Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

# Secure resource allocation against colluding eavesdropping in a user-centric cell-free massive multiple-input multiple-output system*#

Na LI†, Yuanyuan GAO†‡, Kui XU, Xiaochen XIA, Huazhi HU, Yang LI, Yueyue ZHANG

*College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China*

†E-mail: linalala@yeah.net; njyygao@sina.com

Received Nov. 26, 2022; Revision accepted Aug. 9, 2023; Crosschecked Mar. 6, 2024

**Abstract:** We investigate the resource allocation problem of a cell-free massive multiple-input multiple-output system under the condition of colluding eavesdropping by multiple passive eavesdroppers. To address the problem of limited pilot resources, a scheme is proposed to allocate the pilot with the minimum pollution to users based on access point selection and optimize the pilot transmission power to improve the accuracy of channel estimation. Aiming at the secure transmission problem under a colluding eavesdropping environment by multiple passive eavesdroppers, based on the local partial zero-forcing precoding scheme, a transmission power optimization scheme is formulated to maximize the system's minimum security spectral efficiency. Simulation results show that the proposed scheme can effectively reduce channel estimation error and improve system security.

**Key words:** Cell-free; Colluding eavesdropping; Limited pilot resources; Access point selection; Security rate
https://doi.org/10.1631/FITEE.2200599 **CLC number:** TN929.5

## 1 Introduction

A cell-free (CF) massive multiple-input multiple-output (MIMO) system is proposed based on a traditional centralized massive MIMO system. Although centralized massive MIMO systems can significantly improve network transmission performance (Larsson et al., 2014; Björnson et al., 2017), there are also problems such as poor communication quality of edge users, large system energy consumption, high deployment cost, and poor network flexibility (Ngo et al., 2015). By deploying centralized antennas within the network coverage to form a large number of distributed access points (APs), the CF massive MIMO system can alleviate the aforementioned problems (Timilsina et al., 2018; Interdonato et al., 2019a). Not only can the CF massive MIMO system inherit the performance advantages of the centralized massive MIMO system (Bashar et al., 2018), but the user-centric architecture that is formed by dynamically selecting part of the APs can avoid frequent cross-area switching, reduce the cost of wireless resources, and achieve a large range of continuous coverage and on-demand allocation (Huang and Burr, 2017). Moreover, the space time characteristics of the channel in the distributed architecture reduce pilot pollution caused by pilot multiplexing by using the service matching strategy of AP clustering and user grouping, and design the power allocation scheme

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 62071485, 61671472, and 62271503) and the Natural Science Foundation of Jiangsu Province, China (Nos. 20201334 and 20181335)

# Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2200599) contains supplementary materials, which are available to authorized users

ORCID: Na LI, https://orcid.org/0000-0002-3801-274X; Yuanyuan GAO, https://orcid.org/0000-0002-0839-0248

flexibly to further improve the system's spectral efficiency (Buzzi and D'Andrea, 2017a).

The distributed antenna architecture of CF massive MIMO systems multiplies the number of wireless links in the network and makes the network space more open, providing a huge attack entrance for large-scale malicious terminals. In addition, APs are deployed in a certain area in an ultra-dense manner, such that the distance between each AP and the eavesdropper is closer, which improves the eavesdropping capability, leading to a greater threat to the transmission security. With the development of the CF massive MIMO technology and demand for hot high-capacity scenarios, it is a general trend for CF MIMO systems to serve more and more users. With the exponential growth of users, pilot resources are limited and pilot reuse is inevitable. However, in contrast to the CF massive MIMO system where users are served by all APs, the user-centric CF massive MIMO system allocates a service AP set for each user, and each user in the network is served by part of the APs, which facilitates each user to be located in the effective center of the service AP set. Consequently, pilot multiplexing is transformed from the cell-centric to the user-centric mode. The allocation of the pilot frequency must combine the service requirements of users and the construction of dynamic cooperative clusters. Ensuring secure transmission by a system with limited pilot resources is an urgent problem that should be solved.

### 1.1 Previous works

The main difference between the user-centric CF massive MIMO network and traditional CF massive MIMO network lies in the construction of the service APs; a practical user service framework is established according to offline or online information based on different requirements of the user, which can simplify signal processing and reduce the computational complexity, fronthaul load, and channel estimation overhead. The strategies for constructing service clusters can be divided into two categories: (1) constructing service APs for users according to their different needs; (2) optimizing the service APs by user scheduling or power control (Ammar et al., 2022a, 2022b). Buzzi and D'Andrea (2017a) proposed that the AP with the strongest channel was selected to serve the users based on channel estimation information. Ngo et al. (2018) proposed AP selection schemes based on the received power and large-scale fading. Buzzi and D'Andrea (2017b) proposed that each AP serves users whose channel strength was greater than the average channel strength. However, the above-mentioned AP selection algorithms cannot guarantee the optimal performance of the system. Therefore, an AP selection scheme based on sum-rate maximization was proposed by Boroujerdi et al. (2017). Power allocation and AP selection were optimized jointly to maximize the system's energy efficiency (Vu et al., 2020). van Chien et al. (2020) designed a downlink power minimization problem and selected the service APs according to the optimization result based on a threshold. Dong et al. (2019) solved the energy efficiency maximization problem of joint power allocation, user scheduling, and antenna selection using alternating optimization and a successive convex approximation algorithm. The optimal service AP cluster was selected (Mendoza et al., 2020) by optimization based on the fronthaul capacity and quality of service.

At present, there are some mature algorithms for eliminating the influence of non-orthogonal pilots on transmission. Based on a random allocation algorithm (Interdonato et al., 2019b), a greedy allocation algorithm that considers the minimum downlink rate has been proposed previously (Ngo et al., 2017). A tabu search method for pilot allocation (Liu et al., 2020a) and an interference graph determination based on a user-centric AP selection scheme and pilot allocation by graph coloring (Liu et al., 2020b) have been reported. Buzzi et al. (2021) first grouped users according to the number of pilots, and then allocated pilots using the Hungarian algorithm. The distance between users of the same pilot has been minimized by using $K$-means clustering (Attarifar et al., 2018). Continuously selecting the user with the minimum interference to complete the scalable pilot allocation based on assigning pilots to the main AP service users has been attempted (Ngo et al., 2017). Chen et al. (2021) assigned the same pilot to spatially separated users to suppress pilot contamination. To solve the problem that the above methods do not have an upper limit of APs for users, Sarker and Fapojuwo (2021) first grouped users and APs, and then minimized the number of common APs of users with the same pilot to improve the pilot pollution problem. Few studies have been conducted on passive eavesdropping in CF massive MIMO

systems. Xia et al. (2023) proposed an optimization algorithm based on successive convex approximation under the constraint of uplink and downlink transmission power, which improves the security spectral efficiency of non-colluding passive eavesdropping and colluding passive eavesdropping.

### 1.2 Motivation

Some of the previous pilot allocation methods have inherent defects. For example, the random allocation algorithm has a poor ability to suppress pilot contamination, and the performances of the tabu search and $K$-means clustering algorithms depend on the initial pilot allocation. The pilot allocation method based on the minimum interference of the main AP (Ngo et al., 2017) and the pilot method based on the same pilot allocated to spatially separated users (Chen et al., 2021) do not provide upper limits for user service clusters. The method of minimizing the number of public APs (Sarker and Fapojuwo, 2021) is difficult to determine the initial user group in the dense deployment scenario. In addition, only the method proposed in Ngo et al. (2017) is scalable, and other methods have limitations in the application of large-scale access scenarios. Moreover, multiple eavesdroppers who can monitor the channel individually (non-colluding eavesdroppers) or share their eavesdropper information (colluding eavesdroppers) pose a more serious threat to wireless communication security (Yu et al., 2020; Choi and Park, 2021), making the attack more effective (Mirmohseni and Papadimitratos, 2014). From the perspective of security, colluding eavesdroppers pose a greater threat to system security, and thus it is more valuable to study security under colluding eavesdroppers. If there are multiple eavesdroppers in a network with limited pilot resources, security design would be more challenging.

### 1.3 Contributions

Aiming at the shortcomings in the application of the above pilot allocation methods in CF massive MIMO systems and the secure transmission problem under colluding eavesdropping, in this study, we propose a scalable joint AP selection and pilot allocation scheme, and ensure secure transmission by the system by optimizing the transmission power. The contributions of this study are as follows:

1. To address the inevitable pilot reuse problem, a scalable joint AP selection and pilot allocation scheme is proposed. Based on the AP selection, the pilot with the minimum pollution is selected for access users, and the transmission power of the pilot is optimized to improve the accuracy of channel estimation. The complexity of the optimization algorithm does not increase with an increase in the number of users; therefore, the proposed method is scalable.

2. In this study, the security problem of multi-eavesdropper colluding eavesdropping in a user-centric CF massive MIMO system is studied for the first time. To address the secure transmission problem, a secure transmission method based on partial zero-forcing (ZF) precoding is proposed. The closed-form expression of security spectral efficiency under local partial ZF is derived, and the maximization problem of security spectral efficiency based on power allocation is established and solved by stepwise approximation.

Notations: Bold indicates the vectors or matrices. $x\sim\mathcal{CN}(a,b)$ indicates that $x$ is a complex Gaussian random variable with a mean of $a$ and variance of $b$. $(\cdot)^{\mathrm{H}}$ represents the conjugate transpose. $\|\cdot\|$, $|\cdot|$, and $\mathbb{E}\{\cdot\}$ represent the Euclidean norm, modulus, and mathematical expectation, respectively. $\boldsymbol{I}_N$ represents the $N$-dimensional unit matrix.

## 2 System model

In this study, we consider a CF massive MIMO scenario with colluding eavesdropping by multiple eavesdroppers, where $M$ APs (equipped with multi-antenna, i.e., $N$ antennas), $K$ users (equipped with single-antenna), and $N_{\mathrm{e}}$ passive eavesdroppers (equipped with single-antenna) are randomly distributed within the coverage range, as shown in Fig. 1. The channel between AP $m$ and user $k$ is modeled by large- and small-scale fading as follows:

$$\boldsymbol{g}_{m,k}=\beta_{m,k}\boldsymbol{h}_{m,k}, \tag{1}$$

where $\beta_{m,k}$ represents large-scale fading between AP $m$ and user $k$, which changes slowly depending on whether the user moves, and $\boldsymbol{h}_{m,k}\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{I}_N)$ is an independent Rayleigh fading channel. The eavesdropper channel is modeled in the same manner as that for a legitimate user. The channel in the coherence interval is constant, the total transmission

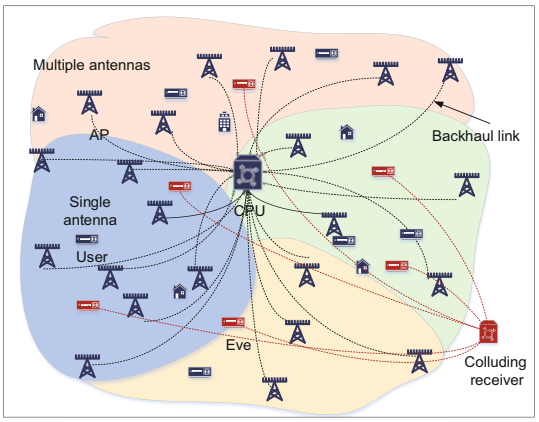


**Fig. 1 System model diagram (AP: access point; CPU: central processing unit)**

is in the time division duplex (TDD) mode, and the uplink and downlink channels have reciprocity.

## 3 Uplink channel estimation under pilot constraints

We consider the channel estimation of non-orthogonal pilots with pilot length $\tau_{\mathrm{p}} < K$ in the TDD mode. In the uplink channel estimation phase, the user sends a pilot signal $\boldsymbol{S} = [\boldsymbol{s}_1; \boldsymbol{s}_2; \ldots; \boldsymbol{s}_K] \in \mathbb{C}^{K\times\tau_{\mathrm{p}}}$ to the AP to estimate the channel, and the pilot signal received by AP $m$ is as follows:

$$\boldsymbol{y}_m = \sum_{k=1}^{K} \sqrt{\tau_{\mathrm{p}} P_{\tau_k}} \boldsymbol{g}_{m,k} \boldsymbol{s}_k + \boldsymbol{N}_m, \tag{2}$$

where $P_{\tau_k}$ represents the pilot power of user $k$, $\|\boldsymbol{s}_k\|^2 = 1$, and $\boldsymbol{N}_m \in \mathbb{C}^{1\times\tau_{\mathrm{p}}}$ is the noise vector whose elements satisfy the complex Gaussian distribution with independent identically distribution $N_{m,k} \sim \mathcal{CN}(0, \sigma^2)$ with $\sigma^2$ the noise power. The estimated channel by the minimum mean square error (MMSE) is

$$\begin{aligned} \hat{\boldsymbol{g}}_{m,k}^{\mathrm{MMSE}} &= \frac{\hat{\boldsymbol{y}}_{m,k}^{\mathrm{H}} \boldsymbol{g}_{m,k}}{|\hat{\boldsymbol{y}}_{m,k}|^2} \hat{\boldsymbol{y}}_{m,k} \\ &= \sqrt{\tau_{\mathrm{p}} P_{\tau_k}} \frac{\beta_{m,k}}{\sum\limits_{j\in\Phi_{\tau_{\mathrm{p}}}} \tau_{\mathrm{p}} P_{\tau_k} \beta_{m,j} + \sigma^2} \hat{\boldsymbol{y}}_{m,k}, \end{aligned} \tag{3}$$

where

$$\begin{aligned} \hat{\boldsymbol{y}}_{m,k} =& \sqrt{\tau_{\mathrm{p}} P_{\tau_k}} \boldsymbol{g}_{m,k} + \sum_{j\in\Phi_{\tau_{\mathrm{p}}},\, j\neq k} \sqrt{\tau_{\mathrm{p}} P_{\tau_j}} \boldsymbol{g}_{m,j} \\ &+ N_{m,k} \boldsymbol{s}_k^{\mathrm{H}} \end{aligned} \tag{4}$$

is the pilot-related signal of user $k$ received by AP $m$, and $\Phi_{\tau_{\mathrm{p}}}$ represents the user set with respect to $\tau_{\mathrm{p}}$. $\hat{\boldsymbol{g}}_{m,k} \sim \mathcal{CN}(\boldsymbol{0}, r_{m,k}\boldsymbol{I}_N)$ is the estimated channel of $\boldsymbol{g}_{m,k}$, where the mean square value of the estimated channel is

$$r_{m,k} = |\hat{\boldsymbol{g}}_{m,k}|^2 = \frac{\tau_{\mathrm{p}} P_{\tau_k} \beta_{m,k}^2}{\sum\limits_{j\in\Phi_{\tau_{\mathrm{p}}}} \tau_{\mathrm{p}} P_{\tau_j} \beta_{m,j} + \sigma^2}. \tag{5}$$

The channel estimation error is

$$\tilde{\boldsymbol{g}}_{m,k} \sim \mathcal{CN}\left(\boldsymbol{0}, (\beta_{m,k} - r_{m,k})\boldsymbol{I}_N\right).$$

In this study, a scalable method combining AP selection and pilot allocation is proposed for pilot reuse and AP selection under pilot constraints. The proposed pilot allocation and AP selection method solves the following three problems:

1. In a CF massive MIMO system, users far away from the APs have a poor quality of service. Therefore, the APs give up serving such users during transmission, thus forming a user-centric CF massive MIMO system.

2. In the pilot allocation of CF massive MIMO systems, if only the impact of pilot pollution is considered and the size of the service APs is not limited, system performance will be degraded (Chen et al., 2021).

3. The channel estimation performance is improved to satisfy the scalability of the system.

According to Eq. (5), the channel estimation error of user $k$ originates from the users who share the same pilot frequency with user $k$, i.e.,

$$I_{m,k} = \sum_{j\in\Phi_{\tau_{\mathrm{p}}},\, j\neq k} \tau_{\mathrm{p}} P_{\tau_j} \beta_{m,j} + \sigma^2. \tag{6}$$

In a user-centric massive MIMO system, a user is served by a part of APs, and the design of the precoding vector in the downlink transmission is related only to the estimated channel on those APs; only the pilot pollution on those APs affects the transmission, and this part of pilot pollution is called effective pilot pollution. Therefore, pilot pollution can be controlled by the user's service APs; i.e., the impact of pilot reuse can be reduced through AP selection.

Because the development of location-based services has made it possible to obtain the exact location of a user, a central processing unit (CPU) can obtain the location information of a legitimate user. The influence of non-orthogonal pilot channel estimation originates from the contamination of the same AP in channel estimation by users sharing the

same pilot. To reduce this impact, it is necessary to reduce the number of intersections on the service AP between users who share the same pilot. In addition, according to a previous report (Alonzo et al., 2019), using APs close to users to provide services can reduce the impact of path loss on transmission and improve transmission efficiency. Therefore, APs close to each user requesting access were assigned according to the distance between APs and users. When the number of access users is smaller than the number of orthogonal pilots, an orthogonal pilot is assigned to the user. When the number of access users is greater than the number of orthogonal pilots, the pilot with the minimum effective pilot pollution is assigned to the user. In this study, we propose a method of joint pilot allocation and AP selection based on the current AP clustering situation, where the corresponding pilots are assigned to users.

To further suppress the impact of pilot contamination and improve the accuracy of channel estimation, the user's normalized channel estimation error is minimized by optimizing the user's pilot power. The normalized mean square error (NMSE) from user $k$ to AP $m$ is as follows:

$$e_{m,k}=\frac{\left|\hat{\boldsymbol{g}}_{m,k}-\boldsymbol{g}_{m,k}\right|^2}{\left|\boldsymbol{g}_{m,k}\right|^2}=\frac{\beta_{m,k}-r_{m,k}}{\beta_{m,k}}. \tag{7}$$

Then, the NMSE of user $k$ is given by the following expression:

$$e_k=\sum_{m\in\mathcal{M}(k)}\frac{\beta_{m,k}-r_{m,k}}{\beta_{m,k}}, \tag{8}$$

where $\mathcal{M}(k)$ represents the service AP set of user $k$. In this study, we propose minimizing the sum of the channel estimation errors of all users who share the APs with the requested access user $k$ to reduce the channel estimation error of the system. The optimization formula is as follows:

$$\begin{aligned}&\min_{P_{\tau_k}}\sum_{k\in\mathcal{K}(m)}e_k\\&\text{s.t.}\sum_{k\in\mathcal{K}(m)}P_{\tau_k}^2\le\left|\mathcal{K}(m)\right|P_\tau,\end{aligned} \tag{9}$$

where $\mathcal{K}(m)$ represents the user set of service AP $m$. Here, the constraint is the total pilot power limit, and $P_\tau$ is the pilot power of each user. First, the auxiliary variable $\boldsymbol{u}$ is introduced to rewrite the above optimization formula as follows:

$$\begin{aligned}&\min_{P_{\tau_k},\,\boldsymbol{u}}\sum_{k\in\mathcal{K}(m)}\sum_{m\in\mathcal{M}(k)}\left(1-\frac{u_{m,k}}{\beta_{m,k}}\right)\\&\text{s.t.}\begin{cases}\text{(C1)}:\sum\limits_{k\in\mathcal{K}(m)}P_{\tau_k}^2\le\left|\mathcal{K}(m)\right|P_\tau,\\\text{(C2)}:\frac{\left(\sqrt{\tau_\text{p}}P_{\tau_k}\beta_{m,k}\right)^2}{\sum\limits_{j\in\Phi_{\tau_\text{p}}}\tau_\text{p}P_{\tau_k}^2\beta_{m,j}+\sigma^2}\le u_{m,k}.\end{cases}\end{aligned} \tag{10}$$

Because constraint (C2) is non-convex, it can be rewritten using definition (a) as follows:

$$\begin{aligned}\frac{\tau_\text{p}P_{\tau_k}^2\beta_{m,k}^2}{u_{m,k}}&\le\sum_{j\in\Phi_{\tau_\text{p}}}\tau_\text{p}P_{\tau_k}^2\beta_{m,j}+\sigma^2\\&\overset{\text{(a)}}{=}\sum_{j\in\Phi_{\tau_\text{p}}}\tau_\text{p}\left(2P_0P_{\tau_j}-P_0^2\right)\beta_{m,j}+\sigma^2,\end{aligned} \tag{11}$$

where (a) denotes the condition when $x=x_0$ and $x^2=2x_0x-x_0^2$, and $P_0=P_{\tau_j}$. The above optimization problem can be rewritten as

$$\begin{aligned}&\min_{P_{\tau_k},\,\boldsymbol{u}}\sum_{k\in\mathcal{K}(m)}\sum_{m\in\mathcal{M}(k)}\left(1-\frac{u_{m,k}}{\beta_{m,k}}\right)\\&\text{s.t.}\begin{cases}\text{(C2)}:\frac{\tau_\text{p}P_{\tau_k}^2\beta_{m,k}^2}{u_{m,k}}\le\sum\limits_{j\in\Phi_{\tau_\text{p}}}\tau_\text{p}\left(2P_0P_{\tau_j}-P_0^2\right)\beta_{m,j}\\\qquad\qquad+\sigma^2,\\\text{(C1)}.\end{cases}\end{aligned} \tag{12}$$

The joint pilot allocation and AP selection method is given in Algorithm 1.

The optimization of pilot power is based on pilot allocation. For user $k$, the number of service APs is limited. Similarly, the number of users served by an AP is limited. We have $|\mathcal{M}(k)|=\Lambda$, where $\Lambda$ is the maximum number of APs serving user $k$, and the complexity of Eq. (8) is $O(\Lambda)$. According to Eq. (9), its complexity is related to the number of users served by each AP, and $|\mathcal{K}(m)|=V$ where $V$ is the maximum number of users served by AP $m$. Overall, the complexity of the optimization algorithm is $O(V\Lambda)$; therefore, the complexity does not increase with the increase of the number of users.

## 4 Downlink information transmission

In the downlink transmission phase, the AP transmits downlink information to the user according to the channel estimation information. The

**Algorithm 1** Joint pilot allocation and AP selection method

1. Initialization: location information of users and APs, maximum number of users served by an AP $V$, maximum number of APs serving a user $\Lambda$, number of pilots $\tau_{\mathrm{p}}$, $P_{\tau_k,0}=P_\tau, k=1,2,...,K$, $e_{k,0}=\left.\sum_{m\in\mathcal{M}(k)}\frac{\beta_{m,k}-r_{m,k}}{\beta_{m,k}}\right|_{P_{\tau_k}=P_{\tau_k,0}}$, threshold $\varepsilon$
2. When $k\le\tau_{\mathrm{p}}$, the user's service AP set $\mathcal{M}(k)$ includes the $\Lambda$ APs closest to user $k$, and the user is assigned pilot $\tau_k$ to respond to user access
3. When $k>\tau_{\mathrm{p}}$, assign the service AP set $\mathcal{M}(k)$ to user $k$ according to step 2. If $|\mathcal{K}(m)|>V$, select the next closest AP to replace the current AP and assign pilot $\tau_k=\arg\min\sum_{m\in\mathcal{M}(k)}I_{m,k}$ to user $k$, finishing accessing
4. Use the CVX toolbox to solve the optimization problem (12)
5. Obtain the optimal pilot transmission power $P_\tau^*$, $e_k^*=\left.\sum_{m\in\mathcal{M}(k)}\frac{\beta_{m,k}-r_{m,k}}{\beta_{m,k}}\right|_{P_{\tau_k}=P_{\tau_k}^*}$. If $e_k^*\le e_{k,0}$, $e_{k,0}=e_k^*$, $P_0=P_\tau^*$; repeat step 2 until $e_k^*>e_{k,0}$. The optimal pilot transmission power is $P_\tau^*=P_0$

signal received by user $k$ is as follows:

$$y_{\mathrm{d},k}=\sum_{m=1}^{M}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{x}_m+n_k, \tag{13}$$

where $\boldsymbol{g}_{m,k}\in\mathbb{C}^{N\times1}$ is the channel between AP $m$ and user $k$, and $n_k\sim\mathcal{CN}\left(0,\sigma^2\right)$ is the noise. Further, $\boldsymbol{x}_m=\sum_{k'=1}^{K}\sqrt{p_{m,k'}}\boldsymbol{w}_{m,k'}q_{k'}\iota_{m,k'}$ is the signal sent by AP $m$ to all service users, where $q_{k'}$ represents the signal sent to user $k'$, $p_{m,k'}$ represents the power sent by AP $m$, and $\boldsymbol{w}_{m,k'}\in\mathbb{C}^{N\times1}$ is the transmission precoding vector. Here, the expression

$$\iota_{m,k'}=\begin{cases}1, & m\in\mathcal{M}(k')\\ 0, & m\notin\mathcal{M}(k')\end{cases} \tag{14}$$

indicates whether AP $m$ serves user $k'$. The user cannot obtain the current accurate instantaneous value but can receive only information through the average value $\mathbb{E}\left\{\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right\}$ of the channel. The signal received by user $k$ is represented as Eq. (15) (on top of the next page), where term (1) is the user-expected signal, which is the useful signal that the user expects to extract when receiving the signal with mixed multi-user interference and noise, while terms (2) and (3) are unrelated to the transmitted signal. According to Theorem 3 in Demir et al. (2021), the signal-to-interference-plus-noise ratio (SINR) of the received signal at the user is given in Eq. (16) (on top of the next page).

Passive eavesdroppers remain silent next to legitimate users and receive useful information from them. The information received by an eavesdropper is the leakage in the unexpected direction during downlink transmission, which is expressed as follows:

$$\begin{aligned}y_{\mathrm{d},e'}&=\sum_{m=1}^{M}\boldsymbol{g}_{m,e'}^{\mathrm{H}}\sum_{k'=1}^{K}\sqrt{p_{m,k'}}\boldsymbol{w}_{m,k'}q_{k'}\iota_{m,k'}+n_{\mathrm{e}}\\&=\sum_{m=1}^{M}\boldsymbol{g}_{m,e'}^{\mathrm{H}}\sqrt{p_{m,k}}\boldsymbol{w}_{m,k}q_k\iota_{m,k}\\&\quad+\sum_{m=1}^{M}\boldsymbol{g}_{m,e'}^{\mathrm{H}}\sum_{k'\ne k}^{K}\sqrt{p_{m,k'}}\boldsymbol{w}_{m,k'}q_{k'}\iota_{m,k'}+n_{\mathrm{e}},\end{aligned} \tag{17}$$

where $n_{\mathrm{e}}$ is the white Gaussian noise of the eavesdropper.

For an eavesdropper, the worst case is considered; in other words, the eavesdropper knows instantaneous channel state information. In this case, the achievable rate of the eavesdropper is the ideal maximum; therefore, the security problem considered is more reliable, and the SINR of the expected legitimate user obtained by the eavesdropper is

$$\Upsilon_{k,e'}=\frac{\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,e'}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right|^2\right\}}{\sum_{k'\ne k}^{K}\mathbb{E}\left\{\left|\sum_{m=1}^{M}\sqrt{p_{m,k'}}\iota_{m,k'}\boldsymbol{g}_{m,e'}^{\mathrm{H}}\boldsymbol{w}_{m,k'}\right|^2\right\}+\sigma^2}. \tag{18}$$

## 5 Performance analysis

According to Eqs. (16) and (18), the transmission power, user scheduling, and precoding mode affect the system performance. To reduce the complexity of the system, we optimize the transmission power based on the given precoding vector to satisfy the requirements of the system. The MMSE precoding method considers not only multi-user interference but also that the uncertainty of channel estimation has the best performance; however, the complexity of this method is high. We ignore the uncertainty of channel estimation in exchange for low-complexity operations, and local ZF precoding is selected for data transmission (i.e., the APs use the estimated channel of the served user to form a

$$y_{\mathrm{d},k}=\underbrace{\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right\}q_k}_{(1)}+\underbrace{\left(\sum_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}-\mathbb{E}\left\{\sum_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right\}\right)q_k}_{(2)}$$
$$+\underbrace{\sum_{m=1}^{M}\sum_{k'=1,\ k'\neq k}^{K}\sqrt{p_{m,k'}}\boldsymbol{w}_{m,k'}q_{k'}\iota_{m,k'}}_{(3)}+n_k. \tag{15}$$

$$\Upsilon_{\mathrm{d},k}=\frac{\left|\mathbb{E}\left\{\sum\limits_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right\}\right|^2}{\sum\limits_{k'=1}^{K}\mathbb{E}\left\{\left|\sum\limits_{m=1}^{M}\sqrt{p_{m,k'}}\iota_{m,k'}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k'}\right|^2\right\}-\left|\mathbb{E}\left\{\sum\limits_{m=1}^{M}\sqrt{p_{m,k}}\iota_{m,k}\boldsymbol{g}_{m,k}^{\mathrm{H}}\boldsymbol{w}_{m,k}\right\}\right|^2+\sigma^2}. \tag{16}$$

precoding vector for downlink information transmission). Therefore, the precoding vector used in this study is as follows:

$$\boldsymbol{w}_{m,k}=\sqrt{\lambda_{m,k}}\hat{\boldsymbol{g}}_m^{\iota}\left((\hat{\boldsymbol{g}}_m^{\iota})^{\mathrm{H}}\hat{\boldsymbol{g}}_m^{\iota}\right)^{-1}\boldsymbol{e}_k, \tag{19}$$

where $\lambda_{m,k}$ is the normalization coefficient of precoding, $\mathbb{E}\left\{|\boldsymbol{w}_{m,k}|^2\right\}\leq 1$, $\hat{\boldsymbol{g}}_m^{\iota}=\left[\hat{\boldsymbol{g}}_{m,1}^{\iota},\hat{\boldsymbol{g}}_{m,2}^{\iota},\ldots,\hat{\boldsymbol{g}}_{m,K}^{\iota}\right]\in\mathbb{C}^{N\times K}$, and $\boldsymbol{e}_k$ represents the $k^{\mathrm{th}}$ column of the unit matrix.

Because ZF precoding has a strong ability to suppress unexpected directions, passive eavesdroppers exhibit poor performance in this network. We consider the security of a system based on the existence of multiple eavesdroppers. The eavesdroppers cooperate to form a colluding eavesdropping environment. When eavesdroppers receive the information, they use the maximal ratio combining (MRC) to process the information. Therefore, the SINR of colluding to eavesdrop is considered to be the superposition of the SINR of a single eavesdropper (Yang et al., 2016):

$$\Upsilon_{\mathrm{E}}=\sum_{e'=1}^{N_{\mathrm{e}}}\Upsilon_{k,e'}. \tag{20}$$

According to the supplementary materials, the SINR of legitimate user $k$ is

$$\Upsilon_{\mathrm{d},k}=\frac{\left|\sum\limits_{m=1}^{M}\sqrt{p_{m,k}\lambda_{m,k}}\iota_{m,k}\right|^2}{\sum\limits_{k'=1}^{K}\sum\limits_{m=1}^{M}p_{m,k'}\iota_{m,k'}\left(\beta_{m,k}-r_{m,k}\right)/\lambda_{m,k'}+\sigma^2}. \tag{21}$$

According to the precoding power limit, i.e., $\mathbb{E}\left\{\|\boldsymbol{w}_{m,k'}\|\right\}\leq 1$, we have

$$\begin{aligned}&\mathbb{E}\left\{\|\boldsymbol{w}_{m,k'}\|^2\right\}\\=&\mathbb{E}\left\{\left\|\sqrt{\lambda_{m,k}}\hat{\boldsymbol{g}}_m^{\iota}\left((\hat{\boldsymbol{g}}_m^{\iota})^{\mathrm{H}}\hat{\boldsymbol{g}}_m^{\iota}\right)^{-1}\boldsymbol{e}_k\right\|^2\right\}\\=&\mathbb{E}\left\{\lambda_{m,k}\left((\hat{\boldsymbol{g}}_m^{\iota})^{\mathrm{H}}\hat{\boldsymbol{g}}_m^{\iota}\right)^{-1}\right\}\\=&1.\end{aligned} \tag{22}$$

Therefore, according to Theorem 2.10 in Tulino and Verdú (2004), we have

$$\lambda_{m,k}=\left(N-|\boldsymbol{\iota}_m|\right)r_{m,k}, \tag{23}$$

where $|\boldsymbol{\iota}_m|=\left|\sum_{k=1}^{K}\iota_{m,k}\right|$ represents the number of users served by AP $m$, and according to the supplementary materials the SINR of legitimate users can be rewritten as follows:

$$\Upsilon_{k,e'}=\frac{\sum\limits_{m=1}^{M}p_{m,k}\iota_{m,k}\beta_{m,e'}}{\sum\limits_{k'\neq k}^{K}\sum\limits_{m=1}^{M}p_{m,k'}\iota_{m,k'}\beta_{m,e'}+\sigma^2}. \tag{24}$$

The spectral efficiency of a legitimate user $k$ is

$$R_{\mathrm{d},k}=\left(1-\frac{\tau_{\mathrm{p}}}{T}\right)\log_2\left(1+\Upsilon_{\mathrm{d},k}\right), \tag{25}$$

where $T$ represents the coherent time. The spectral efficiency of the colluding eavesdroppers of user $k$ is

$$R_{\mathrm{d,E}}=\left(1-\frac{\tau_{\mathrm{p}}}{T}\right)\log_2\left(1+\sum_{e'=1}^{N_{\mathrm{e}}}\Upsilon_{k,e'}\right), \tag{26}$$

and the security spectral efficiency of user $k$ is given by

$$R_k^{\mathrm{E}} = [R_{\mathrm{d},k} - R_{\mathrm{d,E}}]^+, \tag{27}$$

where $[\cdot]^+$ means that the result returns a positive value. The reduction in complexity is a way to improve system performance by comparing with the way of optimizing precoding vectors. The complexity of optimizing the precoding vector is $O\left(M^2K^2\right)$, while the complexity of optimizing the power is $O\left(MK\right)$. Therefore, compared to optimizing the complex precoding vector, optimizing the power will bring lower complexity.

## 6 Transmission power optimization

For a multi-user system, its security depends on the user with the lowest achievable security rate, and we focus on the AP power distribution scheme to maximize the minimum achievable security rate. Therefore, the optimization problem is formulated as follows:

$$\max_{\boldsymbol{P}} \min \boldsymbol{R}_{\mathrm{d}}^{\mathrm{E}}$$

$$\text{s.t. } (\mathrm{C1}') : \|\boldsymbol{P}_m\|^2 \le P_{\mathrm{t}}, \tag{28}$$

where (C1′) refers to the transmission power limit of AP $m$. This study optimizes the power based on the equal power of each AP; i.e., the maximum transmission power of each AP is $P_{\mathrm{t}}$. To solve the non-convex problem, we introduce new variables $\lambda_k$ and $\psi_k$ first; the objective function is rewritten as follows:

$$\max_{\boldsymbol{P}} \min \boldsymbol{\lambda}$$

$$\text{s.t.} \begin{cases} (\mathrm{C2}') : \log_2\left(1+\sum_{e'=1}^{N_{\mathrm{e}}} \Upsilon_{k,e'}\right) \le \psi_k, \\ (\mathrm{C3}') : \log_2(1+\Upsilon_{\mathrm{d},k}) - \psi_k \ge \lambda_k, \\ (\mathrm{C1}'). \end{cases} \tag{29}$$

By introducing a new variable $T_{k,e'}$, (C2′) can be rewritten follows:

$$(\mathrm{C2a}') : \log_2\left(1+\sum_{e'=1}^{N_{\mathrm{e}}} T_{k,e'}\right) \le \psi_k, \tag{30}$$

$$(\mathrm{C2b}') : \Upsilon_{k,e'} \le T_{k,e'}. \tag{31}$$

$\forall t > 0$, $\log_2(1+t)$ is concave according to the following first-order Taylor expansion:

$$\log_2(1+t) \le \log_2(1+\bar{t}) + (t-\bar{t})\frac{1}{(1+\bar{t})\ln 2}. \tag{32}$$

Thus, constraint (C2a′) can be rewritten as follows:

$$\log_2\left(1+\bar{T}_{k,e'}\right) + \left(\sum_{e'=1}^{N_{\mathrm{e}}} T_{k,e'} - \bar{T}_{k,e'}\right)\frac{1}{(1+\bar{T}_{k,e'})\ln 2} \le \psi_k. \tag{33}$$

Both the numerator and the denominator in (C2b′) are quadratically convex, which can be solved by an equivalent transformation. First, $\Upsilon_{k,e'}$ can be expressed as

$$\Upsilon_{k,e'} = \frac{\boldsymbol{p}_k \boldsymbol{\iota}_k \boldsymbol{\beta}_{e'}}{\sum_{k'\ne k}^{K} \boldsymbol{p}_{k'} \boldsymbol{\iota}_{k'} \boldsymbol{\beta}_{e'} + \sigma^2} = \frac{x_{k,e'}}{y_{k,e'}}, \tag{34}$$

where

$$\begin{cases} x_{k,e'} = \sum_{m=1}^{M} p_{m,k}\iota_{m,k}\beta_{m,e'}, \\ y_{k,e'} = \left\|\boldsymbol{\xi}_{e'}\boldsymbol{I}_{-k}\ \sigma^2\right\|^2, \\ \boldsymbol{\xi}_{e'} = [\xi_{1,e'}, \xi_{2,e'}, ..., \xi_{K,e'}], \\ \xi_{k',e'} = \sqrt{\iota_{m,k'}p_{m,k'}\beta_{m,e'}}\ (k'=1,2,\ldots,K), \end{cases}$$

and $\boldsymbol{I}_{-k}$ represents the unit matrix of $K\times(K-1)$, because

$$\begin{aligned} x_{k,e'} &= \sum_{m=1}^{M} p_{m,k}\iota_{m,k}\beta_{m,e'} \\ &\le \left(\sum_{m=1}^{M}\sqrt{p_{m,k}\iota_{m,k}\beta_{m,e'}}\right)^2 \\ &= x'_{k,e'}. \end{aligned} \tag{35}$$

Therefore, when $\frac{x'_{k,e'}}{y_{k,e'}} \le T_{k,e'}$, we have $\frac{x_{k,e'}}{y_{k,e'}} \le T_{k,e'}$. Because $p_{m,k'} \ge 2\sqrt{\bar{p}_{m,k'}p_{m,k'}} - \bar{p}_{m,k'}$ $(\forall \bar{p}_{m,k'} > 0)$, we have

$$\begin{aligned} \xi_{k',e'}^2 &= \sum_{m=1}^{M}\iota_{m,k'}p_{m,k'}\beta_{m,e'} \\ &\ge \sum_{m=1}^{M}\iota_{m,k'}\beta_{m,e'}\Big(\bar{p}_{m,k'} \\ &\quad + 2\sqrt{\bar{p}_{m,k'}}\left(\sqrt{p_{m,k'}} - \sqrt{\bar{p}_{m,k'}}\right)\Big) \\ &= \sum_{m=1}^{M}\iota_{m,k'}\beta_{m,e'}\left(2\sqrt{\bar{p}_{m,k'}p_{m,k'}} - \bar{p}_{m,k'}\right). \end{aligned} \tag{36}$$

When $\sqrt{p_{m,k'}} = \sqrt{\bar{p}_{m,k'}}$, $p_{m,k'} = 2\sqrt{\bar{p}_{m,k'}}\sqrt{p_{m,k'}} - \bar{p}_{m,k'}$; at this time, $\frac{x'_{k,e'}}{y'_{k,e'}} \le T_{k,e'}$, where $y'_{k,e'} =$

$\sum_{k'\neq k}^{K}\xi_{k',e'}^2+\sigma^2$. We rewrite (C3′) through the proof in the supplementary materials:

$$\begin{aligned}&\lg\left(1+\frac{|S|^2}{I}\right)\\ \geq&\lg\left(1+\frac{|\bar{S}|^2}{\bar{I}}\right)+2\frac{S|\bar{S}|}{\bar{I}\ln 2}\\ &-\frac{|\bar{S}|^2}{\bar{I}\ln 2}-\frac{\left(I+|S|^2\right)|\bar{S}|^2}{\left(\bar{I}+|\bar{S}|^2\right)\bar{I}\ln 2},\end{aligned}\tag{37}$$

where $S$ and $I$ are auxiliary symbols for proofing the formula. According to inequality (37), (C3′) can be rewritten as

$$\begin{aligned}&\lg\left(1+\frac{|\bar{S}_k|^2}{\bar{I}_k}\right)+2\frac{S|\bar{S}_k|}{\bar{I}_k\ln 2}\\ &-\frac{|\bar{S}_k|^2}{\bar{I}_k\ln 2}-\frac{\left(I_k+|S_k|^2\right)|\bar{S}_k|^2}{\left(\bar{I}_k+|\bar{S}_k|^2\right)\bar{I}_k\ln 2}\\ \geq&\lambda_k+\psi_k.\end{aligned}\tag{38}$$

According to the above transformation, the optimization problem can be summarized as

$$\max_{\boldsymbol{P}}\min\boldsymbol{\lambda}$$

$$\text{s.t.}\begin{cases}(\text{C2a}'):\log_2(1+\bar{T}_{k,e'})\\ +\left(\sum_{e'=1}^{N_e}T_{k,e'}-\bar{T}_{k,e'}\right)\frac{1}{(1+\bar{T}_{k,e'})\ln 2}\leq\psi_k,\\ (\text{C2b}'):\left(\sum_{m=1}^{M}\sqrt{p_{m,k}\iota_{m,k}\beta_{m,e'}}\right)^2\Big/T_{k,e'}\\ \leq\sum_{k'\neq k}^{K}\sum_{m=1}^{M}\iota_{m,k'}\beta_{m,e'}\left(2\sqrt{\bar{p}_{m,k'}p_{m,k'}}-\bar{p}_{m,k'}\right)+\sigma^2,\\ (\text{C3}'):\lg\left(1+\frac{|\bar{S}_k|^2}{\bar{I}_k}\right)+2\frac{S|\bar{S}_k|}{\bar{I}_k\ln 2}\\ -\frac{|\bar{S}_k|^2}{\bar{I}_k\ln 2}-\frac{\left(I_k+|S_k|^2\right)|\bar{S}_k|^2}{\left(\bar{I}_k+|\bar{S}_k|^2\right)\bar{I}_k\ln 2}\geq\lambda_k+\psi_k,\\ (\text{C1}'),\\ \forall\bar{S}_k,\bar{I}_k,\bar{p}_{m,k},\bar{T}_{k,e'}>0,\end{cases}\tag{39}$$

where $\bar{S}_k$ is the initial expected signal of legitimate user $k$, $\bar{I}_k$ is the initial interference-plus-noise of legitimate user $k$, $\bar{p}_{m,k}$ is the initial power allocation of AP $m$ to user $k$, and $\bar{T}_{k,e'}$ is the initial SINR of eavesdropper $e'$ eavesdropping on legitimate user $k$. The initial values are optional. To reduce the number of iterations, all initial values are set under equal power allocation in this study. The power allocation algorithm is as shown in Algorithm 2.

**Algorithm 2** Power allocation algorithm

1. Initialization: selecting an error tolerance $\varepsilon>0$, calculating the initial value under equal power distribution $\bar{p}_{m,k}=P_{\text{t}}/K$, where

$$\bar{S}_k=\left|\sum_{m=1}^{M}\sqrt{\bar{p}_{m,k}\iota_{m,k}\left(N-|\boldsymbol{\iota}_m|\right)r_{m,k}}\right|^2,$$

$$\bar{I}_k=\sum_{k'=1}^{K}\sum_{m=1}^{M}\frac{\bar{p}_{m,k'}\iota_{m,k'}\left(\beta_{m,k}-r_{m,k}\right)}{\left(N-|\boldsymbol{\iota}_m|\right)r_{m,k}}+\sigma^2,$$

$$\bar{T}_{k,e'}=\frac{\sum_{m=1}^{M}\bar{p}_{m,k}\iota_{m,k}\beta_{m,e'}}{\sum_{k'\neq k}^{K}\sum_{m=1}^{M}\bar{p}_{m,k'}\iota_{m,k'}\beta_{m,e'}+\sigma^2}.$$

2. Use the CVX toolbox to solve problem (39)
3. Obtain the optimized power allocation $p_{m,k}^*$. If $\text{MSE}_p\geq\varepsilon$, $\bar{p}_{m,k}=p_{m,k}^*$; repeat step 2. Once $\text{MSE}_p<\varepsilon$, stop the iteration and output the optimal $p_{m,k}^*$, where

$$\text{MSE}_p=\frac{1}{K}\sum_{k=1}^{K}\|\bar{\boldsymbol{p}}_k-\boldsymbol{p}_k^*\|$$

## 7 Simulations, results, and discussion

In this section, we verify the feasibility and superiority of the proposed method using simulations. Consider a multi-antenna AP CF massive MIMO system, where the network coverage is a rectangular area with length and width of 1 km, antenna height of 15 m, user height of 1.65 m, noise power $\sigma^2=-96$ dBm, system bandwidth $B=20$ MHz, and carrier center frequency $G=2.4$ GHz. The number of pilot multiplexes is represented by $t$. The path loss is calculated using the three-slope model (Ngo et al., 2017):

$$\begin{aligned}&\beta_{m,k}\,[\text{dB}]\\ =&\begin{cases}-35.7-35\lg d_{m,k}+F_{m,k},\ d_{m,k}>d_1,\\ -61.2-20\lg d_{m,k},\ d_0<d_{m,k}\leq d_1,\\ -81.2,\ d_{m,k}\leq d_0,\end{cases}\end{aligned}\tag{40}$$

where $d_0=10$ m and $d_1=50$ m are the reference distances, $d_{m,k}$ is the distance between AP $m$ and user $k$, and shadow fading $F_{m,k}\sim\mathbb{N}\left(0,8^2\right)$ exists only when $d_{m,k}>d_1$.

Figs. 2 and 3 show the NMSE performance of different pilot allocation algorithms under different

pilot multiplexes, where

$$e_{\mathrm{NMSE}}=\frac{\sum_{k=1}^{K}\mathbb{E}\left\{\left|\hat{\boldsymbol{g}}_k-\boldsymbol{g}_k\right|^2\right\}}{\sum_{k=1}^{K}\mathbb{E}\left\{\left|\hat{\boldsymbol{g}}_k\right|^2\right\}}. \tag{41}$$

Reference schemes include pilot allocation method based on the main AP (Björnson and Sanguinetti, 2020), greedy pilot allocation algorithm (Ngo et al., 2017), orthogonal pilot allocation scheme, cluster-based pilot allocation method (Attarifar et al., 2018), and random allocation method (Interdonato et al., 2019b). In the simulations, the AP selection mentioned in Section 3 is used to further verify the advantages of the proposed method. It can be seen from Fig. 2 that in all the above schemes, the orthogonal pilot allocation scheme and the random allocation method act as the upper bound and lower bound, respectively. In addition, the optimization scheme proposed in this study is superior to the pilot allocation method based on main AP, greedy pilot allocation algorithm, and cluster-based pilot allocation method in terms of channel estimation accuracy. Moreover, the proposed optimization scheme is (1) better than the orthogonal pilot allocation scheme in which all APs serve users and (2) close to the orthogonal pilot allocation method based on AP selection, i.e., the upper bound of pilot allocation performance.

It can be seen from Fig. 3 that the channel estimation error increases with the increase of the number of pilot multiplexes. This is because the larger the number of pilot multiplexes, the smaller the average distance between users sharing the same pilot, and the greater the impact of receiving the same pilot interference. The pilot allocation algorithm proposed in this study is almost close to the result of obtaining perfect channel state information for the case without pilot multiplexes.

The influence of the total AP number and user-centric AP selection on the pilot estimation error is shown in Fig. 4. This study adopts the local ZF precoding method, and the precoding vector is related only to the number of users served by each AP; only channel estimates on user-centric APs affect the transmission performance. It can be observed from the simulation diagram that the greater the total number of antennas in the system, the better the

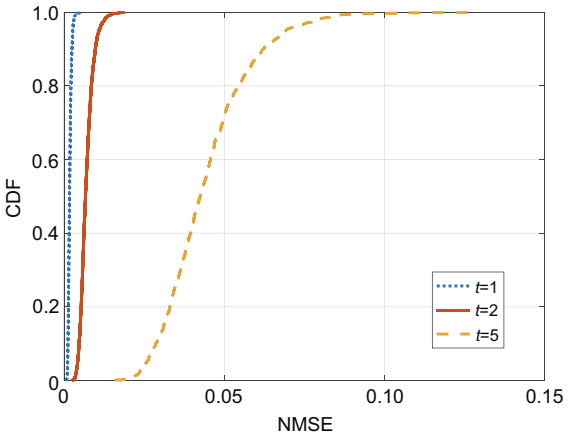


**Fig. 3 NMSE of different pilot multiplexes with $M=16$, $N=8$, and $K=4$ (NMSE: normalized mean square error; CDF: cumulative distribution function)**

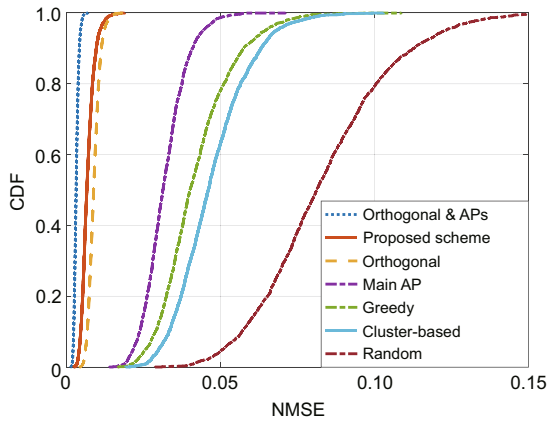


**Fig. 2 NMSE of different pilot assignment algorithms with $M=100$, $N=8$, $K=40$, and $t=2$ (NMSE: normalized mean square error; CDF: cumulative distribution function)**

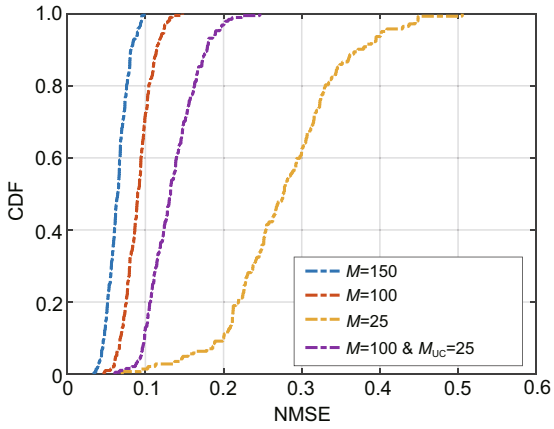


**Fig. 4 Influence of the total AP number and user-centric AP selection on pilot estimation error with $N=8$, $K=40$, and $t=5$ (NMSE: normalized mean square error; CDF: cumulative distribution function)**

channel estimation performance. By comparing the purple curve with the red curve, it can be observed that the user-centric AP selection has less loss in terms of channel estimation error. By comparing the purple curve with the yellow curve, we see that the pilot estimation performance does not deteriorate significantly because the number of service APs per user decreases.

From Figs. 5 and 6, we can see how the number of APs impacts the minimum security rate and sum security rate, respectively. The comparison schemes are the equal power allocation scheme and the proposed max–min power allocation scheme. The simulation results show that with an increase in the number of APs, the minimum security rate and sum security rate of the system are improved, and the proposed max–min scheme is superior to the equal power allocation scheme. In terms of sum security rate, the max–min power optimization method is twice that of the equal power allocation method. In terms of minimum security rate, the probability that the minimum security rate of users in the system $>0$ is 20%, whereas the probability after optimization is 100%, which implies that the max–min power optimization method has obvious advantages in improving the worst transmission by the system.

Figs. 7 and 8 compare the sum security spectral efficiency and the minimum security spectral efficiency under different power optimization schemes, respectively. The schemes are the equal power allocation scheme, max–sum power allocation scheme,

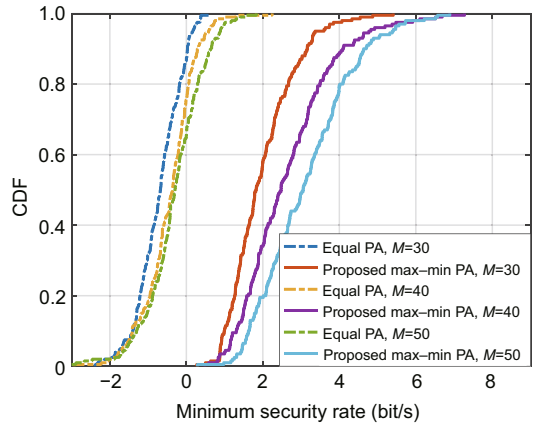


**Fig. 5 Comparison of the minimum security rate under different power allocation and AP selection strategies with $N=12$, $K=10$, $t=5$, and $N_{\mathrm{e}}$=5 (CDF: cumulative distribution function)**

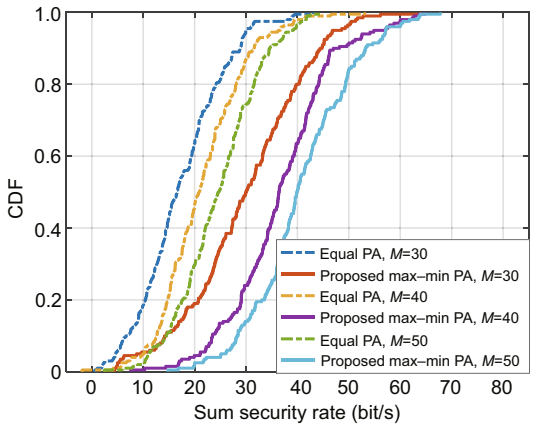


**Fig. 6 Comparison of the sum security rate under different power allocation and AP selection strategies with $N=12$, $K=10$, $t=5$, and $N_{\mathrm{e}}$=5 (CDF: cumulative distribution function)**

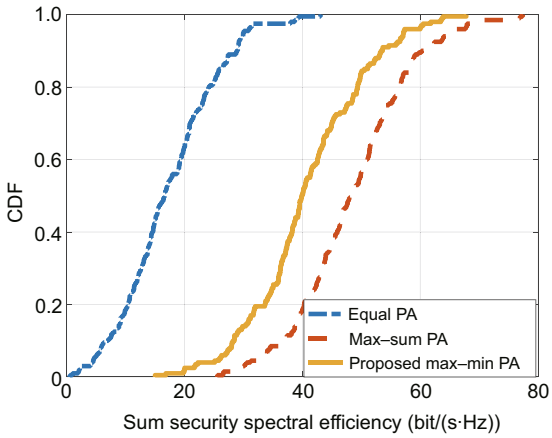


**Fig. 7 Comparison of the sum security spectral efficiency with $M=50$, $N=12$, $K=10$, $t=5$, and $N_{\mathrm{e}}=5$ (CDF: cumulative distribution function)**

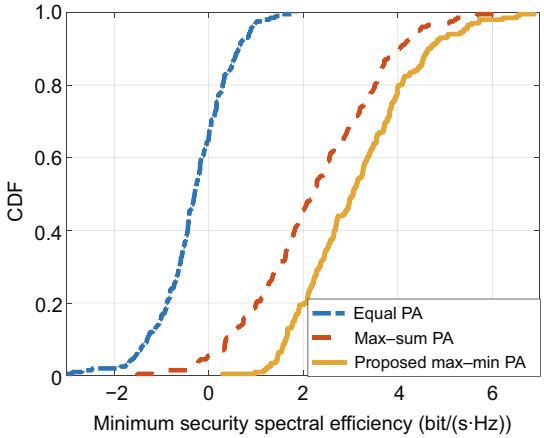


**Fig. 8 Comparison of the minimum security spectral efficiency with $M=50$, $N=12$, $K=10$, $t=5$, and $N_{\mathrm{e}}=5$ (CDF: cumulative distribution function)**

and the proposed max–min power allocation scheme. The max–sum scheme changes the optimization objective of Eq. (28) to $\max_{\boldsymbol{P}} \sum_{k=1}^{K} \boldsymbol{R}_{\mathrm{d}}^{\mathrm{E}}$. The simulation results show that the proposed max–min power allocation scheme can improve the worst security performance of the system; however, it is inferior to the max–sum power allocation scheme in terms of improving the sum security spectral efficiency of the system. This is because the scheme proposed in this study attempts to ensure that each user is in a secure transmission state as much as possible, and the simulation conforms to the theoretical analysis.

## 8 Conclusions

In this paper, we have investigated the resource allocation problem of a CF massive MIMO system under the condition of colluding eavesdropping by multiple passive eavesdroppers. First, given limited pilot resources, a scalable pilot allocation scheme has been proposed. Based on the user-service AP selection, the pilot with the minimum pollution has been selected as the access user, and the transmission power of the pilot has been optimized to improve the accuracy of the channel estimation. To address the situation of multiple colluding eavesdroppers in the network, a secure transmission scheme based on power allocation has been proposed. With the goal of maximizing the minimum security spectral efficiency, based on user scheduling, the power allocation at the transmitter has been optimized to improve the security of the system. Simulation results showed that the proposed pilot allocation scheme can effectively reduce the channel estimation error. Moreover, to maximize the security spectrum efficiency, the proposed scheme not only ensures safe transmission by the system but also significantly improves the security of the system.

### Contributors

Na LI designed the research and processed the data. Yuanyuan GAO drafted the paper. Kui XU and Xiaochen XIA helped organize the paper. Huazhi HU, Yang LI, and Yueyue ZHANG revised and finalized the paper.

### Conflict of interest

All the authors declare that they have no conflict of interest.

### Data availability

The data that support the findings of this study are available from the authors upon reasonable request.

### References

Alonzo M, Buzzi S, Zappone A, et al., 2019. Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave. *IEEE Trans Green Commun Netw*, 3(3):651-663. https://doi.org/10.1109/TGCN.2019.2908228

Ammar HA, Adve R, Shahbazpanahi S, et al., 2022a. Downlink resource allocation in multiuser cell-free MIMO networks with user-centric clustering. *IEEE Trans Wirel Commun*, 21(3):1482-1497. https://doi.org/10.1109/TWC.2021.3104456

Ammar HA, Adve R, Shahbazpanahi S, et al., 2022b. Distributed resource allocation optimization for user-centric cell-free MIMO networks. *IEEE Trans Wirel Commun*, 21(5):3099-3115. https://doi.org/10.1109/TWC.2021.3118303

Attarifar M, Abbasfar A, Lozano A, 2018. Random vs structured pilot assignment in cell-free massive MIMO wireless networks. Proc IEEE Int Conf on Communications Workshops, p.1-6. https://doi.org/10.1109/ICCW.2018.8403508

Bashar M, Ngo HQ, Burr AG, et al., 2018. On the performance of backhaul constrained cell-free massive MIMO with linear receivers. Proc 52nd Asilomar Conf on Signals, Systems, and Computers, p.624-628. https://doi.org/10.1109/ACSSC.2018.8645433

Björnson E, Sanguinetti L, 2020. Scalable cell-free massive MIMO systems. *IEEE Trans Commun*, 68(7):4247-4261. https://doi.org/10.1109/TCOMM.2020.2987311

Björnson E, Hoydis J, Sanguinetti L, 2017. Massive MIMO networks: spectral, energy, and hardware efficiency. *Found Trends Signal Process*, 11(3-4):154-655. https://doi.org/10.1561/2000000093

Boroujerdi MN, Abbasfar A, Ghanbari M, 2017. Antenna assignment in cell free massive MIMO systems. Proc Iranian Conf on Electrical Engineering, p.1747-1751. https://doi.org/10.1109/IranianCEE.2017.7985333

Buzzi S, D'Andrea C, 2017a. Cell-free massive MIMO: user-centric approach. *IEEE Wirel Commun Lett*, 6(6):706-709. https://doi.org/10.1109/LWC.2017.2734893

Buzzi S, D'Andrea C, 2017b. User-centric communications versus cell-free massive MIMO for 5G cellular networks. Proc 21st Int ITG Workshop on Smart Antennas, p.1-6.

Buzzi S, D'Andrea C, Fresia M, et al., 2021. Pilot assignment in cell-free massive MIMO based on the Hungarian algorithm. *IEEE Wirel Commun Lett*, 10(1):34-37. https://doi.org/10.1109/LWC.2020.3020003

Chen SF, Zhang JY, Björnson E, et al., 2021. Structured massive access for scalable cell-free massive MIMO systems. *IEEE J Select Areas Commun*, 39(4):1086-1100. https://doi.org/10.1109/JSAC.2020.3018836

Choi J, Park J, 2021. Sum secrecy spectral efficiency maximization in downlink MU-MIMO: colluding eavesdroppers. *IEEE Trans Veh Technol*, 70(1):1051-1056. https://doi.org/10.1109/TVT.2020.3048736

Demir ÖT, Björnson E, Sanguinetti L, 2021. Foundations of user-centric cell-free massive MIMO. *Found Trends Signal Process*, 14(3-4):162-472. https://doi.org/10.1561/2000000109

Dong GN, Zhang HX, Jin S, et al., 2019. Energy-efficiency-oriented joint user association and power allocation in distributed massive MIMO systems. *IEEE Trans Veh Technol*, 68(6):5794-5808. https://doi.org/10.1109/TVT.2019.2912388

Huang QH, Burr A, 2017. Compute-and-forward in cell-free massive MIMO: great performance with low backhaul load. Proc IEEE Int Conf on Communications Workshops, p.601-606. https://doi.org/10.1109/ICCW.2017.7962724

Interdonato G, Ngo HQ, Frenger P, et al., 2019a. Downlink training in cell-free massive MIMO: a blessing in disguise. *IEEE Trans Wirel Commun*, 18(11):5153-5169. https://doi.org/10.1109/TWC.2019.2933831

Interdonato G, Björnson E, Ngo HQ, et al., 2019b. Ubiquitous cell-free massive MIMO communications. *EURASIP J Wirel Commun Netw*, 2019(1):197. https://doi.org/10.1186/s13638-019-1507-0

Larsson EG, Edfors O, Tufvesson F, et al., 2014. Massive MIMO for next generation wireless systems. *IEEE Commun Mag*, 52(2):186-195. https://doi.org/10.1109/MCOM.2014.6736761

Liu H, Zhang JY, Zhang XD, et al., 2020a. Tabu-search-based pilot assignment for cell-free massive MIMO systems. *IEEE Trans Veh Technol*, 69(2):2286-2290. https://doi.org/10.1109/TVT.2019.2956217

Liu H, Zhang JY, Jin S, et al., 2020b. Graph coloring based pilot assignment for cell-free massive MIMO systems. *IEEE Trans Veh Technol*, 69(8):9180-9184. https://doi.org/10.1109/TVT.2020.3000496

Mendoza CF, Schwarz S, Rupp M, 2020. Cluster formation in scalable cell-free massive MIMO networks. Proc 16$^{\text{th}}$ Int Conf on Wireless and Mobile Computing, Networking and Communications, p.62-67. https://doi.org/10.1109/WiMob50308.2020.9253391

Mirmohseni M, Papadimitratos P, 2014. Colluding eavesdroppers in large cooperative wireless networks. Proc Iran Workshop on Communication and Information Theory, p.1-6. https://doi.org/10.1109/IWCIT.2014.6842500

Ngo HQ, Ashikhmin A, Yang H, et al., 2015. Cell-free massive MIMO: uniformly great service for everyone. Proc 16$^{\text{th}}$ Int Workshop on Signal Processing Advances in Wireless Communications, p.201-205. https://doi.org/10.1109/SPAWC.2015.7227028

Ngo HQ, Ashikhmin A, Yang H, et al., 2017. Cell-free massive MIMO versus small cells. *IEEE Trans Wirel Commun*, 16(3):1834-1850. https://doi.org/10.1109/TWC.2017.2655515

Ngo HQ, Tran LN, Duong TQ, et al., 2018. On the total energy efficiency of cell-free massive MIMO. *IEEE Trans Green Commun Netw*, 2(1):25-39. https://doi.org/10.1109/TGCN.2017.2770215

Sarker M, Fapojuwo AO, 2021. Granting massive access by adaptive pilot assignment scheme for scalable cell-free massive MIMO systems. Proc 93$^{\text{rd}}$ Vehicular Technology Conf, p.1-5. https://doi.org/10.1109/VTC2021-Spring51267.2021.9449014

Timilsina S, Kudathanthirige D, Amarasuriya G, 2018. Physical layer security in cell-free massive MIMO. Proc IEEE Global Communications Conf, p.1-7. https://doi.org/10.1109/GLOCOM.2018.8647876

Tulino AM, Verdú S, 2004. Random matrix theory and wireless communications. *Commun Inform Theory*, 1(1):1-82.

van Chien T, Björnson E, Larsson EG, 2020. Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks. *IEEE Trans Wirel Commun*, 19(10):6798-6812. https://doi.org/10.1109/TWC.2020.3006083

Vu TX, Chatzinotas S, ShahbazPanahi S, et al., 2020. Joint power allocation and access point selection for cell-free massive MIMO. Proc IEEE Int Conf on Communications, p.1-6. https://doi.org/10.1109/ICC40277.2020.9148948

Xia XJ, Fan ZQ, Luo WY, et al., 2023. Joint uplink power control, downlink beamforming, and mode selection for secrecy cell-free massive MIMO with network-assisted full duplexing. *IEEE Syst J*, 17(1):720-731. https://doi.org/10.1109/JSYST.2022.3188514

Yang MQ, Guo DX, Huang YZ, et al., 2016. Physical layer security with threshold-based multiuser scheduling in multi-antenna wireless networks. *IEEE Trans Commun*, 64(12):5189-5202. https://doi.org/10.1109/TCOMM.2016.2606396

Yu Y, Liu SM, Yuan WN, et al., 2020. Robust secure beamforming for multi-receiver multi-eavesdropper MIMO SWIPT systems. Proc IEEE Global Communications Conf, p.1-6. https://doi.org/10.1109/GLOBECOM42002.2020.9348120

## List of supplementary materials

1 Derivation of Eq. (21)

2 Derivation of Eq. (24)

3 Derivation of inequality (37)