

Correspondence

Proactive Eavesdropping in Massive MIMO-OFDM Systems via Deep Reinforcement Learning

Jiale Chen¹, Lan Tang¹, Member, IEEE, Delin Guo¹,
Yechao Bai², Member, IEEE, Lvxi Yang³, Senior Member, IEEE,
and Ying-Chang Liang⁴, Fellow, IEEE

Abstract—It is challenging to monitor suspicious links in massive multiple input multiple output-orthogonal frequency division multiplexing (MIMO-OFDM) systems with directional beamforming. When the monitor and suspicious user are not in the coverage of the same beam, to implement legitimate monitoring, we propose the proactive eavesdropping scheme in which the monitor acts as a spoofing relay to realize beam misleading and data eavesdropping. In the phase of beam sweep, the beam misleading algorithm induces the transmitter to choose the beam which is beneficial to the monitor by optimizing the relay precoding matrices. In the phase of data transmission, the data eavesdropping algorithm elevates the monitoring rate by optimizing the relay power split factors and power gain factors of all subcarriers. Due to the unknown channel state information between suspicious pairs, we propose to find the optimal precoders and power split factors by the multi-agent deep deterministic policy gradient (MADDPG) algorithm. The simulation results verify the effectiveness and superiority of the proposed eavesdropping scheme.

Index Terms—Proactive eavesdropping, massive MIMO-OFDM, deep reinforcement learning, beam misleading, spoofing relay.

I. INTRODUCTION

Massive multiple-input multiple-output orthogonal-frequency-division-multiplexing (MIMO-OFDM) is considered as one of the key technologies of the 5th generation (5G) mobile communication system [1], [2]. However, when the massive antenna array is employed in 5G base stations, the directional narrow beam makes the traditional monitoring methods inefficient or even invalid. Therefore, to realize the legitimate monitoring of suspicious links, it is critical to investigate the monitoring scheme in the narrow beam scene.

The existing literature on eavesdropping can be divided into three categories: passive eavesdropping, proactive eavesdropping by jamming, and proactive eavesdropping by spoofing relay. In passive eavesdropping, the eavesdropper keeps silent while eavesdropping. This approach is effective only when the eavesdropping channel is better than the suspicious channel. In order to overcome this disadvantage,

proactive eavesdropping by jamming has been proposed, in which the eavesdropper transmitted interference signals to the suspicious receiver to force the transmitter to slow down the rate so that information can be decoded by the eavesdropper [3], [4]. To implement proactive eavesdropping flexibly, an eavesdropping method called spoofing relay was proposed [5]. This method can maximize the monitoring rate by disguising the monitor as a relay when the eavesdropping channel was superior to the suspicious channel. However, massive MIMO systems have the property of “secrecy for free” because the beam towards the intended user is so narrow that the information leaked to the eavesdropper tends to zero [6]. Therefore, the eavesdropper outside the beam coverage can hardly eavesdrop successfully using the passive or proactive eavesdropping methods mentioned above.

In the case of operation at higher frequencies with a large number of antennas, the antenna processing will be carried out in the analog domain with a focus on beamforming. In analog beamforming, beam management is implemented to establish and retain a suitable transmitting and receiving beam pair which jointly provide good connectivity [1]. Inspired by the spoofing relay and beam management in 5G new radio (NR), we propose an eavesdropping scheme in massive MIMO-OFDM systems, where a legitimate eavesdropper acts as an amplify-and-forward (AF) relay to realize the beam misleading and data eavesdropping. Different from eavesdropping by jamming, the proposed scheme will not degrade the communication quality of the suspicious link during monitoring, but realize monitoring on the basis of improving the communication quality of the suspicious link. To the best of our knowledge, none of the existing work studied eavesdropping schemes in massive MIMO-OFDM systems with analog beamforming. The main contributions of this paper are as follows

1) To save the total transmit power of the monitor conditioned on successful monitoring, we optimize the relaying parameters in both the beam misleading and data eavesdropping phases of the monitor. In the beam misleading phase, we optimize the precoding matrix on each subcarrier of the monitor to induce the transmitter to choose the beam that is favorable for the monitor. In the data eavesdropping phase, we optimize the power split factors and power gain factors to maximize the monitoring rate under the transmit power constraint. We derive the optimal solution of the relaying parameters on each subcarrier when channel state information (CSI) of the suspicious communication pair is known.

2) Considering that it is difficult for the monitor to obtain CSI of the suspicious link, we further propose parameters optimization methods based on deep reinforcement learning (DRL) to handle the partial CSI issue. In the formulated problem, the dimension of action space is proportional to the number of subcarriers. Large action space will lead to learning difficulties and slow down the convergence speed [7]. To reduce the dimension of the action space, we transform the action of each subcarrier from a precoding matrix into two variables according to the expression of the optimal solution in the phase of beam misleading and adopt the multi-agent deep deterministic policy gradient (MADDPG) to split a large actor network into multiple small actor networks.

3) Simulation results demonstrate that the policies learned by MADDPG can successfully realize beam misleading and data eavesdropping, and achieve better eavesdropping performance than traditional methods.

Manuscript received 22 December 2021; revised 25 April 2022; accepted 2 July 2022. Date of publication 13 July 2022; date of current version 14 November 2022. This work was supported in part by the National Key R&D Program of China under Grant 2019YFE0113200 and in part by the National Natural Science Foundation of China under Grants U1936201, 62072229, 62071220, and 61976113. The review of this article was coordinated by Prof. Namyoon Lee. (Corresponding author: Lan Tang.)

Jiale Chen, Lan Tang, Delin Guo, and Yechao Bai are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China (e-mail: mf1923004@smail.nju.edu.cn; tanglan@nju.edu.cn; 1018899193@qq.com; ychbai@nju.edu.cn).

Lvxi Yang is with the School of information science and engineering, Southeast University, Nanjing 210096, China (e-mail: lxyang@seu.edu.cn).

Ying-Chang Liang is with the Center for Intelligent Networking and Communication (CINC), University of Electronic Science and Technology of China, Chengdu 610056, China (e-mail: liangyc@ieee.org).

Digital Object Identifier 10.1109/TVT.2022.3190382

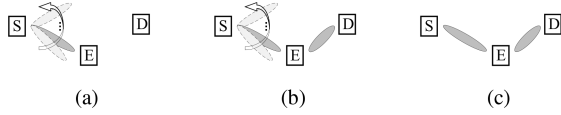


Fig. 1. Eavesdropping process of E.

II. SYSTEM MODEL

We consider a legitimate eavesdropping system which consists of a pair of communication nodes (S and D) and a legitimate eavesdropper (E). S and D are equipped with N_{ts} transmit antennas and N_{rd} receive antennas, respectively. We assume that both S and D adopt large-scale antenna arrays with analog beamforming to transmit and receive information signals. As in [8], the analog beamforming vector of S is selected from a predefined discrete codebook $\mathcal{F}_s$. E acts as a full-duplex spoofing relay, which receives signals from S with N_{re} antennas and forwards signals to D with N_{te} antennas simultaneously. To improve the monitoring quality, E adopts digital beamforming on each subcarrier. We assume all channels in the system keep constant in each transmission block but may vary according to Markov model between different blocks.

As shown in Fig. 1, for source S, the whole process in each block includes two phases: the beam sweeping (BSP) phase and the data transmission (DT) phase. While for monitor E, the eavesdropping process includes three phases: beam selection, beam misleading, and data eavesdropping. In the beam selection phase (Fig. 1(a)), the monitor selects the beam based on the received signal-to-noise ratio (SNR) and beam quality information from the feedback channel. Specifically, when S performs BSP with beamforming vectors in $\mathcal{F}_s$, the received signals by

D and E on the k th subcarrier are $\mathbf{y}_D^{[j,k]} = \sqrt{\frac{P_S^k}{N_{ts}}} \mathbf{H}_{SD}^k \mathbf{f}_j s_k + \mathbf{n}_D^k$ and

$\mathbf{y}_E^{[j,k]} = \sqrt{\frac{P_S^k}{N_{ts}}} \mathbf{H}_{SE}^k \mathbf{f}_j s_k + \mathbf{n}_E^k$, respectively, where s_k is the reference

signal on the k th subcarrier with transmit power P_S^k , $\mathbf{f}_j \in \mathcal{F}_s$ is the constant envelope beamforming vector with $|\mathbf{f}_j(n)| = 1, n = 1, \dots, N_{ts}$, $\mathbf{H}_{SD}^k \in \mathbb{C}^{N_{rd} \times N_{ts}}$ and $\mathbf{H}_{SE}^k \in \mathbb{C}^{N_{re} \times N_{ts}}$ are channel matrices from S to D and S to E on the k th subcarrier, respectively, $\mathbf{n}_D^k \in \mathbb{C}^{N_{rd} \times 1}$ and $\mathbf{n}_E^k \in \mathbb{C}^{N_{re} \times 1}$ are the additive Gaussian white noises (AWGN) with zero mean and covariance matrix $\sigma^2 \mathbf{I}$. At the destination D,

the received signals on all subcarriers are processed using an analog combiner $\mathbf{v}_D \in \mathbb{C}^{N_{rd} \times 1}$ with $\|\mathbf{v}_D\|^2 = N_{rd}$ and at E, the received signals on the k th subcarrier are processed by employing a digital combiner $\mathbf{v}_E^k \in \mathbb{C}^{N_{re} \times 1}$. D and E measure SNR of each beam on the k th subcarrier $\beta_D^{[j,k]} = \frac{P_S^k |(\mathbf{v}_D)^H \mathbf{H}_{SD}^k \mathbf{f}_j|^2}{\sigma^2 N_{ts} N_{rd}}$, and $\beta_E^{[j,k]} = \frac{P_S^k |(\mathbf{v}_E^k)^H \mathbf{H}_{SE}^k \mathbf{f}_j|^2}{\sigma^2 N_{ts} \|\mathbf{v}_E^k\|^2}$, respectively. To select the desired transmit beam, D calculates the average SNR over all subcarriers $\beta_D^j = \frac{1}{K} \sum_{k=1}^K \beta_D^{[j,k]}$ where K is the number of subcarriers, and selects J candidate beams with high β_D^j from $\mathcal{F}_s$, and then feeds back $\{\beta_D^j\}_{j=1}^J$ and corresponding indexes to S [9]. We assume that E can obtain these feedback information by eavesdropping the feedback channel between S and D. When the beams chosen by S result in low $\{\beta_E^{[j,k]}\}_{k=1}^K$, E is difficult to decode the communication information transmitted by S. For E, the ideal beam should provide high SNR for both E and D because the beam with low β_D^j will consume more forwarding power of E. As a result, E determines the index of the desired beam according to

$$j^* = \arg \max_{j \leq J} \beta_E^j + \delta \left(\beta_D^j - \max_j \beta_D^j \right), \quad (1)$$

where δ is a discounted factor used to balance the beam misleading success rate and power consumption of E.

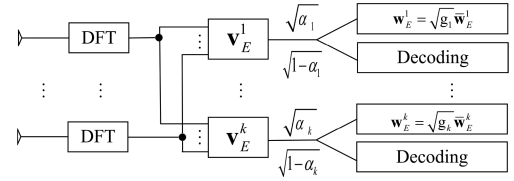


Fig. 2. Transceiver structure of E.

As shown in Fig. 1(b), after determining the desired beam j^* , E will induce S to choose the beam j^* during the following BSP periods (see Section III-A and Section IV-A). Once successful misleading, in the DT phase (Fig. 1(c)), S transmits information to D via beam j^* , and E continues to act as a spoofing AF relay, which eavesdrops while forwarding data.

In Fig. 2, $\mathbf{v}_E^k$, $\mathbf{w}_E^k$, α_k and g_k denote the combining vector, beamforming vector, power splitting ratio and power gain factor of the monitor E on the subcarrier k , respectively. In the beam misleading phase, the received signals are amplified and transmitted with $\alpha_k = 1$. In the phase of data forwarding, the received signal power is divided into two parts for decoding and forwarding. In the following sections, we will analyze how to optimize $\{\mathbf{v}_E^k, \mathbf{w}_E^k, \alpha_k, g_k\}_{k=1}^K$ to achieve the maximum monitoring rate.

III. BEAM MISLEADING AND DATA EAVESDROPPING

A. Beam Misleading

In this subsection, we will study the beam misleading policy which will induce S to choose $\mathbf{f}_{j^*}$ as the beamforming vector. As shown in Fig. 1(b), during the BSP period of S, E keeps silent unless S sweeps the j^* th beam. E will amplify and forward the received reference signals used to measure the beam quality of the j^* th beam. We assume that the latency of the AF relay is much less than the symbol duration so that it can be neglected [9]. Due to the full-duplex property of E, the received signal of E at subcarrier k is

$$\mathbf{y}_E^{[j^*,k]} = \sqrt{\frac{P_S^k}{N_{ts}}} \mathbf{H}_{SE}^k \mathbf{f}_{j^*} s_k + \mathbf{H}_{EE}^k \tilde{\mathbf{x}}_E^k + \mathbf{n}_E^k, \quad (2)$$

where $\mathbf{H}_{EE}^k \in \mathbb{C}^{N_{re} \times N_{te}}$ is the self-interference channel, and $\tilde{\mathbf{x}}_E^k = \mathbf{W}_k \tilde{\mathbf{y}}_E^k$ in which $\mathbf{W}_k \in \mathbb{C}^{N_{te} \times N_{re}}$ is the precoding matrix of E and $\tilde{\mathbf{y}}_E^k$ is the received signal at the last moment. From (2), we can see the constraint $\mathbf{H}_{EE}^k \tilde{\mathbf{x}}_E^k = 0$ will be met if $\mathbf{W}_k$ is in the null space of $\mathbf{H}_{EE}^k$, i.e., $\mathbf{H}_{EE}^k \mathbf{W}_k = 0$. Letting $\bar{\mathbf{V}}_k \in \mathbb{C}^{N_{te} \times r_0}$ ($r_0 = N_{te} - \text{rank}(\mathbf{H}_{EE}^k)$) be the right singular matrix corresponding to zero singular value of $\mathbf{H}_{EE}^k$, the precoding matrix can be given by

$$\mathbf{W}_k = \bar{\mathbf{V}}_k \tilde{\mathbf{W}}_k, \quad (3)$$

where $\tilde{\mathbf{W}}_k \in \mathbb{C}^{r_0 \times N_{re}}$ is the matrix to be optimized. To guarantee $r_0 > 0$, we have $N_{te} > N_{re}$, i.e., E needs more transmit antennas than receive ones to suppress the self-interference. After eliminating the self-interference, the transmitted signal $\mathbf{x}_E^k$ can be expressed as $\mathbf{x}_E^k = \sqrt{\frac{P_S^k}{N_{ts}}} \bar{\mathbf{V}}_k \tilde{\mathbf{W}}_k \mathbf{H}_{SE}^k \mathbf{f}_{j^*} s_k + \bar{\mathbf{V}}_k \tilde{\mathbf{W}}_k \mathbf{n}_E^k$, and the transmit power of E is

$P_E^k = \mathbb{E}[\|\mathbf{x}_E^k\|^2] = \frac{P_S^k}{N_{ts}} \|\tilde{\mathbf{W}}_k \mathbf{H}_{SE}^k \mathbf{f}_{j^*}\|^2 + \sigma^2 \|\tilde{\mathbf{W}}_k\|^2$. The received signal of D from S and E after receive combining can be expressed as

$$\mathbf{y}_D^{[j^*,k]} = (\mathbf{v}_D)^H \left(\sqrt{\frac{P_S^k}{N_{ts}}} \mathbf{H}_{SD}^k \mathbf{f}_{j^*} s_k + \mathbf{H}_{ED}^k \mathbf{x}_E^k + \mathbf{n}_D^k \right)$$

$$= \sqrt{\frac{P_S^k}{N_{ts}}} h_{SD}^{[j^*,k]} s_k + \sqrt{\frac{P_S^k}{N_{ts}}} \mathbf{h}_{ED}^k \widetilde{\mathbf{W}}_k \mathbf{h}_{SE}^{[j^*,k]} s_k + \mathbf{h}_{ED}^k \widetilde{\mathbf{W}}_k \mathbf{n}_E^k + (\mathbf{v}_D)^H \mathbf{n}_D^k, \quad (4)$$

where $\mathbf{H}_{ED}^k \in \mathbb{C}^{N_{rd} \times N_{te}}$ is the channel matrix between E and D on the k th subcarrier, $h_{SD}^{[j^*,k]} = (\mathbf{v}_D)^H \mathbf{H}_{SD}^k \mathbf{f}_{j^*}^k$, $\mathbf{h}_{ED}^k = (\mathbf{H}_{ED}^k \bar{\mathbf{V}}_k)^H \mathbf{v}_D$ and $\mathbf{h}_{SE}^{[j^*,k]} = \mathbf{H}_{SE}^k \mathbf{f}_{j^*}^k$. Then the received SNR of D on the k th subcarrier can be expressed as

$$\beta_k^{j^*} = \frac{P_S^k}{\sigma^2 N_{ts}} \frac{|h_{SD}^{[j^*,k]} + (\mathbf{h}_{ED}^k)^H \widetilde{\mathbf{W}}_k \mathbf{h}_{SE}^{[j^*,k]}|^2}{N_{rd} + \|(\mathbf{h}_{ED}^k)^H \widetilde{\mathbf{W}}_k\|^2}. \quad (5)$$

The average SNR on the beam j^* is $\beta^{j^*} = \frac{1}{K} \sum_{k=1}^K \beta_k^{j^*}$. To make S and D choose beam j^* , we need to improve β^{j^*} to make it higher than β_D , where $\beta_D = \max_j \beta_D^j$. Meanwhile, we expect to achieve beam misleading with the minimum transmit power. That is,

$$\min_{\{\widetilde{\mathbf{W}}_k\}_{k=1}^K} P_E \quad \text{s.t. } \beta^{j^*} \geq \beta_D, \quad (6)$$

where $P_E = \sum_{k=1}^K P_E^k$. To obtain the closed-form expression of $\widetilde{\mathbf{W}}_k$, we decompose the problem (6) into K independent subproblem by narrowing the constraint to $\beta_k^{j^*} \geq \beta_D^k$ for $k = 1, \dots, K$, where $\frac{1}{K} \sum_{k=1}^K \beta_D^k = \beta_D$. Consequently, the subproblems are given by

$$\min_{\widetilde{\mathbf{W}}_k} P_E^k \quad \text{s.t. } \beta_k^{j^*} \geq \beta_D^k, k = 1, \dots, K. \quad (7)$$

To solve (7), we first provide the following lemma.

Lemma 1: The optimal solution to the problem (7) can be expressed as

$$\widetilde{\mathbf{W}}_k^* = \sqrt{w_k} e^{i\theta_k^{j^*}} \bar{\mathbf{h}}_{ED}^k \left(\bar{\mathbf{h}}_{SE}^{[j^*,k]} \right)^H, \quad (8)$$

where $w_k = \|\widetilde{\mathbf{W}}_k^*\|$, $\theta_k^{j^*} = \angle h_{SD}^{[j^*,k]}$, $\bar{\mathbf{h}}_{ED}^k = \mathbf{h}_{ED}^k / \|\mathbf{h}_{ED}^k\|$, $\bar{\mathbf{h}}_{SE}^{[j^*,k]} = \mathbf{h}_{SE}^{[j^*,k]} / \|\mathbf{h}_{SE}^{[j^*,k]}\|$.

Proof: The proof is given in the Appendix.

Substituting (8) into P_E^k and $\beta_k^{j^*}$ in (7), we can see that (7) is a convex problem of w_k . Therefore, the unique unknown variable w_k in (8) can be obtained by solving (7) with CVX toolbox. Since Lemma holds for any given $\{\beta_D^k\}_{k=1}^K$ with $\frac{1}{K} \sum_{k=1}^K \beta_D^k = \beta_D$, the optimal solution to (6) has the same form as (8). To solve (6), we can substitute (8) into P_E and β^{j^*} in (6), and then calculate the optimal $\{w_k\}_{k=1}^K$. However, β^{j^*} in (6) is not a convex function of $\{w_k\}_{k=1}^K$. Therefore, it is difficult to obtain the optimal solution of (6). For given $\{\beta_D^k\}_{k=1}^K$, we can get the suboptimal solution to (6) by solving (7), and the optimal value of (7) is an upper bound of the optimal value of (6).

The monitor E can adopt the precoding matrices in (8) only when E can know all channels $\{h_{SD}^{[j^*,k]}, \mathbf{h}_{ED}^k, \mathbf{h}_{SE}^{[j^*,k]}\}_{k=1}^K$. We assume E can obtain the equivalent channel vectors $\{\mathbf{h}_{ED}^k, \mathbf{h}_{SE}^{[j^*,k]}\}_{k=1}^K$ by overhearing the pilot signals. However, it is difficult to obtain $\{h_{SD}^{[j^*,k]}\}_{k=1}^K$ due to the non-cooperative relationship between S and E. As a result, we can adjust $\{\widetilde{\mathbf{W}}_k\}_{k=1}^K$ to minimize P_E based on the feedback β^{j^*} and β_D with DRL. Note that the derived expression in (8) provides important basis for simplifying the action space of MADDPG in Section IV-A. With (8), the action of each subcarrier is transformed from $\widetilde{\mathbf{W}}_k$ to two variables. As stated in Section IV-A, in the learning framework of MADDPG, $\{w_k, \theta_k^{j^*}\}_{k=1}^K$ are determined in real time to mislead D to select the j^* th beam. Based on (8), $\mathbf{W}_k$ in (3) can be expressed as a product of column vector $\mathbf{w}_E^k = \sqrt{w_k} e^{i\theta_k^{j^*}} \bar{\mathbf{V}}_k \bar{\mathbf{h}}_{ED}^k$ and row vector $\mathbf{v}_E^k = (\bar{\mathbf{h}}_{SE}^{[j^*,k]})^H$, which is shown in Fig. 2.

B. Data Eavesdropping

After the successful beam misleading, S will select beam j^* to send information to D. The superscript j^* is omitted for simplicity in this subsection. Successful beam misleading does not mean that monitoring can be carried out successfully. In the phase of data transmission, if E does not forward data to D, the bit error rate of D may be higher than the threshold and trigger the beam recovery process so that the current beam may be switched by S [1]. Therefore, to achieve simultaneous data relaying and eavesdropping at E, the received signal $\mathbf{y}_E^k$ is split into two parts: one is forwarded to D, and the other is decoded to realize eavesdropping. Due to the introduction of α_k , the power gain factor w_k in $\mathbf{w}_E^k$ needs to be re-optimized. Define $\bar{\mathbf{w}}_E^k = \frac{\mathbf{w}_E^k}{\sqrt{w_k}}$ as the normalized beamforming vector. The transmit signal $\mathbf{x}_E^k$ of E is expressed as [9]

$$\begin{aligned} \mathbf{x}_E^k &= \sqrt{g_k \alpha_k} \bar{\mathbf{w}}_E^k \mathbf{v}_E^k \mathbf{y}_E^k \\ &= \sqrt{\frac{g_k \alpha_k P_S^k}{N_{ts}}} \bar{\mathbf{w}}_E^k \mathbf{v}_E^k \mathbf{h}_{SE}^k s_k + \sqrt{g_k \alpha_k} \bar{\mathbf{w}}_E^k \mathbf{v}_E^k \mathbf{n}_E^k, \end{aligned} \quad (9)$$

where g_k is the power gain factor for controlling transmit power. Note that $\bar{\mathbf{w}}_E^k$ and $\mathbf{v}_E^k$ are consistent with those in beam misleading phase since in both of phases, we aim to improve the SNR of D. Similar to (4), the received signal of D in data transmission phase can be written as

$$\begin{aligned} y_D^k &= \sqrt{\frac{P_S^k}{N_{ts}}} h_{SD}^k s_k + \sqrt{\frac{g_k \alpha_k P_S^k}{N_{ts}}} \mathbf{h}_{ED}^k \bar{\mathbf{w}}_E^k \mathbf{v}_E^k \mathbf{h}_{SE}^k s_k \\ &\quad + \sqrt{g_k \alpha_k} \mathbf{h}_{ED}^k \bar{\mathbf{w}}_E^k \mathbf{v}_E^k \mathbf{n}_E^k + (\mathbf{v}_D)^H \mathbf{n}_D^k. \end{aligned} \quad (10)$$

For given $\bar{\mathbf{w}}_E^k$ and $\mathbf{v}_E^k$, the received SNR of D and E can be calculated as $\gamma_D^k = \frac{P_S^k}{\sigma^2 N_{ts}} \frac{|h_{SD}^k + \sqrt{g_k \alpha_k} e^{i\angle h_{SD}^k} \|\mathbf{h}_{ED}^k\| \|\mathbf{h}_{SE}^k\|}{N_{rd} + g_k \alpha_k \|\mathbf{h}_{ED}^k\|^2}$, and $\gamma_E^k = \frac{P_S^k (1 - \alpha_k)}{\sigma^2 N_{ts}} \|\mathbf{h}_{SE}^k\|^2$. The objective of E is to optimize $\{g_k, \alpha_k\}_{k=1}^K$ so that the eavesdropping rate is maximized under the constraint of transmit power. The optimization problem is given by

$$\max_{\{g_k, \alpha_k\}_{k=1}^K} R_D \quad \text{s.t. } R_E \geq R_D, 0 \leq \{\alpha_k\}_{k=1}^K \leq 1, P_E \leq P_M, \quad (11)$$

where $R_D = \frac{1}{K} \sum_{k=1}^K \log(1 + \gamma_D^k)$, $R_E = \frac{1}{K} \sum_{k=1}^K \log(1 + \gamma_E^k)$, $P_E = \sum_{k=1}^K \left(\frac{g_k \alpha_k P_S^k}{N_{ts}} \|\mathbf{h}_{SE}^k\|^2 + \sigma^2 g_k \alpha_k \right)$, and P_M is the total transmit power of E. We assume that E can realize monitoring only when $R_E \geq R_D$, and the corresponding monitoring rate is R_D [5]. The solution of (11) can be derived by the Lagrangian multiplier method if E knows global CSI. However, we can not obtain the optimal $\{g_k, \alpha_k\}_{k=1}^K$ when h_{SD}^k is unknown. A more reasonable assumption than knowing h_{SD}^k is that γ_D^k can be obtained by monitoring the feedback channel between S and D. Therefore, DRL is adopted to determine $\{g_k, \alpha_k\}_{k=1}^K$ by interacting with the system and regarding $\{\gamma_D^k\}_{k=1}^K$ as corresponding rewards. By training neural networks with MADDPG, $\{g_k, \alpha_k\}_{k=1}^K$ are given in real time to improve the eavesdropping rate of E under controllable transmit power.

IV. BEAM MISLEADING AND DATA EAVESDROPPING BASED ON MADDPG

Based on the analysis in Section III, when the CSI between S and D is unknown, the beam misleading and data eavesdropping problems are formulated as Markov decision process (MDP) problems. A straightforward approach is to regard all subcarriers as one agent and obtain the policy by the single-agent DRL, such as DDPG. Considering training

a policy with large action space is typically more difficult than training several policies with small action space in practical implementations, in both phases, we adopt the learning architecture of MADDPG [10], which includes K individual actors (policies) and one centralized critic (value function). In MADDPG, all subcarriers are regarded as individual agents, which cooperate to achieve a common goal. In the training phase of MADDPG, the actors and critic are updated using the global data, including global state, shared reward, and the joint action, which will be defined later.

A. MADDPG-Based Beam Misleading

In this subsection, the beam misleading problem in (6) is formulated as a multi-agent cooperative MDP problem. We will first give the definition of action, state, and reward, and then present the architecture of MADDPG.

Action: Through the analysis in sec.III-A, searching for the optimal $\bar{\mathbf{W}}_k$ is transformed into finding the optimal w_k and θ_k^{j*} , which will speed up the training process. At time t , the action for the k th subcarrier is denoted by $a_t^k = (w_k, \theta_k^{j*})$. Therefore, the joint action of all subcarriers is $\mathbf{a}_t = (a_t^1, \dots, a_t^K)$.

State: At the time step t , the local state for subcarrier k is $s_t^k = ((\mathbf{h}_{SE}^{[j*,k]})^H, (\mathbf{h}_{ED}^k)^H, \beta^{j*}, \beta_D)$, where β^{j*} and β_D are obtained by eavesdropping and analyzing beam reports on the feedback channel [11]. We set $\beta^{j*} = 0$ if the beam report does not contain the j^* th beam. The global state s_t is the concatenation of non-overlapped information of all local states s_t^k , i.e., $s_t = ((\{\mathbf{h}_{SE}^{[j*,k]}\}_{k=1}^K)^H, (\{\mathbf{h}_{ED}^k\}_{k=1}^K)^H, \beta^{j*}, \beta_D)$.

Reward: The reward r_t is defined as $r_t = -a_1 P_E - a_2 (\beta^{j*} - \beta_D - B)^2 + a_3 I(\beta^{j*}, \beta_D)$, where $\{a_i\}_{i=1}^3$ are positive coefficients used to balance the success rate of misleading and power consumption of relay, and B is a constant used to increase the probability of selecting beam j^* . $I(x, y)$ is an indicator function where $I(x, y) = 1$ if $x \geq y$, and 0 otherwise. The reward function encourages the successful beam misleading, meanwhile, punishes the action of consuming too much power.

In MADDPG, the policy for each subcarrier k is parameterized as $\mu_{\Theta_k}^k(s^k)$ and the centralized critic is $Q_\omega(s, \mathbf{a})$, where Θ_k and ω are parameters of neural networks, respectively. Besides, the target policy $\bar{\mu}_{\bar{\Theta}_k}^k(s^k)$ and critic $\bar{Q}_{\bar{\omega}}(s, \mathbf{a})$ are adopted for stabilizing the training process. At the beginning of each episode in the beam misleading MADDPG, we initialize the environment with a random initial state s_0 and the beam misleading index j^* according to (1). At each time-step t , the action a_t^k for subcarrier k is selected using the individual policy $\mu_{\Theta_k}^k(s_t^k)$ with some exploration. When S sweeps the j^* th beam, actions $\{a_t^k\}_{k=1}^K$ are executed by forwarding the multiplication of the received signal $\mathbf{y}_E^{[j*,k]}$ and $\mathbf{W}_k$ to D. The reward r_t is obtained by utilizing the feedback, and then the system transfers to the next state s_{t+1} . Then the transition $(s_t, \mathbf{a}_t, r_t, s_{t+1})$ is stored into the memory buffer $\mathcal{D}$. During the training, the parameters ω are updated by minimizing the following temporal difference (TD) loss

$$L(\omega) = \hat{\mathbb{E}}_{(s_i, \mathbf{a}_i, r_i, s_{i+1}) \sim \mathcal{D}} \left\{ (y_i - Q_\omega(s_i, \mathbf{a}_i))^2 \right\}, \quad (12)$$

where $\hat{\mathbb{E}}_{(s_i, \mathbf{a}_i, r_i, s_{i+1}) \sim \mathcal{D}} \{\dots\}$ indicates the sample average over the transition $(s_i, \mathbf{a}_i, r_i, s_{i+1})$ uniformly sampled from $\mathcal{D}$, $y_i = r_i + \phi Q_\omega(s_{i+1}, \{\mu_{\Theta_k}^k(s_{i+1}^k)\}_{k=1}^K)$, and ϕ is the discount factor. Meanwhile, each $\mu_{\Theta_k}^k$ is updated by using gradient ascent with the following policy gradient [10]

$$\Delta \Theta_k = \hat{\mathbb{E}}_{(s_i, \mathbf{a}_i) \sim \mathcal{D}} \left\{ \nabla_{\Theta_k} Q_\omega(s_i, \mathbf{a}_i) \big|_{\mathbf{a}_k = \mu_{\Theta_k}^k(s_i^k)} \nabla_{\Theta_k} \mu_{\Theta_k}^k(s_i^k) \right\}. \quad (13)$$

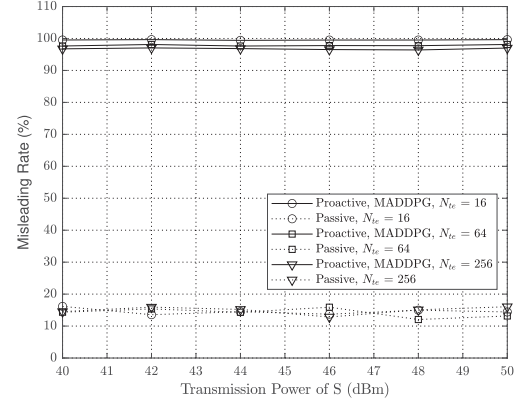


Fig. 3. Beam misleading rate vs. transmit power of S with different N_{te} .

Moreover, the target parameters $\{\bar{\Theta}_k\}_{k=1}^K$ and $\bar{\omega}$ are periodically updated from $\{\Theta_k\}_{k=1}^K$ and ω , respectively. By applying MADDPG, we can find the optimal map from s_t^k to a_t^k , so that transmission parameters of each subcarrier can be quickly determined based on the state in the practical system.

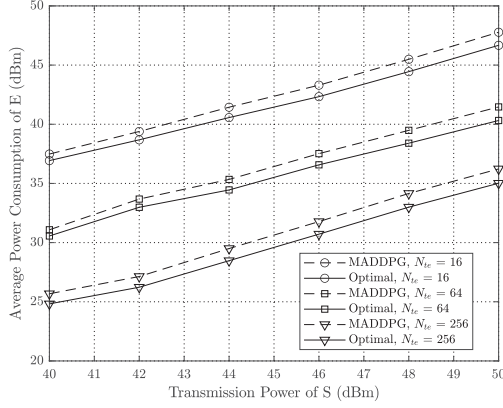
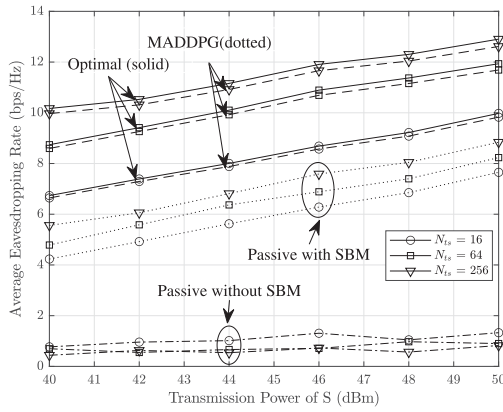
B. MADDPG-Based Data Eavesdropping

After successful beam misleading, the monitor E will forward data for beam maintenance and improving the eavesdropping rate. In this phase, agents need to optimize $\{g_k, \alpha_k\}_{k=1}^K$ to maximize the eavesdropping rate under the transmit power constraint. The action for the k th subcarrier at time t is denoted by $\hat{a}_t^k = (g_k, \alpha_k)$, and the action for all subcarriers is $\hat{\mathbf{a}}_t = (\hat{a}_t^1, \dots, \hat{a}_t^K)$. The state for the k th subcarrier is defined as $\hat{s}_t^k = ((\mathbf{h}_{SE}^k)^H, (\mathbf{h}_{ED}^k)^H, \gamma_D^k)$ and the state for the central critic is $\hat{s}_t = ((\{\mathbf{h}_{SE}^k\}_{k=1}^K)^H, (\{\mathbf{h}_{ED}^k\}_{k=1}^K)^H, \gamma_D^k)$, where $\{\gamma_D^k\}_{k=1}^K$ are obtained by monitoring the feedback channel. Then the reward $\hat{r}_t$ at time t is given by $\hat{r}_t = b_1 R_D - b_2 I(P_E, P_M) - b_3 (R_E - R_D - C)^2 + b_4 I(R_E, R_D)$, where $\{b_i\}_{i=1}^4$ are positive coefficients used to balance the eavesdropping rate and power consumption of the monitor, and C is an eavesdropping rate margin. The reward function encourages the subcarriers to cooperate to maximize R_D under the constraints of $R_E > R_D$ and $P_E \leq P_M$. The optimal map between $\hat{a}_t^k$ and $\hat{s}_t^k$ on each subcarrier can be obtained by MADDPG.

V. NUMERICAL RESULTS

In this section, we provide numerical results to evaluate performances of the proposed scheme. The scattering channel mode in [2] with 6 scatterers is adopted. The codebook $\mathcal{F}_s$ is generated by MATLAB phase array toolbox and the number of beams in $\mathcal{F}_s$ is $4\sqrt{N_{ts}}$. We set $N_{rd} = 4$, $N_{re} = 4$, $K = 4$, and $\sigma^2 = -116$ dBm. The critic and actor neural networks are based on a two-layer and one-layer Multilayer perception (MLP) with relu activation, respectively, and each hidden layer has 100 neurons. The smooth factor is 10^{-2} and the learning rates of critic and actors are 10^{-3} and 10^{-4} , respectively. The length of each episode is 500. In each episode, both $\mathbf{H}_{SE}$ and $\mathbf{H}_{EE}$ are constant, while channels $\mathbf{H}_{SD}$ and $\mathbf{H}_{ED}$ change according to the Markov model. In different episodes, the azimuth (elevation) angles of arrival and departure of each scatterer vary randomly in $[-90^\circ, 90^\circ]$ ($[-45^\circ, 45^\circ]$), and the distances between nodes vary randomly in [260 m, 300 m]. Each point in the following Figures is plotted by averaging 10^5 Monte Carlo simulations using the well-trained policy.

Fig. 3 shows, in the beam misleading phase, the success rate of beam misleading versus the transmit power of S P_S with different N_{te} .

Fig. 4. Transmit power of E vs. transmit power of S with different N_{te} .Fig. 5. The average eavesdropping rate vs. P_S with different N_{ts} .

We set $N_{ts} = 16$ and $P_S^k = \frac{P_S}{K}$. The misleading rate is calculated by counting the number of $\beta^{j*} \geq \beta_D$ in 10^5 simulations. In the passive eavesdropping, the monitor E keeps silent when S executes beam sweep. As shown by dotted line in Fig. 3, in this case, when E and D are far separated, D will select the j^* -th beam (preferred by E) with low probability. It is shown that the successful misleading rates of the proactive scheme are close to 100% under different N_{te} , which verify the effectiveness of the proposed method. Fig. 4 demonstrates the relationship between P_E and P_S with different N_{te} in the beam misleading phase of E. In the optimal scheme, $P_E = \frac{1}{K} \sum_{k=1}^K P_E^k$, where P_E^k is the optimal objective value in (7) with known $\{h_{SD}^{[j^*, k]}\}$ s. In the MADDPG-based scheme, P_E is calculated with parameters learned by MADDPG. We can see that although P_E will increase as P_S increases, more N_{te} can effectively reduce P_E . Combining Figs. 3 and 4, it can be seen that even if $\{h_{SD}^{[j^*, k]}\}_{k=1}^K$ are unknown, we can still achieve a good beam misleading using the policy learned by MADDPG, while the transmit power cost is only slightly higher than the minimum power. Fig. 5 displays the average eavesdropping rate versus P_S with different N_{ts} in the phase of data eavesdropping. We set $N_{te} = 16$ and $P_M = 50$ dBm. The optimal solution is obtained by solving (11) with known $\{h_{SD}^k\}$ s. In the passive eavesdropping, S usually transmits to D with the j -th ($j \neq j^*$) beam, which results in low channel capacity of the S-E link and thus eavesdropping difficulties, and the eavesdropping rate deteriorates with the increase of N_{ts} . The passive method with successful beam misleading (SBM) means that E executes beam misleading in the BSP phase but keeps silent in the DT phase of S. After SBM, the channel capacity of S-E link is improved greatly and

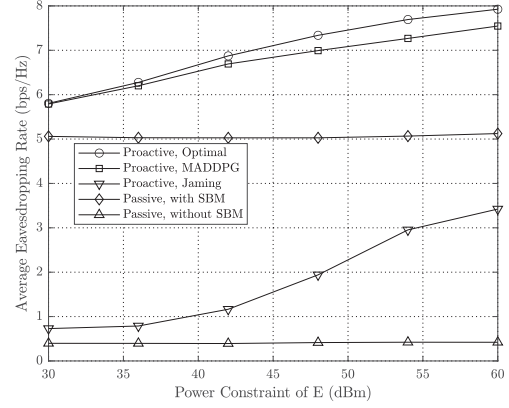


Fig. 6. Comparison of average eavesdropping rates of various eavesdropping methods.

will be larger than the R_D at most of the time. That means, after SBM, the eavesdropping rate is limited by R_D . With proper beam selection (1), R_D will not severely decrease due to beam misleading. Therefore, the passive method with SBM achieves a higher eavesdropping rate than traditional passive eavesdropping. In the proposed proactive scheme, the monitor can further raise the eavesdropping rate by improving R_D under the constraint $R_E \geq R_D$ and the maximum eavesdropping R_E can be achieved with enough power budget and reasonable power splitting. Meanwhile, we can see that the eavesdropping rate of the MADDPG-based method is very close to the optimal solution and increases with the increase of P_S and/or N_{ts} .

Fig. 6 compares the average eavesdropping rate of different eavesdropping methods with the varying power constraint P_M . The passive scheme without beam misleading performs the worst since the channel capacity of S-E link is lower than R_D most of the time. Jamming scheme in [4] is also plotted for comparison. The jamming scheme forces the rate of S below R_E by sending jamming signals, and thus improves the probability of successful monitoring. However, the performances of both passive and jamming schemes are limited by R_E . When the beam selected by S causes low R_E , the two schemes can only achieve a low eavesdropping rate. By inducing S to select the beam preferred by E, R_E can be effectively improved. As shown in Fig. 6, after SBM, passive eavesdropping in the DT phase can achieve better eavesdropping performance than the jamming scheme. In the proactive scheme, E acts as an AF relay to improve the transmission rate of S-D link so that the eavesdropping rate is further improved. It is shown that the eavesdropping rate achieved by the MADDPG scheme is close to the optimal solution. When $P_M > 55$ dBm, the eavesdropping rate of the proposed proactive scheme approaches the maximum R_E with small α^* .

VI. CONCLUSION

In massive MIMO-OFDM systems, when the transmitter transmits to the receiver via narrow beams, the eavesdropping performance is limited by the low channel capacity between the transmitter and the monitor. Our proposed monitoring scheme can effectively improve the channel capacity between transmitter and monitor by beam misleading, and thus raise the eavesdropping rate. By optimizing transmit parameters of the monitor, it can realize monitoring with a minimum power budget. Considering the monitor can hardly obtain the channel information between suspicious nodes, we propose MADDPG-based learning schemes, in which multiple subcarriers cooperate to realize

the beam misleading and maximize the eavesdropping rate by adjusting transmit parameters. Simulation results demonstrated the effectiveness and superiority of the proposed proactive eavesdropping scheme.

APPENDIX

For the sake of brevity, the superscripts of subcarrier k and beam j^* in channel notation are omitted in the following proof. We suppose that a feasible solution to (7) is given by $\mathbf{W}'_k = \sqrt{w'_k} \cdot \bar{\mathbf{W}}_k$ with $\|\bar{\mathbf{W}}_k\| = 1$, and the resulting optimal value is $P_E(\mathbf{W}'_k)$. We then construct an alternative solution $\mathbf{W}''_k = \sqrt{w''_k} e^{i\angle h_{SD}} \bar{\mathbf{h}}_{ED} \bar{\mathbf{h}}_{SE}^H$ with $w''_k = w'_k \frac{(\bar{\mathbf{h}}_{ED})^H \bar{\mathbf{W}}_k \bar{\mathbf{h}}_{SE}}{\|\bar{\mathbf{h}}_{ED}\|^2 \|\bar{\mathbf{h}}_{SE}\|^2} \leq w'_k$, where the inequality follows $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ for any matrices (vectors) $\mathbf{A}$ and $\mathbf{B}$. We will prove that $\mathbf{W}''_k$ is feasible to problem (7), and $\mathbf{W}''_k$ can produce an objective value smaller than $P_E(\mathbf{W}'_k)$. Let $\gamma(\mathbf{W}''_k) = |h_{SD} + \mathbf{h}_{ED}^H \mathbf{W}''_k \mathbf{h}_{SE}|^2$, $\delta(\mathbf{W}''_k) = \|\mathbf{h}_{ED}^H \mathbf{W}''_k\|$. By substituting $\mathbf{W}''_k$ into numerator and denominator of $\beta_k^{j^*}$ in (5), we have

$$\gamma(\mathbf{W}''_k) \geq |h_{SD} + \sqrt{w'_k} \mathbf{h}_{ED}^H \bar{\mathbf{W}}_k \mathbf{h}_{SE}|^2 = \gamma(\mathbf{W}'_k) \quad (14)$$

$$\delta(\mathbf{W}''_k) = \sqrt{w'_k} \frac{|\mathbf{h}_{ED}^H \bar{\mathbf{W}}_k \mathbf{h}_{SE}|}{\|\bar{\mathbf{h}}_{SE}\|} \leq \sqrt{w'_k} \|\mathbf{h}_{ED}^H \bar{\mathbf{W}}_k\| = \delta(\mathbf{W}'_k) \quad (15)$$

where (14) and (15) follow from the triangle inequality. Based on (14) and (15), we can infer that $\beta_k^{j^*}(\mathbf{W}''_k) \geq \beta_k^{j^*}(\mathbf{W}'_k) \geq \beta_D$. The above result demonstrates $\mathbf{W}''_k$ is feasible to problem (7). By substituting $\mathbf{W}''_k$ into the objective function of (7), we obtain

$$\begin{aligned} P(\mathbf{W}''_k) &= \frac{P_S}{N_{ts}} w'_k \frac{|\mathbf{h}_{ED}^H \bar{\mathbf{W}}_k \mathbf{h}_{SE}|^2}{\|\bar{\mathbf{h}}_{ED}\|^2} + w'_k \sigma^2 \\ &\leq \frac{P_S}{N_{ts}} w'_k \|\bar{\mathbf{W}}_k \mathbf{h}_{SE}\|^2 + w'_k \sigma^2 = P(\mathbf{W}'_k), \end{aligned} \quad (16)$$

where (16) follows from Cauchy-Schwarz inequality, i.e., $|\mathbf{a}^H \mathbf{b}|^2 \leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2, \forall \mathbf{a}, \mathbf{b}$. In summary, for any solution $\mathbf{W}'_k$ to the problem

(7), we can always construct another feasible solution in the form of $\mathbf{W}''_k = \sqrt{w'_k} e^{i\angle h_{SD}} \bar{\mathbf{h}}_{ED} \bar{\mathbf{h}}_{SE}^H$ that achieves smaller objective value. This completes the proof.

REFERENCES

- [1] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. Hoboken, NJ, USA: Academic Press, 2020.
- [2] O. El 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, Mar. 2014.
- [3] J. Xu, L. Duan, and R. Zhang, "Proactive eavesdropping via cognitive jamming in fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2790–2806, May 2017.
- [4] H. Cai, Q. Zhang, Q. Li, and J. Qin, "Proactive monitoring via jamming for rate maximization over MIMO rayleigh fading channels," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2021–2024, Sep. 2017.
- [5] Y. Zeng and R. Zhang, "Wireless information surveillance via proactive eavesdropping with spoofing relay," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1449–1461, Dec. 2016.
- [6] A. Bereyhi, S. Asaad, R. R. Muller, R. F. Schaefer, and A. M. Rabiei, "On robustness of massive MIMO systems against passive eavesdropping under antenna selection," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*. IEEE, 2018, pp. 1–7.
- [7] G. Weisz, P. Budzianowski, P.-H. Su, and M. Gašić, "Sample efficient deep reinforcement learning for dialogue systems with large action spaces," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 11, pp. 2083–2097, Nov. 2018.
- [8] F. Sahrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.
- [9] *NR; Physical Layer Procedures for Control*, 3rd Generation Partnership Project (3GPP), 3GPP: TS 38.213, 2020.
- [10] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. Advances Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6379–6390.
- [11] *NR; Physical Layer Measurements*, 3rd Generation Partnership Project (3GPP), 3GPP: TS 38.215, 2020.