.

dynamics of the AoI at the 4-th sensor and the corresponding AoI at the server. The trajectory of the blocker for this trace is illustrated in Fig. 2. The LoS path between the BS and the 4-th sensor is blocked since the 45-th frame. Compared with BM1, the proposed policy can detect future channel degradation, and keep the AoIs at the 4-th sensor and the BS at a low level before the 45-th frame, which reduces the time duration with outdated AoI at the server.
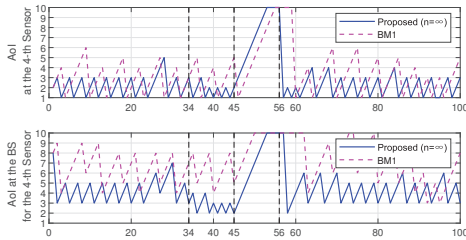

Fig. 4. AoI dynamics of the 4-th sensor.

The average per-frame cost versus the number of sensors is studied in Fig. 5. The average per-frame cost of the proposed scheme is always lower than the benchmarks, which verifies the better performance of the proposed scheme and its robustness against the number of sensors.
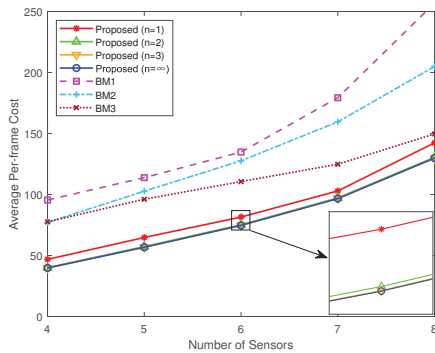

Fig. 5. Average per-frame cost versus the number of sensors.

VI. CONCLUSION

In this paper, we formulate the dynamic scheduling of sampling and uploading in an mmWave-based WSN with random human blockage as an infinite-horizon MDP with

discounted cost, where the weighted sum of average AoI and system energy consumption is the minimization objective. Since the system state space grows exponentially with respect to the number of sensors, an approximate MDP solution framework is proposed to address the curse of dimensionality. In the proposed framework, the optimal value function is approximated by an analytical expression derived from a reference policy. The numerical evaluation of the value function in conventional approximate MDP solutions can be eliminated. Finally, the policy iteration based on analytical value function is significantly more efficient than lookup-table-based value functions in conventional solution frameworks. Thus, the solution complexity is greatly reduced. Moreover, the impact of random human blockage on future cost is predicted in the approximate value function, and hence, mitigated in the scheduling of the current frame. In future work, we shall consider low-complexity scheduling for multi-blocker scenarios, where how to address the curse of dimensionality caused by large location state space is challenging.

APPENDIX A: PROOF OF LEMMA 1

According to [24], with sufficient large N_R and N_T, $f_R(\phi) \to 1$ and $f_T(\theta) \to 1$ if $\phi = 0$ and $\theta = 0$, otherwise $f_R(\phi) \to 0$ and $f_T(\theta) \to 0$, where $f_R(\phi) = |\mathbf{a}_R^H(0)\mathbf{a}_R(\phi)|$ and $f_T(\theta) = |\mathbf{a}_T^H(0)\mathbf{a}_T(\theta)|$.

Hence, substituting (2) into (7), we can derive

$$
\begin{aligned}
(\mathbf{w}_{t,k}, \mathbf{f}_{t,k}) &= \arg\max_{\substack{\mathbf{a}_R(\phi_q) \in \mathcal{W} \\ \mathbf{a}_T(\phi_p) \in \mathcal{F}}} \mathbb{E}_{(\alpha_{t,k,i}) \in \mathcal{M}} \Big| \sum_{i \in \mathcal{M}} B_{t,k,i}(l_t^B)\alpha_{t,k,i} \\
&\quad \times f_R(\phi_{k,i} - \phi_q) f_T(\theta_p - \theta_{k,i}) \Big|^2 \quad (27) \\
&= \arg\max_{\mathbf{a}_R(\phi_{k,i}), \mathbf{a}_T(\phi_{k,i})} B_{t,k,i}(l_t^B)\rho_{k,i}^{-1}. \quad (28)
\end{aligned}
$$

Let $i^\star = \arg\max_i B_{t,k,i}(l_t^B)\rho_{k,i}^{-1}$, (28) can be written by $(\mathbf{w}_{t,k}, \mathbf{f}_{t,k}) = (\mathbf{a}_R(\phi_{k,i^\star}), \mathbf{a}_T(\phi_{k,i^\star}))$. Then the baseband gain can be represented by $Y_{t,k}(l_t^B) = |\alpha_{t,k,i^\star}|^2/(N_0 W)$.

Given $\alpha_{t,k,i^\star} \sim \mathcal{CN}(0, \rho_{k,i^\star}^{-1})$, $|\alpha_{t,k,i^\star}|^2 = (\Re[\alpha_{t,k,i^\star}])^2 + (\Im[\alpha_{t,k,i^\star}])^2$ follows an exponential distribution, i.e., $|\alpha_{t,k,i^\star}|^2 \sim \text{Exp}(\rho_{k,i^\star})$. Then $Y_{t,k}(l_t^B) \sim \text{Exp}(\rho_{k,i^\star}N_0 W)$ and thus the CDF of $Y_{t,k}(l_t^B)$ can be written by

$$
\Pr[Y_{t,k}(l_t^B) \leq x] = 1 - \exp(-\rho_{k,i^\star}N_0 W x). \quad (29)
$$

Therefore, the PMF of departure packet number is

$$
\begin{aligned}
\Pr\left[D_{t,k}^\Pi = d \big| l_t^B\right] &= \Pr\left[D_{t,k}^\Pi \leq d+1 \big| l_t^B\right] - \Pr\left[D_{t,k}^\Pi \leq d \big| l_t^B\right] \\
&= \Pr\left[Y_{t,k}(l_t^B) \leq \frac{2^{\frac{(d+1)N_b}{W\tau_{t,k}^\Pi}} - 1}{P^\Pi}\right] - \Pr\left[Y_{t,k}(l_t^B) \leq \frac{2^{\frac{dN_b}{W\tau_{t,k}^\Pi}} - 1}{P^\Pi}\right].
\end{aligned}
$$

APPENDIX B: PROOF OF LEMMA 2

In fact, $\mathbf{M}_k^{(\ell)}$ represents the transition probability matrix of local abstract state eliminating l_t^B, i.e., $(Q_{t,k}, A_{t,k}^s, A_{t,k}^d)$, conditioned on $l_{t+1}^B = \ell$. Then the derivation of (21) is straightforward. We have the following discussion on all possible cases for $\mathbf{M}_k^{(\ell)}$ defined in Table II.

- **Case 1** ($Q_{t,k} = 0, 1 \leq A_{t,k}^s \leq A_{\max}, 1 \leq A_{t,k}^d \leq A_{\max}, Q_{t+1,k} = 0, A_{t+1,k}^s = 1, A_{t+1,k}^d = 1$)**:** This means that sampling and trans-

mission of all packets of the new sample are accomplished with the probability of L_k departure packets.

- **Case 2** ($Q_{t,k}=0, 1 \leq A_{t,k}^{\mathrm{s}} \leq A_{\max}, 1 \leq A_{t,k}^{\mathrm{d}} \leq A_{\max}, 1 \leq Q_{t+1,k} \leq L_k, A_{t+1,k}^{\mathrm{s}}=1, A_{t+1,k}^{\mathrm{d}}=\min\{A_{t,k}^{\mathrm{d}}+1, A_{\max}\}$)**:** This means that sampling and transmission of $(L_k - Q_{t+1,k})$ packets of the new sample are accomplished with the probability of $(L_k - Q_{t+1,k})$ departure packets.

- **Case 3** ($1 \leq Q_{t,k} \leq L_k, 1 \leq A_{t,k}^{\mathrm{s}} \leq A_{\max}, 1 \leq A_{t,k}^{\mathrm{d}} \leq A_{\max}, Q_{t+1,k} =0, A_{t+1,k}^{\mathrm{s}}=\min\{A_{t,k}^{\mathrm{s}}+1, A_{\max}\}, A_{t+1,k}^{\mathrm{d}}=\min\{A_{t,k}^{\mathrm{s}}+1, A_{\max}\}$)**:** This means that transmission of all remaining $Q_{t+1,k}$ packets of the current sample is accomplished with the probability of $Q_{t+1,k}$ departure packets.

- **Case 4** ($1 \leq Q_{t,k} \leq L_k, 1 \leq A_{t,k}^{\mathrm{s}} \leq A_{\max}, 1 \leq A_{t,k}^{\mathrm{d}} \leq A_{\max}, 1 \leq Q_{t+1,k} \leq Q_{t,k}, A_{t+1,k}^{\mathrm{s}}=\min\{A_{t,k}^{\mathrm{s}}+1, A_{\max}\}, A_{t+1,k}^{\mathrm{d}}=\min \{A_{t,k}^{\mathrm{d}}+1, A_{\max}\}$)**:** This means that transmission of $(Q_{t,k} - Q_{t+1,k})$ packets of the current sample is accomplished with the probability of $(Q_{t,k} - Q_{t+1,k})$ departure packets.

APPENDIX C: PROOF OF THEOREM 1

The approximate value function is given by

$$W^{\Pi}(\widetilde{\mathcal{S}}_t) = \lim_{T \to \infty} \sum_{t=1}^{T} \gamma^{t-1} \Big[\sum_{k \in \mathcal{K}} \big[\mathbf{e}_{\kappa(l_t^{\mathrm{B}}, \epsilon(Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}}))}^{|\mathbb{L}|(L_k+1)A_{\max}^2} \big]^{\mathsf{T}} \mathbf{P}_k^{t-1} \mathbf{g}_k$$
$$+ w_{\mathrm{P}} T_{\mathrm{F}} P^{\Pi} \Big]. \tag{30}$$

Since the reference policy adopts constant transmission time and power allocations, the per-frame cost for transmission power consumption, and thus its discounted cumulative sum corresponding to the second term in (22), results into a constant. In fact, $\mathbf{e}_{\kappa(l_t^{\mathrm{B}}, \epsilon(Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}}))}^{|\mathbb{L}|(L_k+1)A_{\max}^2}$ represents the probability vector for a deterministic local abstract state $\widetilde{\mathcal{S}}_{t,k} = (l_t^{\mathrm{B}}, Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}})$, $\mathbf{P}_k$ represents the transition probability matrix, and the $\kappa(l_t^{\mathrm{B}}, \epsilon(Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}}))$-th entry of $\mathbf{g}_k$ represents the per-frame cost for sampling energy cost, and AoIs and outdated AoI penalties at the server. Since the reference policy samples only when the queue becomes empty, the per-frame cost for sampling is counted only for $Q_{t,k}=0$. Therefore, $\big[\mathbf{e}_{\kappa(l_t^{\mathrm{B}}, \epsilon(Q_{t,k}, A_{t,k}^{\mathrm{s}}, A_{t,k}^{\mathrm{d}}))}^{|\mathbb{L}|(L_k+1)A_{\max}^2} \big]^{\mathsf{T}} \mathbf{P}_k^{t-1} \mathbf{g}_k$ represents the expected per-frame cost in the t-th frame. The derivation from (30) to (22) resembles the proof in Appendix B(3) of [25].

REFERENCES

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[2] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2011, pp. 350–358.

[3] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 193–204, 2018.

[4] M. Sheikhi and V. Hakami, "AoI-aware status update control for an energy harvesting source over an uplink mmWave channel," in *2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2021, pp. 01–06.

[5] B. Zhou and W. Saad, "Joint status sampling and updating for minimizing age of information in the internet of things," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7468–7482, 2019.

[6] ——, "Minimum age of information in the internet of things with non-uniform status packet sizes," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1933–1947, 2020.

[7] X. Zheng, S. Zhou, Z. Jiang, and Z. Niu, "Closed-form analysis of non-linear age of information in status updates with an energy harvesting transmitter," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4129–4142, 2019.

[8] X. He, S. Wang, X. Wang, S. Xu, and J. Ren, "Age-based scheduling for monitoring and control applications in mobile edge computing systems," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 1009–1018.

[9] B. Zhou and W. Saad, "Optimal sampling and updating for minimizing age of information in the internet of things," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6.

[10] J. Palacios, P. Casari, and J. Widmer, "JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[11] T. Wei, A. Zhou, and X. Zhang, "Facilitating robust 60 GHz network deployment by sensing ambient reflectors," in *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, 2017, pp. 213–226.

[12] Y. Sun, J. Li, T. Zhang, R. Wang, X. Peng, X. Han, and H. Tan, "An indoor environment sensing and localization system via mmwave phased array," *Journal of Communications and Information Networks*, vol. 7, no. 4, pp. 383–393, 2022.

[13] Y. Sun, B. Lv, R. Wang, H. Tan, and F. C. M. Lau, "Predictive resource allocation in mmWave systems with rotation detection," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 2753–2759.

[14] G. Li, S. Wang, J. Li, R. Wang, X. Peng, and T. X. Han, "Wireless sensing with deep spectrogram network and primitive based autoregressive hybrid channel model," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021, pp. 481–485.

[15] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, 2022.

[16] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. R. Yu, "Device-free wireless sensing for human detection: The deep learning perspective," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2517–2539, 2021.

[17] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[18] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 204–219, 2019.

[19] D. Bertsekas, *Dynamic programming and optimal control*, 4th ed. Belmont, MA, USA: Athena scientific, 2012, vol. 2.

[20] L. Li, T. J. Walsh, and M. L. Littman, "Towards a unified theory of state abstraction for MDPs." in *AI&M*, 2006.

[21] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-efficient radio resource allocation for federated edge learning," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[22] A. Maltsev, V. Erceg, E. Perahia, C. Hansen, R. Maslennikov, A. Lomayev, A. Sevastyanov, A. Khoryaev, G. Morozov, M. Jacob *et al.*, "Channel models for 60 GHz WLAN systems, doc.: Ieee 802.11-09/0334r8," *IEEE 802.11 document 09/0334r8*, 2010.

[23] L. Georgiadis, M. J. Neely, L. Tassiulas *et al.*, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends® in Networking*, vol. 1, no. 1, pp. 1–144, 2006.

[24] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.

[25] S. Huang, B. Lv, R. Wang, and K. Huang, "Scheduling for mobile edge computing with random user arrivals—an approximate MDP and reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7735–7750, 2020.