

Age of Information-Oriented Probabilistic Link Scheduling for Device-to-Device Networks

Lixin Wang¹, Qian Wang², He (Henry) Chen³, Shidong Zhou¹

¹Department of Electronic Engineering, Tsinghua University, Beijing, China

²School of Communication Engineering, Hangzhou Dianzi University, Zhejiang, China

³Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China

Email: wanglx19@mails.tsinghua.edu.cn, qian.wang@hdu.edu.cn, he.chen@ie.cuhk.edu.hk, zhousd@tsinghua.edu.cn

Abstract—This paper focuses on optimizing the long-term average age of information (AoI) in device-to-device (D2D) networks through age-aware link scheduling. The problem is naturally formulated as a Markov decision process (MDP). However, finding the optimal policy for the formulated MDP in its original form is challenging due to the intertwined AoI dynamics of all D2D links. To address this, we propose an age-aware stationary randomized policy that determines the probability of scheduling each link in each time slot based on the AoI of all links and the statistical channel state information among all transceivers. By employing the Lyapunov optimization framework, our policy aims to minimize the Lyapunov drift in every time slot. Nonetheless, this per-slot minimization problem is nonconvex due to cross-link interference in D2D networks, posing significant challenges for real-time decision-making. After analyzing the permutation equivariance property of the optimal solutions to the per-slot problem, we apply a message passing neural network (MPNN), a type of graph neural network that also exhibits permutation equivariance, to optimize the per-slot problem in an unsupervised learning manner. Simulation results demonstrate the superior performance of the proposed age-aware stationary randomized policy over baselines and validate the scalability of our method.

Index Terms—Device-to-device networks, age of information, Lyapunov optimization, messaging passing neural network.

I. INTRODUCTION

With the rapid advancement of the Internet of Things (IoT), a multitude of time-sensitive applications, including autonomous driving, monitoring systems, and industrial control, have emerged. Compared with conventional metrics like delay, these applications prioritize the timeliness of information from the perspective of the receiving end, which can be effectively characterized by the novel metric known as age of information (AoI) [1]. On the other hand, device-to-device (D2D) communication, a pivotal technology for supporting these applications in IoT networks, can significantly increase spectral efficiency and network capacity by enabling devices to communicate directly with each other without routing through a base station [2].

The work of L. Wang and S. Zhou is supported in part by National Natural Science Foundation of China under Grants (62394294, 62394290) and the Fundamental Research Funds for the Central Universities under Grant 2242022k60006. The work of H. Chen is supported in part by RGC General Research Funds (GRF) under Project 14205020, the CUHK direct grant for research under Project 4055166.

However, the increasing number of mobile devices leads to a scarcity of frequency resources for D2D communication, making it essential to adopt frequency reuse mechanisms. This means that different D2D links may occupy the same frequency band to transmit simultaneously, leading to cross-link interference. Consequently, it is crucial to properly schedule D2D links and investigate the impact of potential interference between different D2D links on the AoI metric. Previous works have addressed the link scheduling problems in D2D networks using various metrics, such as sum-rate [3], throughput [4], and fairness [4], [5]. However, these metrics differ from AoI, which is inherently Markovian and often necessitates consideration of long-term average performance. This calls for the design of new link scheduling strategies with a focus on optimizing AoI in D2D networks.

Some existing works have put some effort into investigating AoI-oriented link scheduling problems [6]–[9]. However, they often fail to consider a practical physical interference model. Specifically, a portion of these works focus on the collision interference model, where only one device is scheduled to access the shared channel at a time, and multiple simultaneous accesses result in failure for all scheduled devices [6], [7]. Given the spatial distribution of different devices in D2D networks, such an oversimplified setup can miss many transmission opportunities, leading to increased AoI. The authors in [8], [9] considered abstracting the *feasible activation set* from a general interference model and focused on scheduling devices within that pre-established set. However, this feasible activation set is assumed to be fixed and may not align with the time-varying channel common in real-world scenarios.

This paper focuses on optimizing the long-term average AoI in D2D networks with practical Rayleigh fading channels. The problem presents two main challenges: 1) The optimization objective must be evaluated over the long term, meaning we cannot focus solely on current scheduling decisions but must also consider their future impact on AoI; 2) Scheduling one D2D link causes interference with other links. This cross-link interference is time-varying and must be carefully managed. For example, in a D2D network with two links, significant mutual interference usually prevents simultaneous scheduling. However, channel variations occasionally allow both links to be scheduled together. Effectively exploiting these changing

transmission opportunities to minimize AoI has not been thoroughly investigated. A common approach to addressing this issue is to model it as a Markov Decision Process (MDP) in order to derive an optimal decision policy. However, the high-dimensional nature of the considered problem renders standard value iteration methods impractical, making deep reinforcement learning (DRL) a viable alternative [10]. For instance, the authors in [11] formulated a Markov game and developed a multi-agent DRL algorithm to obtain a power control policy for minimizing the expected average AoI in a wireless ad-hoc network. Similarly, the authors in [12] proposed a graph neural network (GNN)-based DRL algorithm to minimize the average AoI by optimizing power control policy in D2D networks. Nonetheless, when a large number of devices are involved, DRL methods often suffer from sample inefficiency, requiring extensive interactions and imposing a heavy burden on real-world applications, not to mention the potential for slow convergence and poor performance.

Given the challenges associated with MDP-based approaches for the considered link scheduling problem, an alternative strategy to reduce the long-term average AoI is to narrow the search space of the policy. The authors in [13], [14] focused on a class of stationary randomized policies, where each device is scheduled based on a time-invariant probability. These policies can be efficiently obtained through convex optimization using statistical channel state information (CSI) and can be deployed in a distributed manner. However, such stationary randomized policies are *age-independent* because they do not leverage instantaneous system AoI, potentially sacrificing performance. Therefore, we are motivated to incorporate AoI into our scheduling policy. The rationale behind this approach can be intuitively illustrated with a simple 2-link scenario: suppose there are two links in a D2D network that cannot be scheduled simultaneously due to significant interference with each other, and scheduling either one causes the same level of interference on other links. In such a scenario, when only one of these two links can be scheduled, prioritizing the link with the larger AoI will result in a lower average AoI. Based on this insight, this paper aims to develop an *age-aware* stationary randomized policy, where each link is scheduled with a probability determined by the instantaneous system AoI and statistical CSI. To the best of our knowledge, this is the first work to develop an age-aware link scheduling policy for D2D networks using practical interference models.

The main contributions of this paper are two-fold: (1) By applying the Lyapunov optimization technique [15], we transform the long-term average AoI minimization problem into a per-slot Lyapunov drift minimization problem. We derive an explicit expression for the Lyapunov drift of the age-aware stationary randomized policy. (2) The per-slot Lyapunov drift minimization problem is non-convex, making it inefficient to solve with traditional optimization methods for real-time decision-making. To circumvent this issue, we propose a *learn-to-optimize* approach, which involves mapping the critical parameters of the problem directly to the solution using an artificial neural network. To that end, we carefully analyze the per-

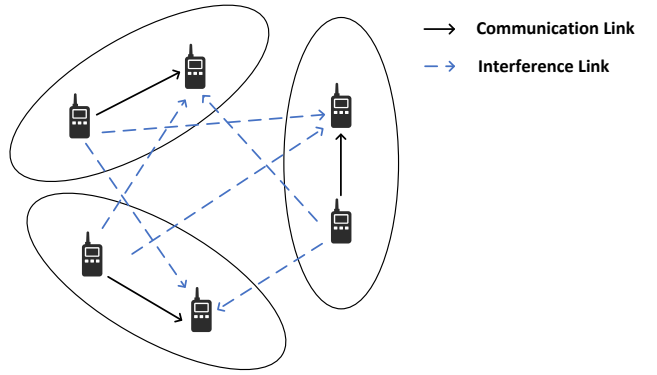


Fig. 1: A D2D communication network with $M = 3$ transmitter-receiver pairs.

mutation equivariance (PE) property of the optimal solutions to the per-slot problem and propose applying a message passing neural network (MPNN) [16], a popular framework within Graph Neural Networks (GNNs), to directly learn the targeted policy. The MPNN is trained in an unsupervised manner and can produce the scheduling probability for each link based on the system AoI and statistical CSI. We conduct computer simulations to evaluate the age-aware stationary randomized policy generated by the MPNN in various D2D networks, demonstrating its superior performance compared to baseline approaches. Notably, the policy maintains robust performance even when generalized to large-scale D2D networks.

The remainder of this paper is organized as follows. Sec. II describes the system model, the definition of age-aware stationary randomized policy, and problem formulation. Sec. III derives the Lyapunov drift for the age-aware policy and elaborates on how to optimize it via MPNN. Sec. IV presents simulation results, and Sec. V concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We consider a D2D wireless network where M independently and randomly distributed transmitter-receiver pairs can simultaneously transmit their respective status updates, as illustrated in Fig. 1. Each transmitter-receiver pair forms a D2D communication link. We use $\mathcal{M} \triangleq \{1, 2, \dots, M\}$ to denote the index set of D2D links, transmitters and receivers, i.e., link i contains transmitter i and receiver i . Each transmitter and receiver is equipped with a single antenna. The considered D2D network operates in a time-slotted manner, where the time synchronization among all D2D links could be achieved through the infrastructure (e.g., a base station). At the beginning of every slot, each transmitter can be scheduled to generate a new status update and transmit it to the receiver using the same frequency band. The entire status update process completes within one time slot. We assume that a centralized controller can be employed to schedule the status updates of D2D links, thereby alleviating the negative impact of excessive interference on successful status updates. Similar assumptions have been considered in [12], [17], [18].

The channel gain between transmitter i and receiver j in slot t is denoted by $h_{ij}(t) = \sqrt{h_{ij}^l h_{ij}^s(t)}$, where h_{ij}^l denotes the time-invariant large-scale power gain, $h_{ij}^s(t)$ represents small-scale Rayleigh fading channel gain, which follows a complex Gaussian distribution $\mathcal{CN}(0, 1)$. We assume the instantaneous channel state $h_{ij}(t)$ remains unchanged during each time slot but varies independently and identically distributed (*i.i.d.*) across different time slots. Specifically, $h_{ii}(t)$ is the channel gain of the direct D2D communication link of the i -th transmitter-receiver pair, while $h_{ij}(t)$ represents the channel gain of the interference link between transmitter i and receiver j , $j \neq i$. Denote the large-scale power gain matrix by $\mathbf{H}^l$, where $\mathbf{H}_{(i,j)}^l = h_{ij}^l$. Hereafter, we refer to $\mathbf{H}^l$ as the statistical CSI, which is available at the centralized controller.

The link scheduling action in time slot t is denoted by $\mathbf{a}(t) = [a_1(t), a_2(t), \dots, a_M(t)]^T$, where $a_i(t)$ is a binary indicator which is equal to 1 when transmitter i is scheduled, and 0 otherwise. Upon performing scheduling action $\mathbf{a}(t)$, the signal-to-interference-and-noise ratio (SINR) for receiver i in the t -th time slot can be expressed as

$$\xi_i(t) = \frac{P_{tx} |h_{ii}(t)|^2 a_i(t)}{\sum_{j \neq i} P_{tx} |h_{ji}(t)|^2 a_j(t) + \sigma^2}, \quad (1)$$

where P_{tx} is the identical transmission power of every transmitter, σ^2 is the noise power at each receiver.

We consider a widely adopted transmission model where a status update is successfully delivered only if the SINR of the receiver is not smaller than a pre-defined threshold β [4], [8], [11], [13], [14]. Denote by $d_i(t)$ the transmission result of device i in slot t . We thus have $d_i(t)$ equals to 1 if $\xi_i(t) \geq \beta$, and 0 otherwise. Given the transmission model, we are now ready to define the AoI of receiver i as $g_i(t)$. Generally, AoI measures the time elapsed since the latest update received by the receiver was generated [1]. Assuming that all scheduled devices can sample and generate fresh information packets for transmission, known as the *generate-at-will* model [19], the AoI $g_i(t+1)$ of receiver i at the beginning of the $(t+1)$ -th slot evolves as follows:

$$g_i(t+1) = g_i(t) + 1 - g_i(t)d_i(t). \quad (2)$$

B. Problem Formulation

Define the vector $\mathbf{g}(t) = [g_1(t), g_2(t), \dots, g_M(t)]^T$ to represent all receivers' AoI at the beginning of time slot t . Denote Π as the set of non-anticipated policies, in which link scheduling action $\mathbf{a}(t)$ is made based on the scheduling decision history $\{\mathbf{a}(k)\}_{k=1}^{t-1}$, the evolution of AoI $\{\mathbf{g}(k)\}_{k=1}^t$, and statistical CSI $\mathbf{H}^l$. We use statistical CSI for decision-making instead of instantaneous CSI, because instantaneous CSI requires channel estimation for each link in each time slot, which incurs a large signaling overhead. In contrast, statistical CSI is easier to obtain and provides important information about the communication quality of different links. The objective is then to find the optimal scheduling policy $\pi \in \Pi$ such that the following long-term average AoI is minimized,

$$\min_{\pi \in \Pi} J(\pi), \quad (3)$$

where $J(\pi) = \lim_{T \rightarrow \infty} \frac{1}{TM} \mathbb{E}_\pi \left[\sum_{t=1}^T \sum_{i=1}^M g_i(t) \right]$, and the expectation is with respect to the random channel states and possible random operations in policy π .

While problem (3) can be formulated as an MDP, searching for the optimal policy within the set Π using conventional MDP-based methods is challenging due to the curse of dimensionality. To address this issue, we narrow the policy search space by focusing on the age-aware stationary randomized policy π_A . For convenience, we may also refer to this policy as age-aware policy in the following sections.

Under the policy π_A , each transmitter i will be scheduled independently with probability $p_i(t)$ at any time slot t . Define the probability vector $\mathbf{p}(t) = [p_1(t), p_2(t), \dots, p_M(t)]^T$, which is determined by the current AoI vector $\mathbf{g}(t)$ and statistical CSI $\mathbf{H}^l$. Allowing for some abuse of notation, we can also denote the mapping between $\mathbf{g}(t)$, $\mathbf{H}^l$ and $\mathbf{p}(t)$ as π_A , i.e., $\mathbf{p}(t) = \pi_A(\mathbf{g}(t), \mathbf{H}^l)$. This mapping will not change over time, i.e., the policy is stationary. Let Π_A denote the set of all such age-aware policies. With this notation, we can reformulate the problem (3) as follows:

$$\min_{\pi_A \in \Pi_A} J(\pi_A). \quad (4)$$

We remark that unlike age-independent stationary randomized policy π_S developed in [13], [14], the dynamic nature of AoI in our problem results in a time-varying probability vector. This complexity makes it challenging to derive a closed-form solution for $J(\pi_A)$ and complicates the task of directly obtaining the optimal age-aware policy π_A^* . In the following section, we first transform problem (4) by applying the Lyapunov optimization framework and then elaborate on how to find a well-performing and computationally efficient age-aware policy.

III. AGE-AWARE POLICY BASED ON LYAPUNOV OPTIMIZATION AND MPNN

In this section, we first derive an explicit expression for the Lyapunov drift of the age-aware stationary randomized policy, and then prove that the solution to the Lyapunov drift minimization problem possesses the property of permutation equivariance. Finally, we introduce a learn-to-optimize method based on message passing neural network for real-time decision-making.

A. Lyapunov Drift for Age-aware Policy

To proceed, we transform the long-term optimization problem (4) into a per-slot optimization problem using Lyapunov optimization [15]. This approach is inspired by the Max-Weight policy proposed in [6], [7], which minimizes per-slot Lyapunov drift to achieve near-optimal AoI performance. By designing a Lyapunov function and reducing its expected increase in each time slot, we can maintain the system in a desirable state, where each D2D pair has a relatively low AoI. Before demonstrating the problem transformation, we introduce some key definitions, including network state, Lyapunov function, and Lyapunov drift.

and the edge feature ε_{ij} is h_{ij}^l . The set of edges is denoted by $\mathcal{E}$. The adjacency feature matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$ can be defined as

$$\mathbf{A}_{(i,j)} = \begin{cases} 0, & \text{if } \{i, j\} \notin \mathcal{E}, \\ \varepsilon_{ij}, & \text{otherwise.} \end{cases} \quad (8)$$

Based on all the definitions above, the directed graph associated with our problem can be represented as $G = (\mathcal{V}, \mathcal{E}, \mathbf{Z}, \mathbf{A})$. Denote the graph optimization variable by $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_M]^T \in [0, 1]^M$, where μ_i represents the scheduling probability associated with node v_i . Based on the directed graph $G = (\mathcal{V}, \mathcal{E}, \mathbf{Z}, \mathbf{A})$, the problem (7) in every slot can be reformulated as the following equivalent graph optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\mu}} \hat{f}(\boldsymbol{\mu}, \mathbf{Z}, \mathbf{A}) &= - \sum_{i=1}^M \mathbf{Z}_{(i,1)} \rho_i \mu_i \prod_{j \neq i} \left(1 - \frac{\mu_j}{1 + D_{ji}} \right), \\ \text{s.t. } \rho_i &= \exp \left(- \frac{\beta \sigma^2}{P_{tx} \mathbf{Z}_{(i,2)}} \right), \quad D_{ji} = \frac{\mathbf{Z}_{(i,2)}}{\beta \mathbf{A}_{(j,i)}}, \\ \boldsymbol{\mu} &= [\mu_1, \mu_2, \dots, \mu_M]^T, 0 \leq \mu_i \leq 1, \forall i \in \mathcal{M}. \end{aligned} \quad (9)$$

As will become clear later, formulating the graph problem (9) will facilitate the training process of the MPNN.

To proceed, we provide a brief introduction of MPNN. An MPNN consists of multiple propagation layers and a readout layer [16]. The structure of the MPNN is illustrated in Fig. 2. Each propagation layer consists of a message passing phase, an aggregation phase, and an update phase. The readout layer may have different forms depending on the specific graph-based task. The primary objective of MPNN is to acquire an informative embedding vector for every node in the graph through the propagation layers and then generate a target value via the readout layer. During the propagation phase, a node updates its embedding using its own feature and embedding, and the features of neighboring nodes and connecting edges, and the embeddings of neighboring nodes. The detailed process is depicted in Fig. 3. Let $\mathbf{e}_i^{(n)} \in \mathbb{R}^E$ denote the embedding representation of node i after the n -th propagation layer, where E is the node embedding size. For the 0-th layer or the input layer, the graph embedding of each node is initialized to zero vector, i.e., $\mathbf{e}_i^{(0)} = \mathbf{0}$. Then the update rule of $\mathbf{e}_i^{(n)}$ can be described as

$$\begin{aligned} \text{Message: } \mathbf{m}_{ji}^{(n)} &= \Phi^{(n)} \left(\mathbf{e}_j^{(n-1)}, \mathbf{z}_j, \varepsilon_{ji} \right), \forall j \in \mathcal{N}(i), \\ \text{Aggregation: } \mathbf{a}_i^{(n)} &= \bigoplus_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}^{(n)}, \\ \text{Update: } \mathbf{e}_i^{(n)} &= U^{(n)} \left(\mathbf{e}_i^{(n-1)}, \mathbf{z}_i, \mathbf{a}_i^{(n)} \right), \end{aligned}$$

where $\bigoplus$ denotes a differentiable, permutation invariant aggregation function such as *SUM*, *MIN* or *MAX*, $\mathcal{N}(i)$ denotes the set of neighbors of i . The message passing function $\Phi^{(n)}(\cdot)$ and update function $U^{(n)}(\cdot)$ are multi-layer perceptions (MLPs) with parameters φ_n and ψ_n , respectively. If not specified, the MLPs in this paper all adopt ReLU as the activation function. Aggregation function *MAX* is adopted in

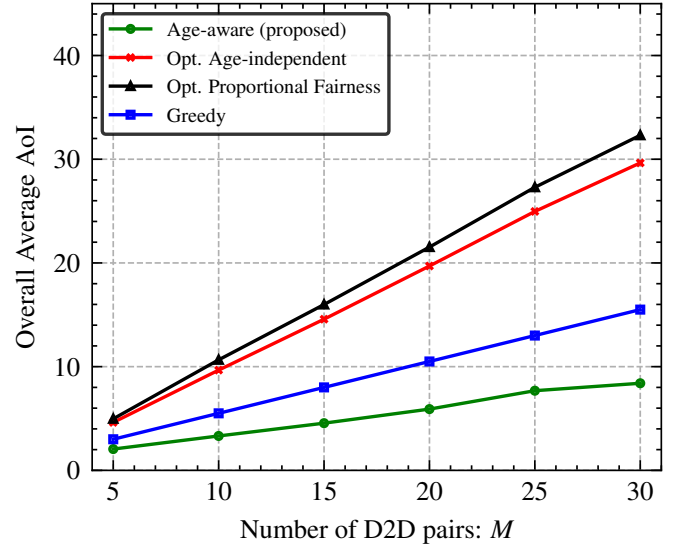


Fig. 4: Overall AoI performance comparison for different M over 500 layouts.

this paper for its robustness to feature corruptions [23]. In MPNN, we will set up N propagation layers and N should not be too large to avoid over-smoothing, i.e., all the nodes will have similar embeddings. An efficient design is that we let the message passing functions and update functions of N layers share the same MLP parameters φ and ψ respectively. After N -layer propagation, the readout layer needs to map the learned embedding $\mathbf{e}_i^{(N)}$ to the scheduling probability μ_i of node i . Since $\mu_i \in [0, 1]$, the readout layer can be designed as an MLP with parameters Ω and a Sigmoid activation function.

The entire MPNN is trained using the unsupervised learning paradigm. It employs the loss function $\mathbb{E} \left\{ \hat{f}(\boldsymbol{\mu}(\varphi, \psi, \Omega), \mathbf{Z}, \mathbf{A}) \right\}$ and utilizes stochastic gradient descent to update network parameters $\{\varphi, \psi, \Omega\}$, where $\hat{f}$ is defined in the graph problem (9). During the training phase, a batch of different training samples $\{\mathbf{Z}, \mathbf{A}\}$ is used to compute gradients. During the inference phase, we use statistical CSI and the system's AoI to generate the randomized scheduling policy for each time slot.

IV. SIMULATION RESULTS

In this section, we compare the performance of the proposed age-aware policy based on Lyapunov optimization and MPNN with the following baselines:

- **Optimal Age-independent Stationary Randomized policy** [13], [14]: Obtain π_S^* by minimizing $J(\pi_S)$ via convex optimization.
- **Optimal Proportional Fairness** [14]: An age-independent stationary randomized policy π_S that minimizes the sum log of average AoI.
- **Greedy Policy**: Only schedule the D2D pair with the highest AoI.

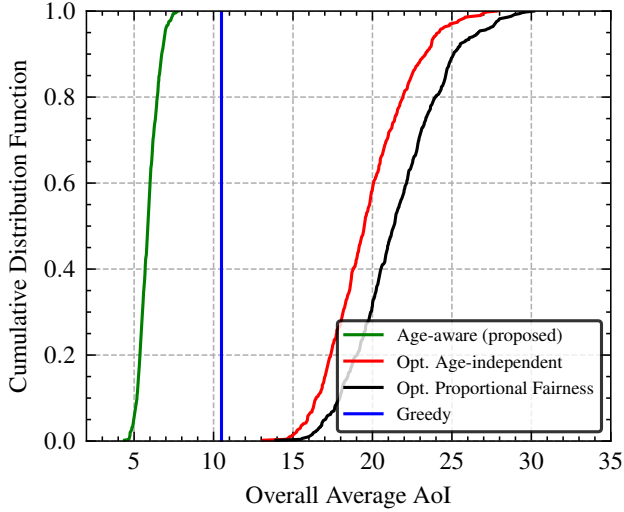


Fig. 5: CDF of overall average AoI over 500 layouts: $M = 20$.

A. Parameter Setting

The D2D network is within an $L \times L$ area, where L is the area length. The link density is set to M/L^2 links/m². If not specified, $L = 500$ m for all the training and testing scenarios. All transmitters are randomly distributed in the area and each receiver is uniformly distributed within 2m $\sim$ 40m from the corresponding transmitter. The D2D communication channel follows ultra high frequency model in ITU-1411 [24]. The large-scale power gain in dB, i.e., path-loss, is captured by

$$L_{pl} = L_{bp} + 6 + \begin{cases} 20 \log_{10} \left(\frac{d}{R_{bp}} \right), & \text{for } d \leq R_{bp} \\ 40 \log_{10} \left(\frac{d}{R_{bp}} \right), & \text{for } d > R_{bp} \end{cases}, \quad (10)$$

where d is the distance between the transmitter and the receiver, R_{bp} is the breakpoint distance defined as $R_{bp} = \frac{4h_1h_2}{\lambda}$ with h_1 and h_2 being antenna heights of the transmitter and receiver, and λ is the carrier wavelength. L_{bp} in (10) is defined as $L_{bp} = \left\lceil 20 \log_{10} \left(\frac{\lambda^2}{8\pi h_1 h_2} \right) \right\rceil$, representing the basic transmission loss at the breakpoint. In our simulation, the used carrier frequency is 2.4 GHz, with a bandwidth of 5 MHz. The antenna heights of both transmitter and receiver are set to be 1.5m, and the antenna gain is 2.5 dBi. The transmission power P_{tx} is 40 dBm, and the noise power spectral density is -169 dBm/Hz. The threshold $\beta = 1023$. We generate weight w_i from a uniform distribution in $[0, 1]$ for any $i \in \mathcal{M}$.

For the parameter setting of MPNN, we set the dimension E of node embedding to 8, the layer number $N = 3$, the number of hidden neurons in all the message passing functions as $\{11, 32, 32\}$, the number of hidden neurons in all the update functions as $\{42, 16, 8\}$, the number of hidden neurons in readout layer as $\{8, 16, 1\}$. We used the Adam optimizer [25] with a learning rate of 0.002 and a decay factor of 0.9. The training process consists of 100 epochs of batch training with batch size 50. For the input data of MPNN, we note that the interference link is considered only if the transmitter i and receiver j are within 500 meters. For the data normalization, both the large-scale power gain h_{ii}^l in

node feature and the large-scale power gain h_{ij}^l in edge feature are in dB and normalized by z-score $\hat{x} = (x - \bar{x})/\hat{\sigma}$, where x is the input data, $\bar{x}$ is the average value of data, $\hat{\sigma}$ is the corresponding standard deviation. The weight w_i is normalized by $\hat{w}_i = w_i/\max\{w_1, w_2, \dots, w_M\}$.

B. Simulation Results

The long-term average AoI performance of each scheduling policy is evaluated over 20,000 time slots for every given D2D layout. For our proposed age-aware policy, the MPNN for solving problem (9) is trained with 50000 samples of $\{\mathbf{Z}, \mathbf{A}\}$ for $M = 5, 10, 15, 20, 25, 30$ in $500\text{m} \times 500\text{m}$ area, each of which contains independently randomly generated D2D layout and weights. In the testing phase, the trained MPNN generates the scheduling probability vector at the beginning of each time slot, using statistical CSI and the current system AoI as inputs. By averaging 500 testing layouts, Fig. 4 depicts the overall AoI performance of the age-aware policy and other baselines versus M . It can be observed that the proposed policy outperforms other policies, while the optimal age-independent stationary randomized policy can even exhibit worse performance than the greedy policy. One possible reason is that the optimal age-independent stationary randomized policy may often schedule links with low AoI while ignoring links with very high AoI. This observation validates the importance of incorporating AoI for link scheduling. Moreover, We also present the cumulative distribution function (CDF) of overall AoI performance in Fig. 5 for $M = 20$ over 500 D2D testing layouts. The CDF curve of the age-aware policy always locates in the upper left compared to other policies, demonstrating the stable inference performance of the MPNN.

Next, we evaluate the generalization capability of the proposed method, specifically whether the MPNN trained on D2D networks with tens of D2D pairs can be directly deployed on D2D networks with hundreds of D2D pairs without fine-tuning. To this end, we train the MPNN in a $500\text{m} \times 500\text{m}$ layout with 20 D2D pairs using 50000 training samples and test its scalability in the following two cases: same link density and different link densities.

For the same link density case, we test the trained MPNN across 100 testing layouts, each containing different numbers of D2D pairs, with L increasing as M increases. The results are shown in Fig. 6. Although the number of D2D pairs in the testing stage is different from that in the training stage, the performance of the age-aware policy remains stable and significantly outperforms the baselines.

For the different link density cases, we test the trained MPNN in 100 layouts, all within a $500\text{m} \times 500\text{m}$ area but with varying numbers of D2D pairs. The results, shown in Fig. 7, indicate that the age-aware policy maintains a performance advantage over the baselines even when the link density is increased tenfold. These results collectively demonstrate the scalability of MPNN in terms of AoI performance.

V. CONCLUSION

In this paper, we developed an age-aware link scheduling policy for long-term average AoI minimization in device-to-

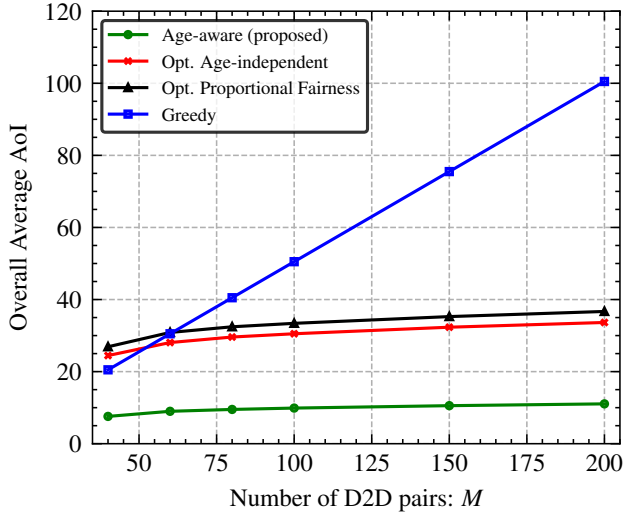


Fig. 6: Generalize to the same density.

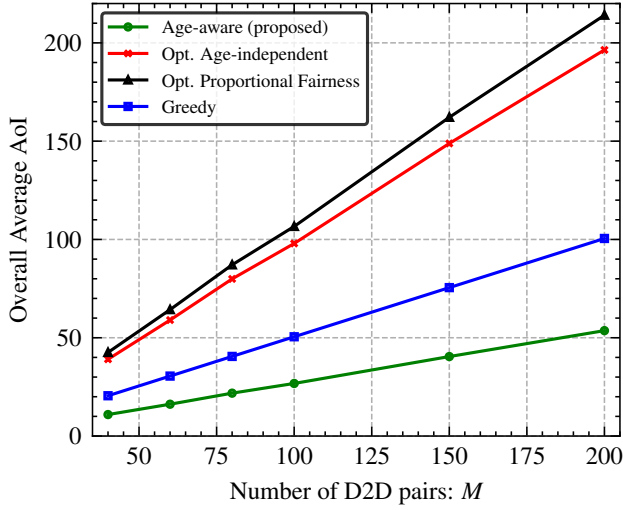


Fig. 7: Generalize to different densities.

device (D2D) networks. Specifically, we proposed a message passing neural network (MPNN)-based method to generate the age-aware stationary randomized policy for real-time link scheduling, using only statistical CSI and system age of information (AoI). The MPNN was trained in an unsupervised manner by minimizing the Lyapunov drift of the original long-term average AoI minimization problem. The performance of the age-aware stationary randomized policy has been demonstrated to be stable and superior to several baselines. Additionally, the proposed method exhibits strong scalability capability.

REFERENCES

- [1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.
- [2] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86–92, May 2014.
- [3] K. Shen and W. Yu, "FPLinQ: A cooperative spectrum sharing strategy for device-to-device communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2323–2327.
- [4] F. Baccelli and C. Singh, "Adaptive spatial aloha, fairness and stochastic geometry," in *Proc. 11th Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw. (WiOpt)*, May 2013, pp. 7–14.
- [5] F. Baccelli, B. Błaszczyszyn, and C. Singh, "Analysis of a proportionally fair and locally adaptive spatial aloha in Poisson networks," in *Proc. IEEE INFOCOM*, Jul. 2014, pp. 2544–2552.
- [6] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.
- [7] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Trans. Mobile Comput.*, vol. 20, no. 3, pp. 1173–1185, Mar. 2021.
- [8] Q. He, D. Yuan, and A. Ephremides, "Optimizing freshness of information: On minimum age link scheduling in wireless systems," in *Proc. 14th Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw. (WiOpt)*, Jun. 2016, pp. 1–8.
- [9] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 15–28, Dec. 2019.
- [10] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [11] S. Leng and A. Yener, "Age of information minimization for wireless ad hoc networks: A deep reinforcement learning approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [12] Y. Liu, C. She, W. Hardjawana, and B. Vucetic, "Graph neural networks for timely updates of short packets in interference-limited networks," in *Proc. Asilomar Conf. Signals Syst. Comput. (ACSSC)*, Oct. 2022, pp. 1050–1054.
- [13] Z. Liu, Z. Chen, L. Luo, M. Hua, W. Li, and B. Xia, "Age of information-based scheduling for wireless device-to-device communications using deep learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.
- [14] N. Jones and E. Modiano, "Minimizing age of information in spatially distributed random access wireless networks," in *Proc. IEEE INFOCOM*, May 2023, pp. 1–10.
- [15] M. Neely, *Stochastic network optimization with application to communication and queueing systems*. Springer Nature, 2022.
- [16] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Aug. 2017, pp. 1263–1272.
- [17] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlying cellular networks," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2752–2763, Jun. 2011.
- [18] P. Phunchongharn, E. Hossain, and D. I. Kim, "Resource allocation for device-to-device communications underlying LTE-advanced networks," *IEEE Wireless Commun. Mag.*, vol. 20, no. 4, pp. 91–100, Aug. 2013.
- [19] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksall, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.
- [20] M. Eisen and A. Ribeiro, "Optimal wireless resource allocation with random edge graph neural networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 2977–2991, April. 2020.
- [21] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 101–115, Jan. 2021.
- [22] J. Guo and C. Yang, "Learning power allocation for multi-cell-multi-user systems with heterogeneous graph neural networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 884–897, Aug. 2021.
- [23] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3d classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 77–85.
- [24] *Propagation Data and Prediction Methods for the Planning of Short-Range Outdoor Radiocommunication Systems and Radio Local Area Networks in the Frequency Range 300 MHz to 100 GHz*, document I. R. P.1411-11 2021.
- [25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Dec. 2014, pp. 1–15.

APPENDIX A

PROOF OF THEOREM 1

First, by the definition of Lyapunov function $L(S(t))$ and incorporating (2) and (5), we have

$$\Delta(S(t)) = - \sum_{i=1}^M \mathbb{E}_{\pi_A} \{d_i(t) | S(t)\} W_i(t) + B(t) \quad (11)$$

for any age-aware policy $\pi_A \in \Pi_A$.

To proceed, we first deal with the term $\mathbb{E}_{\pi_A} \{d_i(t) | S(t)\}$, which represents the transmission success probability of the transmitter i in slot t by the definition of $d_i(t)$. Under the scheduling policy π_A , each transmitter is scheduled to generate and transmit updates independently with probability $p_i(t)$, which is the i -th entry in the probability vector $\mathbf{p}(t) = \pi_A(\mathbf{g}(t), \mathbf{H}^l)$. Then the transmission success probability of the transmitter i is the product of its scheduling probability and conditional success probability, thus we have

$$\mathbb{E}_{\pi_A} \{d_i(t) | S(t)\} = p_i(t) \mathbb{P}[\xi_i(t) \geq \beta | a_i(t) = 1]. \quad (12)$$

The conditional success probability can be obtained using a procedure similar to that in [4], [13], [14]. Specifically, we first condition on the channel randomness of interference links and the randomized transmissions of other transmitters to have

$$\begin{aligned} & \mathbb{P}[\xi_i(t) \geq \beta | a_i(t) = 1, h_{ji}(t), a_j(t), \forall j \neq i] \\ & \stackrel{(a)}{=} \mathbb{P}\left[|h_{ii}(t)|^2 \geq \frac{\beta(\sum_{j \neq i} P_{tx} |h_{ji}(t)|^2 a_j(t) + \sigma^2)}{P_{tx}}\right] \\ & \stackrel{(b)}{=} \exp\left(-\frac{\beta}{h_{ii}^l} \left(\sum_{j \neq i} |h_{ji}(t)|^2 a_j(t) + \frac{\sigma^2}{P_{tx}}\right)\right) \\ & \stackrel{(c)}{=} \rho_i \prod_{j \neq i} \exp\left(-\frac{|h_{ji}^s(t)|^2 a_j(t)}{D_{ji}}\right), \end{aligned} \quad (13)$$

where equation (a) is derived from the definition of SINR for receiver i in (1), equation (b) is due to the fact that $|h_{ii}(t)|^2 = h_{ii}^l |h_{ii}^s(t)|^2$ and $|h_{ii}^s(t)|^2$ follows an exponential distribution with a mean of 1, equation (c) is by the notation $\rho_i = \exp(-\beta\sigma^2/(P_{tx}h_{ii}^l))$ and $D_{ji} = h_{ii}^l/(\beta h_{ji}^l)$. After that, by taking expectation on (13) with respect to the independent randomized transmission of other transmitters, we have

$$\begin{aligned} & \mathbb{P}[\xi_i(t) \geq \beta | a_i(t) = 1, h_{ji}(t), \forall j \neq i] \\ & = \mathbb{E}_{\{a_j(t), \forall j \neq i\}} \left\{ \rho_i \prod_{j \neq i} \exp\left(-\frac{|h_{ji}^s(t)|^2 a_j(t)}{D_{ji}}\right) \right\} \\ & \stackrel{(d)}{=} \rho_i \prod_{j \neq i} \mathbb{E}_{\{a_j(t), \forall j \neq i\}} \left\{ \exp\left(-\frac{|h_{ji}^s(t)|^2 a_j(t)}{D_{ji}}\right) \right\} \\ & \stackrel{(e)}{=} \rho_i \prod_{j \neq i} \left[1 - p_j(t) + p_j(t) \exp\left(-\frac{|h_{ji}^s(t)|^2}{D_{ji}}\right) \right], \end{aligned} \quad (14)$$

where equation (d) follows by the independent scheduling between transmitters, equation (e) is obtained because $a_j(t)$

follows Bernoulli distribution $\text{Ber}(p_j(t))$. Finally, by taking expectation on (14) with respect to the channel randomness of interference links, we have

$$\begin{aligned} & \mathbb{P}[\xi_i(t) \geq \beta | a_i(t) = 1] \\ & = \mathbb{E}_{\{h_{ji}(t), \forall j \neq i\}} \left\{ \rho_i \prod_{j \neq i} \left[1 - p_j(t) + \frac{p_j(t)}{\exp(|h_{ji}^s(t)|^2/D_{ji})} \right] \right\} \\ & = \rho_i \prod_{j \neq i} \left(1 - \frac{p_j(t)}{1 + D_{ji}} \right), \end{aligned} \quad (15)$$

where the last equation is because small-scale channel gain $h_{ji}^s(t)$ is independent across interference links and it follows an exponential distribution with a mean of 1.

We complete the proof by substituting (12) and (15) into (11).

APPENDIX B

PROOF OF THEOREM 2

Before proving Theorem 2, we first show the permutation invariance of f in problem (7).

Lemma 1 (Permutation Invariance of f). *Consider the problem (7) with critical parameter $\{\mathbf{w}, \mathbf{H}^l\}$, and $\mu = [\mu_1, \mu_2, \dots, \mu_M]$ is any feasible solution to the problem (7). For any permutation matrix $\mathbf{P} \in \{0, 1\}^{M \times M}$ satisfies $\mathbf{P}\mathbf{1} = \mathbf{1}$ and $\mathbf{P}^T\mathbf{1} = \mathbf{1}$, the function f in problem (7) is permutation invariant, i.e.,*

$$f(\mu, \mathbf{w}, \mathbf{H}^l) = f(\mathbf{P}\mu, \mathbf{P}\mathbf{w}, \mathbf{P}\mathbf{H}^l\mathbf{P}^T).$$

Proof. Define $\xi_i = W_i(t)\rho_i\mu_i \prod_{j \neq i} \left(1 - \frac{\mu_j}{1 + D_{ji}}\right)$, and we have $f(\mu, \mathbf{w}, \mathbf{H}^l) = \sum_{i=1}^M \xi_i$. By observing the expression of ξ_i , any permutation to the variable and critical parameters only leads to a reordering of the sequence $\{\xi_1, \xi_2, \dots, \xi_M\}$, while the overall summation remains unchanged. This implies that the function f is permutation invariant. $\square$

Based on Lemma 1, we can prove the permutation equivariance of Υ . Define a set $\hat{\Upsilon} = \{\mathbf{P}\mu | \mu \in \Upsilon\}$. To prove $\Upsilon' = \hat{\Upsilon}$, it suffices to prove $\hat{\Upsilon} \subseteq \Upsilon'$ and $\Upsilon' \subseteq \hat{\Upsilon}$. We first show $\hat{\Upsilon} \subseteq \Upsilon'$ holds. Any optimal solution $\mu \in \Upsilon$ achieves the optimum of problem (7) based on critical parameters $\{\mathbf{w}, \mathbf{H}^l\}$. The permutation invariance of f promises that solution $\mathbf{P}\mu$ will also achieve the same optimum of the reformulated problem (7) based on the permuted parameters $\{\mathbf{P}\mathbf{w}, \mathbf{P}\mathbf{H}^l\mathbf{P}^T\}$. Therefore, $\mathbf{P}\mu \in \Upsilon'$, which verifies $\hat{\Upsilon} \subseteq \Upsilon'$. We turn to prove $\Upsilon' \subseteq \hat{\Upsilon}$. Specifically, based on the above definitions, D2D network $\mathcal{D}'$ can be transformed into D2D network $\mathcal{D}$ with some inverse permutation operations, and the corresponding critical parameters $\{\mathbf{P}\mathbf{w}, \mathbf{P}\mathbf{H}^l\mathbf{P}^T\}$ can also be permuted into critical parameters $\{\mathbf{w}, \mathbf{H}^l\}$ with permutation matrix $\mathbf{P}^T$. Thus, for any optimal solution $\mu' \in \Upsilon'$, the permutation invariant function f guarantees $\mathbf{P}^T\mu' \in \Upsilon$. Recall the definition of set $\hat{\Upsilon}$ and $\mathbf{P}\mathbf{P}^T = \mathbf{I}$, we then have $\mu' = \mathbf{P}(\mathbf{P}^T\mu') \in \hat{\Upsilon}$. This completes the proof.