Pilot Assignment and Power Control in Secure UAV-Enabled Cell-Free Massive MIMO Networks

Yong Chen, Xianyu Zhang, Fuqiang Yao, Kang An, Gan Zheng, *Fellow, IEEE,* and Symeon Chatzinotas, *Fellow, IEEE*

***Abstract*—This paper investigates the pilot assignment and power control problems for secure UAV communications in cell-free massive MIMO network with the user-centric scheme, where numerous distributed access points (APs) simultaneously serve multiple UAVs and terminal users. Meanwhile, there exists one UAV acting as an eavesdropper which can perform pilot spoofing attack. Considering a mixture of Rayleigh and Ricean fading channels, the APs respectively perform MMSE estimation and distributed conjugate beamforming for uplink training and downlink data transmission. Using random matrix theory, the closed-form expression for a tight lower bound on the achievable secrecy rate is derived, which enables the impact analysis of key parameters, such as power, antenna configuration, UAV height, etc. Taking into account both performance and complexity, a novel pilot assignment scheme is proposed by combining weighted graphic framework and genetic algorithm, which can actualize global search with limited iterations. The max-min power control with security constraints is then studied in parallel, which can not only enhance the network fairness but also ensure the security. Accordingly, successive convex approximation and fractional optimization are jointly utilized to solve this non-convex problem. Simulation results numerically verify the analytical results and indicate the superiority of the proposed pilot assignment and power control schemes.**

***Index Terms*—Cell-Free, massive MIMO, UAV, physical layer security, user-centric, pilot assignment, power control.**

I. Introduction

NUMEROUS recent works have witnessed the superior inherent advantages of cell-free massive multi-input multi-output (MIMO) technology, such as enhanced spectral efficiency (SE) and energy efficiency (EE), larger coverage ratio, improved reliability and security, etc [1], [2]. A cell-free massive MIMO architecture comprises of plenty of distributed APs coherently serving several users within the same time-frequency resource, which can jointly reap the merits of traditional collocated massive MIMO and network MIMO. As an appealing candidate for beyond 5G network, much research on cell-free massive MIMO networks has emerged in terms of performance analysis [3], beamformer design [4], power control [5], and pilot decontamination [6], etc. Most notably, to reduce the system's complexity, the user-centric (UC) rule has been recently introduced in cell-free massive MIMO systems to provide a virtual-cell alternative architecture [7], [8].

Due to the limited available pilots, pilot contamination constitutes a bottleneck for system performance. Thus, the acquisition of excellent pilot allocation strategies has gained attention in some recent literatures. One straightforward random pilot assignment method was utilized in [2]. However, this scheme is inefficient due to the deficiency of optimization strategy. In response, literature [2] proposed a greedy pilot assignment to iteratively refine the achievable minimum rate. But this algorithm cannot guarantee a stable global convergence. Instead, to avoid the pilot reuse in nearby users, the works in [9], [10], [11], [12] designed corresponding pilot assignment schemes by considering the geographical proximity. However, owing to limited information, these approaches cannot accurately measure pilot contamination. Note that a novel pilot assignment scheme by using weighted graphic framework was investigated in [13], where a metric was established to capture the effects of pilot contamination. Nevertheless, this algorithm only can yield a local optimal solution.

Recently, unmanned aerial vehicles (UAVs) have been achieving unprecedented interest from industry and academia driven by the inherent advantages including agility, mobility, versatility, autonomy, etc. Correspondingly, UAVs have been integrated into a wide range of application scenarios, such as military missions, industrial inspection, surveillance, cargo delivery, search and rescue, etc. Particularly, thanks to the flexibility and favorable propagation condition, the integration of UAVs and wireless communication systems has been one of the research hotspots and widely investigated in many seminal works, where UAVs mainly act as aerial base stations or aerial mobile users [14], [15]. On the other hand, the basic requirements of UAV communication should be divided into two categories: high reliable command and control links and ubiquitous payload data services. To support reliable, low-latency and high-rate communication links, cell-free massive MIMO serving UAVs is one effective solution and has attracted enormous attentions in recent years. Although current works have validated the benefits of this application, this novel communication paradigm still faces many fundamental issues and challenges, including security, pilot contamination, resource allocation, etc [16], [17].

Conventionally, information security was established based

Y. Chen is with College of Communication Engineering, Army Engineering University of PLA, Nanjing, 210007 China, and also with the Sixth-third/63rd Research Institute, National University of Defense Technology, Nanjing, 210007 China. (e-mail: chy63s@126.com.)

X. Zhang, F. Yao and K. An are with the Sixth-third/63rd Research Institute, National University of Defense Technology, Nanjing, 210007 China. (e-mail: zhangxy_sat@126.com, yfq2030@163.com, ankang89@nudt.edu.cn.)

G. Zheng is with the School of Engineering, University of Warwick, Coventry, CV4 7AL, UK. (e-mail: gan.zheng@warwick.ac.uk).

S. Chatzinotas is with Institute of Informatics and Telecommunications, NCSR 'Demokritos', Ag. Paraskevi 153 10, Athens, Greece. (e-mail: schatzin@ieee.org)

This work was supported in part by Project funded by China Postdoctoral Science Foundation (No. 2021MD703980), in part by the National Natural Science Foundation of China (No. 61901502 and 62071352).

on higher-layer cryptographic encryption techniques. However, these traditional approaches result in high computational complexity and energy consumption. As a valuable alternative, physical layer security (PLS) can provide security enhancement without secret key manipulations, which has been extensively studied over past decades [18], [19]. Especially in massive MIMO systems, existing contributions has verified that PLS has great potential and advantages [20], [21]. In general, most researches focused on PLS in co-located massive MIMO networks including performance analysis, secrecy precoding, security enhancement, attack countermeasures, and so on [22], [23], [24]. Meanwhile, some efforts have been presented to investigate the secure communication in cell-free massive MIMO systems. The seminal work [25] firstly investigated the security aspects in cell-free massive MIMO with pilot spoofing attack, which provided various power allocation approaches. Considering the multigroup multicast scenario, reference [26] analyzed the achievable secrecy rate and proposed one active attack detection scheme. To tackle the extremely-costly and energy-hungry problem, some researchers explored their efforts on studying the impact of the RF impairments, phase noise or coarse ADCs/DACs on secure cell-free massive MIMO networks [27], [28], [29]. Aside from that, integration of reconfigurable intelligent surface (RIS) and secure cell-free massive MIMO networks has been considered in recent works [30], [31], which verified the potential of RIS. However, the aforementioned research all focused on terrestrial networks. Actually, works on secret UAV communication assisted by massive MIMO system are scarce. The work [32] jointly considered the physical layer security and authentication in UAV co-located massive MIMO network. The previous work [33] focused on the secure cell-free massive MIMO network for multi-user communication, where the closed-form achievable secrecy rate was derived for performance analysis. Especially noteworthy is that integrating UAVs and ground users with future wireless networks has been a hotspot recently. However, no prior works have studied the secrecy aspect in cell-free massive MIMO networks jointly serving both UAVs and terrestrial terminals.

Additionally, due to the lack of sufficient pilot sequences, pilot reuse is inevitable in cell-free massive MIMO system, which would lead to pilot contamination and degrade the system performance. To alleviate this deficiency, it is crucial to design an effective pilot assignment scheme. Besides, power control is commonly implemented to enhance the performance of cell-free massive MIMO system. Inspired by the above considerations, this paper focuses on the secure cell-free massive MIMO system with user-centric (UC) rule [34] serving both UAVs and ground users, wherein active UAV eavesdropper attempts to intercept confidential information intended for some target user. The novelties and contributions of the article can be summarized as follows:

- Assuming the mixture of Ricean and Rayleigh channels for terminal users (TUEs) and UAVs, the secure UAV communication served by cell-free massive MIMO network is studied taking into account the UC approach, pilot contamination and active pilot spoofing attack. This paper derives the closed-form expressions of the achievable ergodic secrecy rate, which provides an efficient tool to analyze the impacts of various critical system configurations, such as pilot assignment schemes, power control metrics, network architectures, and so on.
- A novel pilot assignment scheme is proposed to reduce the pilot contamination effect in the considered secure communication system. Firstly, the interference graph among the users is constructed based on the large-scale fading coefficients. On the basis of this metric, initial pilot assignment schemes are obtained using a weighted graphic framework. Then, an iterative procedure based on the genetic algorithm is implemented to alleviate the pilot contamination and enhance the fairness.
- Finally, based on the closed-form expressions of the achievable secrecy rates, the power control strategy is proposed aiming to maximize the minimum achievable rate of all users subject to the constraint on per-AP total power and information leakage to the eavesdropper. More specifically, the fractional optimization strategy combining with successive convex approximation is applied to solve the corresponding non-convex problem, which can make a tradeoff between the performance improvement and computational complexity.

The rest of this article is outlined as follows. Section II introduces the system model including of uplink training and down data transmission. Section III performs secrecy performance analysis and derives some analytical results. The pilot assignment scheme is discussed and solved in Section IV. In parallel, the max-min power control optimization problem is proposed and solved in Section V. All presented discussions are numerically verified by simulations in Section VI. Finally, Section VII concludes the work.

Notation: Throughout this article, non-bold lowercase, boldface lowercase and boldface uppercase letters stand for scalars, vectors and matrices, i.e. a, $\mathbf{a}$ and $\mathbf{A}$. Superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and $(\cdot)^{-1}$ are used to denote transpose, conjugate, conjugate transpose, inverse operator, respectively. Besides, $\mathbb{E}\{\cdot\}$, $\mathrm{Var}\{\cdot\}$, $\mathrm{R}\{\cdot\}$, $\mathrm{tr}(\cdot)$ and $\|\cdot\|$ refer to expectation, variance, real part, trace and norm operations, respectively. $\mathbf{z} \sim \mathcal{CN}(\mathbf{a}, \mathbf{A})$ represents a complex Gaussian random vector ($\mathbf{z} \in \mathbb{C}^{1\times N}$) with mean $\mathbf{a} \in \mathbb{C}^{1\times N}$ and covariance matrix $\mathbf{A} \in \mathbb{C}^{N\times N}$. $\mathbf{I}_N$ refers to $N \times N$ identity matrix. Finally, $[x]^+ = \max(x, 0)$.

II. System model

As illustrated in Fig.1, this paper concerns a cell-free massive MIMO network consisting of M outdoor APs that cooperatively serve K_U UAVs and K_T TUEs with the same time-frequency resource. All APs are assumed to be equipped with N omni-directional antennas, while all UAVs and TUEs are single-antenna terminals. Moreover, this paper supposes that one active single-antenna UAV-Eve exists and attempts to intercept the confidential data intended to one legitimate user. On the other hand, all APs, UAVs and TUEs are randomly located over a large geographic region, where a back-haul network connects the APs to one central processing unit

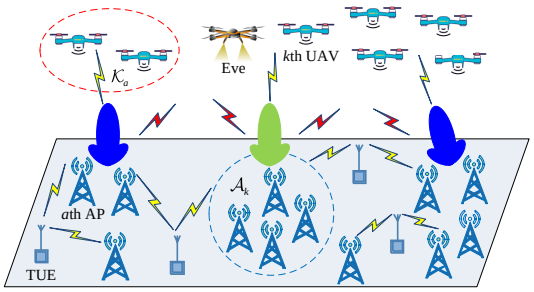


Fig. 1. Illustration of the secure UAV communication in Cell-Free massive MIMO network.

(CPU) and jointly perform data processing and information exchanging. Similar to previous works on cell-free massive MIMO system, this work focuses on time division duplexing (TDD) mode in which channel state information (CSI) can be obtained via uplink training because of the channel reciprocity. Thus, one coherence interval τ_c would be split into uplink training and downlink data transmission with interval durations τ_p and τ_d.

A. Channel Model

In the considered secure cell-free massive MIMO system, it is challenging to build an effective model to characterize the propagation channels of the hybrid TUE-UAV network. That is because the channel characteristics of UAV-to-AP and TUE-to-AP are profoundly different. As aerial users, UAVs are more likely to experience favorable line-of-sight (LoS) transmission conditions. Conversely, due to the numerous obstacles and complicated propagation environment, TUEs would undergo more severe path loss, shadowing fading and multi-path effect. According to literature [35], current work employs the mixture of Rician- and Rayleigh-faded channels to model the channels of the UAVs and TUEs. In the following, propagation models of UAV-to-AP and TUE-to-AP are introduced, respectively.

As aerial devices, flying UAVs generally undergo excellent LoS propagation condition. Importantly, the probability of LoS on the UAV-to-AP link is mainly determined by flying height, propagation distance and environment, etc.. Therefore, Rician fading channels with some variable parameters could be used to describe the UAV-to-AP propagation. Let $\mathbf{g}_{k,a} \in \mathbb{C}^{N\times 1}$ be the generic channel between the kth UAV and the ath AP, which should consist of LOS component and non LoS (NLoS) component. Consequently, $\mathbf{g}_{k,a}$ can be written as

$$\mathbf{g}_{k,a} = \sqrt{\frac{\beta_{k,a}}{K_a^{(k)}+1}}\left[\sqrt{K_a^{(k)}}e^{j\vartheta_{k,a}}\mathbf{a}_{k,a} + \mathbf{h}_{k,a}\right], \quad (1)$$

where $\beta_{k,a}$ and $K_a^{(k)}$ denote the large-scale fading coefficient and the Ricean K-factor, respectively. Besides, vector $\mathbf{h}_{k,a} \sim \mathcal{CN}(0, \mathbf{I}_N)$ stands for the small-scale fading component, and $\vartheta_{k,a}$ represents the random phase shift of the LoS link which follows the uniform distribution within $[0, 2\pi]$. Here vector $\mathbf{a}_{k,a} \in \mathbb{C}^{N\times 1}$ describes the direct path which depends on the 3D relative positions between the ath AP and the antenna elements at the kth UAVs. Thus, vector $\mathbf{a}_{k,a}$ can be represented as

$$\mathbf{a}_{k,a} = \left[1, \cdots, e^{j\frac{2\pi}{\lambda}d_{k,a,1}^{(3D)}}, \cdots, e^{j\frac{2\pi}{\lambda}d_{k,a,N}^{(3D)}}\right]^T, \quad (2)$$

with λ being the carrier wavelength and $d_{k,a,n}^{(3D)}$ standing for the 3D distance difference from the kth UAV to the nth and the first antenna element of the ath AP.

Moreover, the appearing probability of LoS path, represented by P_{ak}^{LoS}, relies on the height of kth UAV, relative distance between the two nodes(target UAV and AP), and the practical environment. As in the literature [35], P_{ak}^{LoS} can be given by

$$P_{ak}^{LoS} = \begin{cases} \frac{d_1}{d_{ak}} + \exp\left(\frac{-d_{ak}}{p_1}\right)\left(1-\frac{d_1}{d_{ak}}\right), d_{ak} > d_1 \\ 1, d_{ak} \le d_1 \end{cases}, \quad (3)$$

Here

$$d_1 = \max\left(294.05\log_{10}(h_U) - 432.94, 18\right), \quad (4)$$

$$p_1 = 233.98\log_{10}(h_U) - 0.95 \quad (5)$$

are about the function of the altitude of the kth UAV h_U.

Consequently, it is worth noting the Rician factor $K_a^{(k)}$ can be calculated as follows

$$K_a^{(k)} = \frac{P_{ak}^{LoS}}{1 - P_{ak}^{LoS}}. \quad (6)$$

Further, referring to existing literature [35], the large-scale fading $\beta_{k,a}$ could be calculated using the following model as

$$\begin{aligned}\beta_{k,a} = &-\max\left(PL_{f_c}, 30.9 + 20\log_{10}(f_c) + \right. \\ &\left.(22.25 - 0.5\log_{10}(h_U))\log_{10}(d_{ak})\right) + \varsigma_{ak}.\end{aligned} \quad (7)$$

Here f_c and PL_{f_c} respectively denote the carrier frequency and corresponding free-space path loss, and ς_{ak} defines the shadow fading component whose standard deviation is δ_{ak}^U.

Additionally, considering the typical UAV operating scenarios, the frame duration is much smaller than the coherence interval caused by Doppler frequency shift. Thus, this work neglects the Doppler effect.

Different from the aerial propagation environment, a plurality of buildings or obstacles would severely disrupt the radio propagation on the ground. As a result, TUEs could experience rich-scattered multipath (numerous NLoS links) and severe pathloss. Similar to literature, this paper characterizes the TUE channels through Ricean fading model with variable Rician factor, where the appearing probability of the LoS link can be modeled as [35]

$$P_{ak}^{LoS} = \begin{cases} \frac{300-d_{ak}}{300}, 0 < d_{ak} \le 300 \\ 0, \quad d_{ak} > 300 \end{cases} \quad (8)$$

with d_{ak} being the distance between the ath AP and the kth TUE.

Similarly, the channel between the kth TUE and the ath AP, denoted by $\mathbf{g}_{k,a}$, is consequently expressed as

$$\mathbf{g}_{k,a} = \sqrt{\frac{\beta_{k,a}}{K_a^{(k)}+1}}\left[\sqrt{K_a^{(k)}}e^{j\vartheta_{k,a}}\mathbf{a}_{k,a} + \mathbf{h}_{k,a}\right], \quad (9)$$

where $\beta_{k,a}$, $K_a^{(k)}$, $\vartheta_{k,a}$, $\mathbf{a}_{k,a}$ and $\mathbf{h}_{k,a}$ stand for large-scale fading coefficient, Rician factor, phase offset, LoS steering

vector and scattered random component, respectively. In addition, these scalars and vectors can be obtained via similar definitions or computing methods as UAV channels.

Especially, to jointly capture pathloss and shadowing fading, $\beta_{k,a}$ can be given as

$$\beta_{ak} = \begin{cases} -30.9 - 26\log_{10}(d_{ak}) + \varsigma_{ak}, P_{ak}^{LoS} \neq 0 \\ -34.53 - 38\log_{10}(d_{ak}) + \varsigma_{ak}, P_{ak}^{LoS} = 0 \end{cases}. \quad (10)$$

Here ς_{ak} models the shadow fading with standard deviation δ_{ak}^{T}.

From (3), it can be noted that the presence or absence of the LoS component mainly depends on the distance from the target TUE to the APs. Correspondingly, the TUE channels would be described as Ricean or Rayleigh fading model. Obviously, this mixture of Rician-Rayleigh fading channel model would drastically complicate the performance analysis and optimal design. However, the considered model is more generic and practical, which can be regarded as a valuable reference for network design.

B. Uplink training

For a general cell-free massive MIMO system working in TDD mode, the APs usually need to acquire the CSI for data detection via uplink channel estimation. Generally, all users would transmit their own pilot sequences to the APs in the uplink training phase. For clarity, both UAVs and TUEs are regarded as generic users. Typically, orthogonality among pilots is hardly achieved because of the limited predesigned pilot sequences. Thus, pilot reuse is often unavoidable in many practical situations, which would cause pilot contamination and impair the system performance. Besides, in many commercial wireless networks, pilot sequences are predesigned and repeatedly utilized. Thus, as an intelligent eavesdropper, it is not hard for the active Eve to obtain the knowledge of pilot signals and jam the training. Considering the deleteriousness of active eavesdropping, this paper concerns the pilot spoofing attack scenario, where the Eve synchronously transmits identical pilot signal as that of the target legitimate user in the uplink training phase.

In general, let $\boldsymbol{\phi}_i \in \mathbb{C}^{\tau_p \times 1}$ and $\boldsymbol{\phi}_E \in \mathbb{C}^{\tau_p \times 1}$ be the normalized training sequence used by kth user and the Eve respectively, with $\|\boldsymbol{\phi}_i\| = 1$ and $\|\boldsymbol{\phi}_E\| = 1$. Also, different pilot sequences are assumed to be mutually orthogonal, i.e. $\boldsymbol{\phi}_i^H \boldsymbol{\phi}_j = 0$ if $\boldsymbol{\phi}_i \neq \boldsymbol{\phi}_j$. For convenience, this work defines $\mathcal{K}_U$, $\mathcal{K}_T$ and $\mathcal{K} = \mathcal{K}_U \cup \mathcal{K}_T$ with cardinalities K_U, K_T and $K = K_U + K_T$ being the sets of UAVs, TUEs and all serving users. Assuming the presence of an active Eve, this work assumes Eve attempts to overhear the information intended for kth user, i.e. $\boldsymbol{\phi}_E = \boldsymbol{\phi}_k$. Then, the received signal of the ath AP in the uplink phase can be expressed as

$$\mathbf{Y}_a = \sum_{i \in \mathcal{K}} \sqrt{\tau_p p_u} \mathbf{g}_{i,a} \boldsymbol{\phi}_i^H + \sqrt{\tau_p p_E} \mathbf{g}_{E,a} \boldsymbol{\phi}_E^H + \mathbf{W}_a, \quad (11)$$

where p_u and p_E respectively represent the uplink pilot power employed by the users and Eve, and $\mathbf{W}_a \in \mathbb{C}^{N \times \tau_p}$ corresponds to additive white Gaussian noise (AWGN) with independent and identically distributed (i.i.d) $\mathcal{CN}\left(0, \delta_w^2\right)$ elements.

To perform channel estimations, each AP should project the observable signal matrix $\mathbf{Y}_a$ onto $\boldsymbol{\phi}_i$ to obtain one post-processing statistics as

$$\hat{\mathbf{y}}_{k,a} = \mathbf{Y}_a \boldsymbol{\phi}_k = \sum_{i \in \mathcal{K}} \sqrt{\tau_p p_u} \mathbf{g}_{i,a} \boldsymbol{\phi}_i^H \boldsymbol{\phi}_k + \sqrt{\tau_p p_E} \mathbf{g}_{E,a} \boldsymbol{\phi}_E^H \boldsymbol{\phi}_k + \mathbf{W}_a \boldsymbol{\phi}_k, \quad (12)$$

To avoid the information exchange among APs, each AP could locally perform the minimum mean-square error (MMSE) estimation approach, which yields that [36]

$$\hat{\mathbf{g}}_{k,a} = \mathbf{D}_{k,a} \hat{\mathbf{y}}_{k,a}, \quad (13)$$

with

$$\mathbf{D}_{k,a} = \sqrt{\tau_p p_u} \mathbf{G}_{k,a} \mathbf{B}_{k,a}^{-1} \in \mathbb{C}^{N \times N},$$

$$\mathbf{G}_{k,a} = \frac{\beta_{k,a}}{K_{k,a} + 1} \left[K_{k,a} \mathbf{a}_{k,a} \mathbf{a}_{k,a}^H + \mathbf{I}_N \right],$$

$$\mathbf{B}_{k,a} = \sum_{i \in \mathcal{K}} \tau_p p_u \beta_{i,a} \mathbf{G}_{i,a} \left| \boldsymbol{\phi}_i^H \boldsymbol{\phi}_k \right|^2 + \tau_p p_E \beta_{E,a} \mathbf{G}_{E,a} \left| \boldsymbol{\phi}_E^H \boldsymbol{\phi}_k \right|^2 + \delta_w^2 \mathbf{I}_N.$$

Subsequently, according to the MMSE approach property, the expectation of the norm-square of estimated channel vector can be described as follows

$$\gamma_{k,a} = \mathbb{E}\left\{ \hat{\mathbf{g}}_{k,a}^H \hat{\mathbf{g}}_{k,a} \right\} = \sqrt{\tau_p p_u} \mathrm{tr}\left(\mathbf{G}_{k,a} \mathbf{D}_{k,a} \right), \quad (14)$$

For quantitatively evaluating the performance of the estimator, one parameter named normalized minimum square error (NMSE) can be defined as follows

$$E_N = 1 - \frac{1}{MK} \sum_{a=1}^{M} \sum_{k \in \mathcal{K}} \frac{\gamma_{k,a}}{N \beta_{k,a}}. \quad (15)$$

Assuming the pilots are evenly allocated to users, the cardinality of the users set using the identical pilot sequence is defined as reuse factor K_p. From (13) and (14), it can be noted that E_N decreases with respect to the reuse factor and the Eve's power, respectively. For the legitimate users, one straightforward method to improve the estimator's accuracy is to increase pilot power p_u. However, due to the pilot contamination, there will still be a non-zero floor of E_N even with unlimited pilot power in noise-free channels.

C. Downlink Data Transmission

With the estimated local channels via uplink training, the APs could perform precoding to transmit the downlink data to intended users. To avoid information interaction among APs and obtain tractable expressions for performance analysis, this work adopts the distributed conjugate beamforming technique referring to various existing researches. By denoting the data intended for kth user as s_k with $\mathbb{E}\left\{ |s_k|^2 \right\} = 1$, the transmitted vector at ath AP can thus be established by

$$\mathbf{x}_a = \sum_{k \in \mathcal{K}_a} \sqrt{\eta_{k,a}} \hat{\mathbf{g}}_{k,a} s_k \quad (16)$$

with $\eta_{k,a}$ denoting the power control coefficient at the ath AP corresponding to the kth user, and $\mathcal{K}_a$ representing the set of the users served by ath AP.

With the UC approach, one AP generally only provides service for partial users. Letting K_a be the cardinality of $\mathcal{K}_a$, that is, $K_a \leq M$. In general, the users with best propagation conditions can be selected as the service targets at one given AP. Referring to literature [36], this paper utilizes the large-scale fading coefficients to appraise the channel transmission quality, that is, one AP only serves the users with smallest large-scale loss. Additionally, the special case $K_a = M$ denotes the conventional cell-free architecture that each AP simultaneously serves all users.

Meanwhile, taking into account the maximum power constraint at each AP, it is needed to impose one critical restrictive condition on the power control coefficients, which can be described as follows

$$\mathbb{E}\left\{\left\|\mathbf{x}_a\right\|^2\right\} = \sum_{k\in\mathcal{K}_a} \eta_{k,a}\gamma_{k,a} \leq p_d. \tag{17}$$

Here p_d represents the power constraint at each AP.

Subsequently, the signal received by kth user and Eve can be respectively written as

$$r_k = \sum_{a\in\mathcal{A}} \mathbf{g}_{k,a}^H \mathbf{x}_a + z_k \tag{18}$$

and

$$r_E = \sum_{a\in\mathcal{A}} \mathbf{g}_{E,a}^H \mathbf{x}_a + z_E. \tag{19}$$

Here $\mathcal{A}$ denotes the set of all APs, and $z_k \sim \mathcal{CN}\left(0, \delta_k^2\right)$ and $z_E \sim \mathcal{CN}\left(0, \delta_E^2\right)$ indicate the AWGN at the kth user and the Eve, respectively.

In order to clearly distinguish desired signal, interference and noise, the receipt signals can be mathematically reformulated as

$$r_k = \sum_{a\in\mathcal{A}_k} \sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{k,a} s_k + \sum_{i\in\mathcal{K}\setminus k} \sum_{a\in\mathcal{A}} \sqrt{\eta_{i,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{i,a} s_i + z_k, \tag{20}$$

$$r_E = \sum_{a\in\mathcal{A}_k} \sqrt{\eta_{k,a}}\mathbf{g}_{E,a}^H \hat{\mathbf{g}}_{k,a} s_k + \sum_{i\in\mathcal{K}\setminus k} \sum_{a\in\mathcal{A}} \sqrt{\eta_{i,a}}\mathbf{g}_{E,a}^H \hat{\mathbf{g}}_{i,a} s_i + z_E, \tag{21}$$

where $\mathcal{A}_j$ $(j \in \mathcal{K})$ represents the set of the APs that provides service for the ith user.

III. Secrecy Performance Merits

This section concentrates on the secrecy performance analysis on this considered network while deriving some tight tractable closed-form expressions. According to available literature, ergodic secrecy rate is always regarded as the secrecy performance metric in various delay-tolerant scenarios, where the coding can be performed within several coherence intervals. Thereby, the achievable ergodic secrecy rate is bounded by the difference of achievable information of target user and the eavesdropper, which can be defined as

$$R_{\text{sec}}^k = \left[R_k - R_{E,k}\right]^+. \tag{22}$$

Here R_k represents the achievable information rate of kth user and $R_{E,k}$ denote the information leakage to the Eve seeking to intercept the data of kth user.

Prior to performing detailed analysis and implementing further countermeasures, it needs to respectively investigate the achievable data rates of target user and the Eve, which are described as follows in the following.

A. Achievable rate of the legitimate user

With the employed transmission protocol, it is almost unlikely for the users to acquire the perfect CSI. To yield insightful results, this work takes into account a general realistic case that all legitimate users utilize statistical CSI instead of instantaneous CSI to decode the desired data. In consequence, the received signal at the kth user can be represented as

$$r_k = D_k \cdot s_k + B_k \cdot s_k + \sum_{i\in\mathcal{K}\setminus k} I_{i,k} \cdot s_i + z_k. \tag{23}$$

where

$$D_k = \mathbb{E}\left\{\sum_{a\in\mathcal{A}_k} \sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{k,a}\right\},$$

$$B_k = \sum_{a\in\mathcal{A}_k} \sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{k,a} - \mathbb{E}\left\{\sum_{a\in\mathcal{A}_k} \sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{k,a}\right\},$$

$$I_{i,k} = \sum_{a\in\mathcal{A}} \sqrt{\eta_{i,a}}\mathbf{g}_{k,a}^H \hat{\mathbf{g}}_{i,a}.$$

Evidently, the terms D_k, B_k and $I_{i,k}$ respectively signify the strength of the desired signal, beamforming gain uncertainty and interference brought by ith user. Notably, it can be demonstrated that these terms and noise component z_k are pairwise uncorrelated. On the other hand, the terms approximately follow Gaussian distribution, especially when MN is extremely large.

By considering the worst case that all terms are Gaussian distributed, the lower bound for the achieved information rate at the kth user can be given by

$$R_k = \frac{\tau_d}{\tau_c}\log_2\left(1 + SINR_k\right). \tag{24}$$

Here $SINR_k$ denotes the signal-to-interference-plus-noise ratio (SINR) of the kth user, whose expression is listed as follows.

$$SINR_k = \frac{\left|D_k\right|^2}{\mathbb{E}\left\{\left|B_k\right|^2\right\} + \sum\limits_{j\in\mathcal{K}\setminus k} \mathbb{E}\left\{\left|I_{j,k}\right|^2\right\} + \delta_k^2}. \tag{25}$$

Additionally, these terms in (25) are deterministic via computing over various independent channel realizations. However, the resulting closed-form expression of is always unavailable due to huge computational complexity. In general cases, the approximation of $SINR_k$ can be presented by means of Monte Carlo simulations. Specifically, by virtue of random matrix theory and general bounding technique [36], the tight lower bound on $SINR_k$ is presented in Theorem 1.

Theorem 1: Taking into account both the MMSE channel estimation and distributed conjugate beamforming, the deterministic expression of the lower bound on $SINR_k$ can be

$$SINR_k = \left(\sum_{a \in \mathcal{A}_k} \sqrt{\eta_{k,a}} \gamma_{k,a} \right)^2 \cdot \left\{ \sum_{a \in \mathcal{A}_k} \eta_{k,a} \left(\tau_p p_k^u \lambda_{k,a}^{(k)} - \gamma_{k,a}^2 \right) + \sum_{j \in \mathcal{K}} \sqrt{\tau_p p_j^u} \sum_{a \in \mathcal{A}_j} \eta_{j,a} \mathrm{tr} \left(\mathbf{G}_{j,a} \mathbf{D}_{j,a}^H \mathbf{G}_{k,a} \right) + \delta_k^2 + \sum_{j \in \mathcal{K} \setminus k} \tau_p p_k^u \left| \boldsymbol{\phi}_k^H \boldsymbol{\phi}_j \right|^2 \sum_{a \in \mathcal{A}_j} \left(\eta_{j,a} \lambda_{k,a}^{(j)} + \sum_{b \in \mathcal{A}_j, b \neq a} \sqrt{\eta_{j,a} \eta_{j,b}} \mathrm{tr} \left(\mathbf{D}_{j,a} \mathbf{G}_{k,a} \right) \mathrm{tr} \left(\mathbf{D}_{j,b} \mathbf{G}_{k,b} \right) \right) \right\}^{-1} . \quad (26)$$

derived as (26) shown on the top of next page with

$$\lambda_{i,a}^{(j)} = \left(\frac{\beta_{i,a}}{K_{i,a}+1} \right)^2 \mathrm{tr}^2 \left(\mathbf{D}_{j,a} \right) + 2K_{i,a} \left(\frac{\beta_{i,a}}{K_{i,a}+1} \right)^2 \cdot \mathrm{R} \left\{ \mathrm{tr} \left(\mathbf{a}^H \left(\theta_{i,a} \right) \mathbf{D}_{j,a} \mathbf{a} \left(\theta_{i,a} \right) \mathbf{D}_{j,a}^H \right) \right\} . \quad (27)$$

Proof: Please refer to Appendix A.

B. Information rate leaked to the Eve

Different from the cooperative legitimate users, it is practically impossible for the network and the APs to acquire the Eve's perfect knowledge, such as accurate CSI, signal processing capability, information perception level, etc. Significantly, it is more crucial for secrecy appraisal and network planning design to cognizance the greatest threat caused by the evesdroppers. With common assumptions in many prior literature on secure cell-free massive MIMO network [25], [26], [27], [28], [29], this work takes into account the worst-case scenario that Eve can decode the desired confidential data by using its instantaneous CSI.

Thereby, the received signal r_E is represented as

$$r_E = D_{E,k} \cdot s_k + \sum_{j \in \mathcal{K} \setminus k} I_{E,j} \cdot s_j + z_E, \quad (28)$$

where $z_E \sim \mathcal{CN} \left(0, \delta_E^2 \right)$ stands for AWGN component, $D_{E,k}$ and $I_{E,j}$ respectively indicate the strength of desired symbol and inter-user interference given by

$$D_{E,k} = \sum_{a \in \mathcal{A}_k} \sqrt{\eta_{k,a}} \mathbf{g}_{E,a}^H \hat{\mathbf{g}}_{k,a}, \quad (29)$$

$$I_{E,j} = \sum_{a \in \mathcal{A}_j} \sqrt{\eta_{j,a}} \mathbf{g}_{E,a}^H \hat{\mathbf{g}}_{j,a}. \quad (30)$$

By treating the inter-user interference and AWGN as effective noise, the corresponding information rate obtained by Eve can be expressed as

$$R_{E,k} = \frac{\tau_d}{\tau_c} \log_2 \left(1 + SINR_{E,k} \right) . \quad (31)$$

Here $SINR_{E,k}$ denotes the achievable average SINR on Eve.

Besides, desired signal, inter-user interference and AWGN components approximately obey Gaussian distribution according to central limit theorem. Thus, $SINR_{E,k}$ is formulated as

$$SINR_{E,k} = \frac{\mathbb{E} \left\{ \left| D_{E,k} \right|^2 \right\}}{\sum_{j \in \mathcal{K} \setminus k} \mathbb{E} \left\{ \left| I_{E,j} \right|^2 \right\} + \delta_E^2}. \quad (32)$$

On the basis of the stringent assumptions, the upper bound on can be provided via a series of mathematical operations, which is shown in Theorem 2 below.

Theorem 2: Considering the case that Eve can detect its desired data by exploiting the instantaneous CSI, the deterministic expression of $SINR_{E,k}$ is provided as (33) listed on the top of next page, where $\lambda_{E,a}^{(k)}$ has the similar definition as

$$\lambda_{E,a}^{(j)} = \left(\frac{\beta_{E,a}}{K_{E,a}+1} \right)^2 \mathrm{tr}^2 \left(\mathbf{D}_{j,a} \right) + 2K_{E,a} \left(\frac{\beta_{E,a}}{K_{E,a}+1} \right)^2 \cdot \mathrm{R} \left\{ \mathrm{tr} \left(\mathbf{a}^H \left(\theta_{E,a} \right) \mathbf{D}_{j,a} \mathbf{a} \left(\theta_{E,a} \right) \mathbf{D}_{j,a}^H \right) \right\} .$$

Proof: Please see Appendix B.

IV. Pilot Assignment Optimization

According to the previous description, pilot reuse would inevitably yield coherent interference as well as channel estimation error in uplink training, which also leads to deterioration of network throughput. Fortunately, optimizing pilot assignment scheme is regarded as an effective pilot contamination control technique, and has been widely investigated in recent years. Notably, most current literatures solve this challenging issue based on positions, system throughput or network fairness [9], [10], [11], [12]. However, these metrics cannot accurately quantify the severity of pilot contamination. Interestingly, the work [13] described a pilot assignment strategy based on weighted graphic framework (WGF), in which a novel metric is established to define the effects of pilot contamination. Nevertheless, the proposed algorithm in [13] only can acquire a local optimization solution. Hence, this paper aims to investigate an improved pilot assignment algorithm on the basis of WGF. For simplicity, the proposed algorithm is abbreviated to IWGF.

To begin with, it is crucial to construct an effective metric to appraise the potential interference among users. Let the kth user be the target user, and the mth AP be one of the serving AP, that is, $m \in \mathcal{A}_k$. Inspired by the previous analysis, it is noted that the available gain is proportional to $|\beta_{k,m}|^2$, while the interference caused by the k'th user is approximately proportional to $|\beta_{k',m}|^2$. Thus, the corresponding quantity $|\beta_{k',m}/\beta_{k,m}|^2$ can be utilized to measure the interference caused by the k'th user. Referring to literature, the potential effect of pilot contamination between the kth user and the k'th user can be indicated by this following defined factor

$$\omega_{k,k'} = \left| \frac{\sum_{m \in \mathcal{A}_k} \beta_{k',m}}{\sum_{m \in \mathcal{A}_k} \beta_{k,m}} \right|^2 + \left| \frac{\sum_{m' \in \mathcal{A}_{k'}} \beta_{k,m'}}{\sum_{m' \in \mathcal{A}_{k'}} \beta_{k',m'}} \right|^2 . \quad (34)$$

Obviously, this defined parameter can be computed with a low complexity. Besides, as the quantity depends on the large-scale coefficients and serving AP set, it can dynamically describe the interference of the corresponding network

$$SINR_{E,k}=\left(\begin{array}{l}\sum\limits_{a\in\mathcal{A}_k}\eta_{k,a}\left(\tau_p p_E\lambda_{E,a}^{(k)}+\sqrt{\tau_p p_k^u}\mathrm{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{E,a}\right)\right)\\ +\tau_p p_E\sum\limits_{a\in\mathcal{A}_k}\sum\limits_{b\in\mathcal{A}_k,b\neq a}\sqrt{\eta_{k,a}\eta_{k,b}}\mathrm{tr}\left(\mathbf{D}_{k,a}\mathbf{G}_{E,a}\right)\mathrm{tr}\left(\mathbf{D}_{k,b}\mathbf{G}_{E,b}\right)\end{array}\right)\cdot$$
$$\left\{\sum_{j\in\mathcal{K}\setminus k}\sum_{a\in\mathcal{A}_j}\left[\begin{array}{l}\tau_p p_E\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2\left(\eta_{j,a}\lambda_{E,a}^{(j)}+\sum\limits_{b\in\mathcal{A}_j,b\neq a}\sqrt{\eta_{j,a}\eta_{j,b}}\mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{E,a}\right)\mathrm{tr}\left(\mathbf{D}_{j,b}\mathbf{G}_{E,b}\right)\right)\\ +\eta_{j,a}\sqrt{\tau_p p_j^u}\mathrm{tr}\left(\mathbf{G}_{j,a}\mathbf{D}_{j,a}^H\mathbf{G}_{E,a}\right)\end{array}\right]+\delta_E^2\right\}^{-1}. \quad (33)$$

Algorithm 1 WGF Algorithm

Input: system parameters: number of users K, number of pilots ζ, AP selection scheme $\mathcal{A}_k$, $\forall i\in\mathcal{K}$, large scale fading β_{aj}, $\forall a\in\mathcal{A}$, $\forall j\in\mathcal{K}$.
Output: optimized pilot assignment sets $\{V_1,\cdots,V_\zeta\}$.
Begin:
(1) Initialize: arbitrarily choose ζ users and assign different user to each set V_j, $j=1,2,\cdots,\zeta$; let $O_{un}=\{U_k:1\leq k\leq K-\zeta\}$ be the set of unassigned user.
(2) While $O_{un}\neq\emptyset$ do
(3) Arbitrarily choose $U_k\in O_{un}$
(4) For $1\leq i\leq\zeta$ do
(5) Calculate $S_{k,i}=\sum\limits_{U_{k'}\in V_i}\omega_{k,k'}$
(6) Update $V_{i'}=V_{i'}+U_k$ with $i'=\arg\min\limits_i S_{k,i}$ and O_{un}=O_{un}−U_k.
(7) End while
(8) Return the eventual sets $\{V_1,V_2,\cdots,V_\zeta\}$ as the pilot assignment scheme.

topology. Moreover, the parameters are symmetrical, that is, $\omega_{k,k'}$=$\omega_{k',k}$.

Methodologically, the essential problem of pilot assignment is how to assign the limited pilot sequences according to some optimality criterion. Let u_i denotes the ith user. Assuming $\{V_1,V_2,\cdots,V_\zeta\}$ being the pilot allocation scheme, the set V_i represents the set of the legitimate users with the same pilot sequence, i.e. $\boldsymbol{\phi}_p=\boldsymbol{\phi}_q$ if $u_p,u_q\in V_i$. Herein, ζ stands for the number of the pilots. Intuitively, the aim of a pilot assignment scheme is to alleviate pilot contamination. Then, it is reasonable to establish the optimization objective function on the basis of the potential interference. Particularly, the pilot assignment scheme can be optimized to maximize the total weights among the disjoint sets or to minimize the total weights within the same sets. Accordingly, the pilot allocation issue is mathematically formulated as [13]

$$\arg\max_{\{V_1,V_2,\cdots,V_\zeta\}}\sum_{1\leq p<q\leq\zeta}\sum_{u_i\in V_p,u_j\in V_q}\omega_{ij} \quad (35)$$

or

$$\arg\min_{\{V_1,V_2,\cdots,V_\zeta\}}\sum_{1\leq p\leq\zeta}\sum_{u_i,u_j\in V_p,i\neq j}\omega_{ij}. \quad (36)$$

Referring to [32], one heuristic algorithm can be adopted to solve the optimization problem, which is named WGF algorithm for the sake of simplicity. Then, the procedure of the WGF algorithm is introduced in Algorithm 1.

It is noted that the WGF method can avoid the exhaustive search and has a low computational complexity. However, this simple search method can only obtain a local optimum solution. To seek a better solution, it is crucial to implement some iteration or update approaches to break away from the local optimum values. As well known, the genetic algorithm is suitable for tackling the resource scheduling problem via global search. Motivated by this consideration, this paper attempts to provide one novel pilot allocation scheme by integrating the WGF method with a genetic algorithm [37], [38]. To achieve a faster convergence rate, partial solutions of WGF algorithm can be chosen as the initial generation of the genetic algorithm. By defining and comparing desired fitness values, some genetic operations including selection, crossover and mutation can be performed to generate next generation population. Then, by implementing several iterations, a near-optimal solution can be obtained. The proposed pilot assignment algorithm, called Improved WGF (IWGF) algorithm is described in Algorithm 2 as follows.

Algorithm 2 Proposed IWGF algorithm

(1) Use the Algorithm 1 to generate several initial pilot assignment solutions, and do encoding to initialize the population (establishing mapping relationship between the indexes of user and assigned pilot);
(2) Calculate and evaluate the fitness values: total weights of the inter-sets according to (35) or total weights among the intra-sets according to (36).
(3) Do selection: rank the fitness values and compute the probability of corresponding gene code.
(4) Perform genetic operations including copy, crossover and mutation.
(5) Update the population with the new derived scheme.
(6) Repeat the steps 2-5 until the number of iterations reaches the preset value or the objective objective fitness value converges.
(7) Return the best obtained pilot assignment sets as the optimal solution.

V. Power Control Optimization

Using the derived closed-form tractable expressions, this section concentrates on max-min fairness power allocation problem. That is, this work should formulate and solve one optimization problem which can provide optimal power control coefficients that can maximize the minimum achievable rate of the legitimate users. For a practical secure transmission

network, two additional critical conditions should be imposed to the optimization problem: power constraint on each AP and greatest threshold of information leakage to the Eve.

Therefore, the max-min optimization problem under power and secrecy constraints can be formulated as follows

$$\begin{aligned} &\max_{\eta_{k,a}} \min_{i\in\mathcal{K}} R_i \\ &\text{s.t.} \sum_{i\in\mathcal{K}_a} \eta_{i,a}\gamma_{i,a} \le p_d, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad \eta_{i,a} \ge 0, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad R_{E,k} \le \bar{R}_{E,k}^T \end{aligned}, \tag{37}$$

where $\bar{R}_{E,k}^T$ is the predetermined threshold for Eve's information rate.

Considering that both R_k and $R_{E,k}$ are monotonically increasing nonnegative function with respect to corresponding SINRs, the optimization framework (37) can be reformulated as follows

$$\begin{aligned} \text{P1}: &\max_{\eta_{k,a}} \min_{i\in\mathcal{K}} SINR_i \\ &\text{s.t.} \sum_{i\in\mathcal{K}_a} \eta_{i,a}\gamma_{i,a} \le p_d, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad \eta_{i,a} \ge 0, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad SINR_{E,k} \le \bar{\theta}_E \end{aligned} \tag{38}$$

Here $\bar{\theta}_E$ represents the corresponding required threshold of $SINR_{E,k}$, which can be straightforwardly achieved via exponentiation computation to $\bar{R}_{E,k}^T$.

To facilitate analysis, this paper defines a sequence of variables as follows.

$$u_{k,a} = \sqrt{\eta_{k,a}}, \forall k \in \mathcal{K}, a \in \mathcal{A}. \tag{39}$$

Besides, by introducing another slack variable t, the power allocation problem P1 can be converted to a more tractable framework as follows.

$$\begin{aligned} \text{P2}: &\max_{u_{i,a}} t \\ &\text{s.t.} \quad SINR_i \ge t, \forall i \in \mathcal{K} \\ &\quad \sum_{i\in\mathcal{K}_a} u_{i,a}^2\gamma_{i,a} \le p_d, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad u_{i,a} \ge 0, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad SINR_{E,k} \le \bar{\theta}_E \end{aligned} \tag{40}$$

Note that the objective function of P2 is affine and the transmit power constraints are convex. Obviously, the problem P2 is non-concave, which cannot be solved directly through existing optimization approaches. Moreover, the similar max-min fairness problems have been investigated and efficiently solved in several existing works. Inspired by the research in [18], [20], a fractional optimization algorithm along with successive convex approximation is proposed to solve this non-convex optimization problem.

To begin with, suitable convex approximation approaches should be performed firstly to tackle the non-convexity issue. To facilitate description and analysis, the troublesome constraints of SINRs at the legitimate user can be represented as

$$SINR_i = \frac{\left(\sum_{a\in\mathcal{A}_i} u_{i,a}\gamma_{i,a}\right)^2}{\varphi_i(\mathbf{\Psi})}, \forall i \in \mathcal{K}, \tag{41}$$

where matrix $\mathbf{\Psi}$ incorporates all power control variables whose (i,a) th $(\forall i \in \mathcal{K}, a \in \mathcal{A})$ element is $u_{i,a}$, and $\varphi_i(\mathbf{\Psi})$ denotes the corresponding denominator component of $SINR_i$. Notably, functions $\varphi_i(\mathbf{\Psi})$ are obviously convex with regard to variables $u_{i,a}$, $\forall i \in \mathcal{K}, a \in \mathcal{A}$.

For positive variables x and y as well as designative values $\bar{x}$ and $\bar{y}$, there exists an inequality as follows [25]

$$\frac{x^2}{y} \ge 2\frac{\bar{x}}{\bar{y}}x - \frac{\bar{x}^2}{\bar{y}^2}y. \tag{42}$$

As such, assuming $\mathbf{\Psi}^{(k)}(u_{i,a}^{(k)}, \forall i \in \mathcal{K}, a \in \mathcal{A})$ is a given feasible point of P2, the concave lower-bounds on $SINR_i$ can be achieved as

$$SINR_i \ge f_i^{(k)}(\mathbf{\Psi}) = a^{(k)} \sum_{a\in\mathcal{A}_i} u_{i,a}\gamma_{i,a} - b^{(k)}\varphi_i(\mathbf{\Psi}) \tag{43}$$

with $a^{(k)} = 2\sum_{a\in\mathcal{A}_i} u_{i,a}^{(k)}\gamma_{i,a} \Big/ \varphi_i\left(\mathbf{\Psi}^{(k)}\right)$ and $b^{(k)} = \left(\frac{a^{(k)}}{2}\right)^2$.

For clarity, $SINR_{E,k}$ is similarly rewritten as

$$SINR_{E,k} = \frac{g_E(\mathbf{\Psi})}{\varphi_E(\mathbf{\Psi})} \tag{44}$$

with

$$\begin{aligned} g_E(\mathbf{\Psi}) = &\sum_{a\in\mathcal{A}_k} \eta_{k,a}\left(\tau_p p_E \lambda_{E,a}^{(k)} + \sqrt{\tau_p p_k^u}\text{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{E,a}\right)\right) \\ &+ \tau_p p_E \sum_{a\in\mathcal{A}_k}\sum_{b\in\mathcal{A}_k, b\ne a} \sqrt{\eta_{k,a}\eta_{k,b}}\text{tr}\left(\mathbf{D}_{k,a}\mathbf{G}_{E,a}\right)\text{tr}\left(\mathbf{D}_{k,b}\mathbf{G}_{E,b}\right) \end{aligned}$$

and $\varphi_E(\mathbf{\Psi})$ represents the corresponding denominator in $SINR_{E,k}$.

Recalling the characteristic that any convex function is lower-bounded by its first-order Taylor expansion, convex function $\varphi_E(\mathbf{\Psi})$ would satisfy the following inequality [25]

$$\begin{aligned} \varphi_E(\mathbf{\Psi}) \ge &\zeta_E(\mathbf{\Psi}) = \varphi_E\left(\mathbf{\Psi}^{(k)}\right) + \\ &\sum_{i\in\mathcal{K}, a\in\mathcal{A}} \frac{\partial\varphi_E(\mathbf{\Psi})}{\partial u_{i,a}}\Big|_{\mathbf{\Psi}^{(k)}} \left(u_{i,a} - u_{i,a}^{(k)}\right), \end{aligned} \tag{45}$$

where $\mathbf{\Psi}^{(k)}$ denotes one feasible point of $\varphi_E(\mathbf{\Psi})$.

With the approximations (43) and (45), the max-min optimization problem can be approximately transformed to the following tractable form

$$\begin{aligned} \text{P3}: &\max_{u_{i,a}} t \\ &\text{s.t.} \quad f_i^{(k)}(\mathbf{\Psi}) \ge t, \forall i \in \mathcal{K} \\ &\quad \sum_{i\in\mathcal{K}_a} u_{i,a}^2\gamma_{i,a} \le p_d, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad u_{i,a} \ge 0, \forall i \in \mathcal{K}, a \in \mathcal{A} \\ &\quad g_E(\mathbf{\Psi}) \le \bar{\theta}_E\zeta_E(\mathbf{\Psi}) \end{aligned} \tag{46}$$

Obviously, the sub-problem P3 can be easily solved via available convex optimization techniques. Consequently,

TABLE I
SETUP OF SOME KEY SYSTEM PARAMETERS

Parameter	Value
Carrier frequency	1.9GHz
Height of the APs	10m
Height of the UAVs	$22.5\text{m} \sim 300\text{m}$
Height of the TUEs	1.65m
Number of UAVs and TUEs	$K_U = K_T = K/2$
Uplink power on each user	p_u=100mW
Transmit power of the Eve	p_E=100mW
Downlink Power of each AP	p_d=100mW
Coherence length	$\tau_c = 100$
Duration of the uplink training	τ_p=10
Thermal noise	-174dBm/Hz
Noise figure	Transmit signal power at 9dB
Network architectures	cell-free (CF) or user-centric (UC)

the optimized power allocation strategy is conceived via a fractional programme, which needs to incorporate successive convex approximation and sequential iteration optimization. For the sake of clarity, the proposed optimization procedure is summarized in Algorithm 3. Notably, the optimization problem (P3) involves MK_U real variables. Besides, there are also ε=$M + K_U + 1$ quadratic constraints in (P3). Thus, the per-iteration complexity for solving problem (P3) is $\mathcal{O}\left((MK_U)^2(\varepsilon-1)^{2.5} + (\varepsilon-1)^{3.5}\right)$ [25], [35].

Algorithm 3 Framework of the Iterative Optimization Algorithm

(1) Initialization: Choose feasible $u_{i,a}, \forall i \in \mathcal{K}, a \in \mathcal{A}$ and formulate feasible point $\mathbf{\Psi}^{(0)}$; Set $k = 0$, $t_{\min}$ and $t_{\max}$.
(2) Repeat
(3) Set $t = (t_{\min} + t_{\max})/2$, then solve the problem P3.
(4) While P3 has the feasible solution $\tilde{\mathbf{\Psi}}$, do that: $k = k+1$, $\mathbf{\Psi}^{(k)} = \tilde{\mathbf{\Psi}}$ and $t_{\min} = t$; otherwise make the change that $t_{\max} = t$.
(5) Perform the iteration until convergence that $t_{\max} - t_{\min} \leq \varepsilon$, where ε is the designative threshold.
(6) Return $\mathbf{\Psi}^{(k)}$ as the desired power allocation scheme.

VI. NUMERICAL RESULTS

This section provides various representative numerical studies to validate the derived results and proposed algorithms. It is noteworthy that the simulations are performed by considering the secure communicant network shown in Fig. 1, where a cell-free massive MIMO system simultaneously provides service for several aerial UAVs and ground terrestrial terminals and one UAV-Eve aims to wiretap the confidential data for one target UAV user. Similar to prior works [27], [28], [35], [36], the simulations take into account a practical network setup that all APs, UAVs and TUEs are randomly located in a $1\text{km}\times1\text{km}$ square area, while aerial UAVs would fly to different heights. Unless otherwise stated, some critical parameters used in simulations are listed in Table I.

To begin with, Fig. 2 presents the achievable secrecy rate comparison between the "analytical values (Ana.)" and "simulated results (Sim.)" under different networks configurations. In particular, the results are obtained via 2000 different Monte Carlo realizations. Additionally, unless considering the power optimization, all APs exploit the fractional power control rule, that is, $\eta_{k,a}$=$p_d \Big/ \sum_{k\in\mathcal{K}_a} \gamma_{k,a}$. It is obvious that the curves in Fig. 2 could well confirm the tightness and accuracy of the achievable expression of secrecy rate, which also imply that the derived results can be used for secrecy performance evaluation. Further, due to the deleteriousness of the pilot spoofing attack, the secrecy rates decrease with the increase in Eve's power, which is consistent with expectations. It is also observed that secrecy performance would be dramatically deteriorated with a larger pilot reuse factor K_p. This phenomenon evidently accounts for the effects of severe pilot contamination.

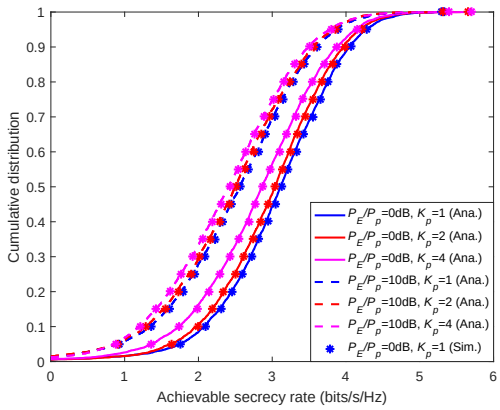


Fig. 2. Simulated and analytical cumulative distribution of the achievable secrecy rates with different pilot power and K_p.

Fig. 3 and Fig. 4 are devoted to evaluate the effects of both antenna configurations and network architectures, respectively. By comparing the cumulative distribution of the secrecy rates, the curves in Fig. 3 demonstrates that the network can reap remarkable performance improvement by employing multi-antenna APs. It can be explained that system can obtain increasing multiplexing gain and diversity gain from multiple antennas configuration. Moreover, the results in Fig. 3 also indicate that the system would undergo a considerable performance degradation in "with TUEs" scenario. That is because that the existence of TUEs would immensely increase the inter-user interference as well as deteriorate the achievable secrecy rate. From Fig. 4, it is easy to note that the achievable secrecy rate is an increasing function with respect to the number of APs M. As expected, it is coincident with some existing findings. Specifically, the curves in Fig. 4 also confirm that the UC deployment with "$A_k = 10$" generally outperforms the CF architecture. The reason is that limited power resources would be allocated to users with best propagation conditions in UC approach, which can yield a better directional beamforming and less interference power. Notably, as the ratio A_k/M increases, the average achievable secrecy rate firstly increases, and then decreases. Thus, it is needed to make tradeoff between the complexity and performance for practical system design.

Next, Fig. 5 shows the impact of UAV's and Eve's Heights on network secrecy performance. Note that the results indicate

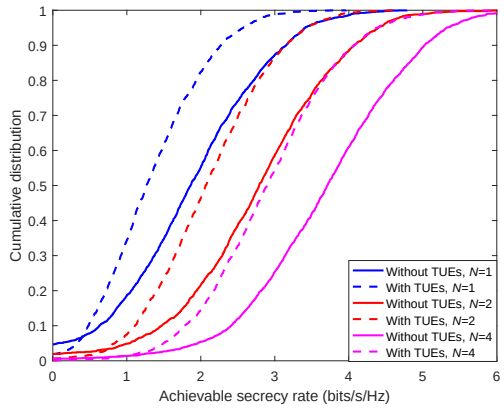


Fig. 3. Cumulative distribution of the achievable secrecy rate against different N.

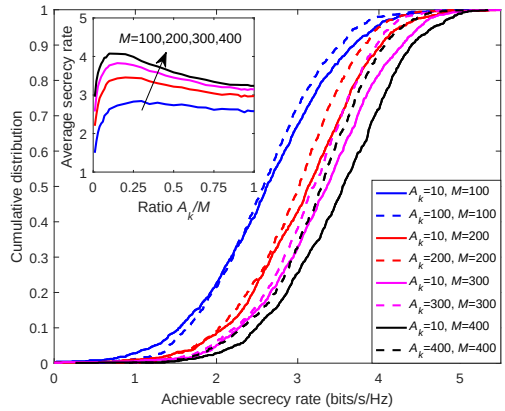


Fig. 4. Comparison of the achievable secrecy rate with user-centric (UC) or cell-free (CF) approach against different M.

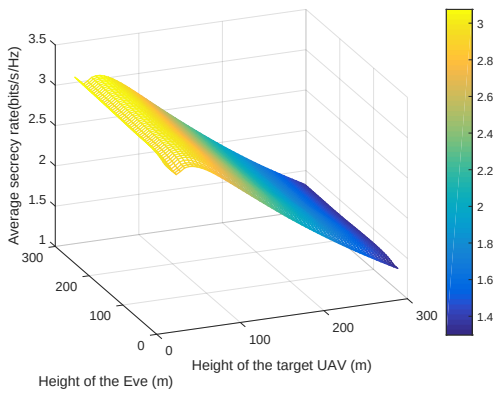


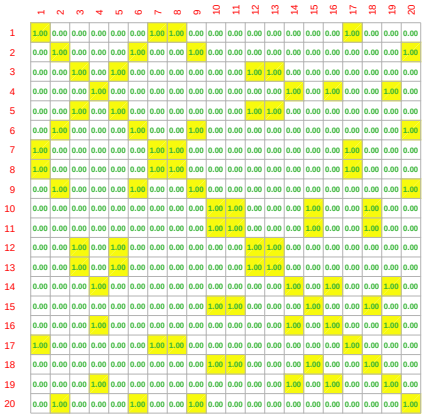
Fig. 5. Achievable average secrecy rate versus target UAV's height and UAV-Eve's height.

that the secrecy rate is scarcely affected by varying the heights of the Eve-UAV. This is because of the strong-LoS characteristic of high-altitude UAV channels, which leads to a less beamforming offset than that of the conventional ground propagation. Particularly, Fig. 5 reflects that secrecy performance is mainly restricted by the heights of the legitimate UAVs. For high-altitude UAVs, the results reveal that the achievable average secrecy rate gradually degrades with the increasing of legitimate UAVs' height. In fact, this phenomenon is consistent with expecting due to the enlarged path loss. However, the variation characteristic is distinct at lower height, where the secrecy rate curve increases first and then decreases. In fact, these changes are due to the variation of P_{ak}^{LoS} and $K_a^{(k)}$. Overall, the secrecy performance is primarily determined by the characteristics of propagation channels.

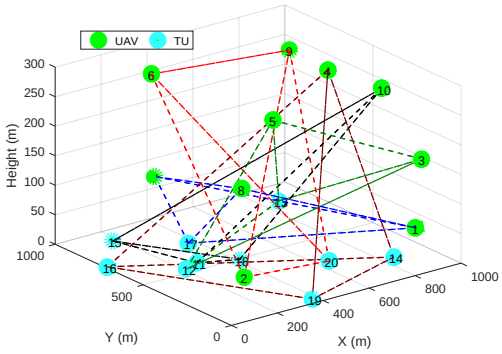
Considering the determined network layout, the four subfigures in Fig. 6 present the pilot assignment schemes and pilot reuse graphs of the WGF and IWGF algorithms, where the numbers "1" (yellow squares) in subfigure Fig. 6(a) and Fig. 6(c) denote "identical pilot while the numbers "0" stand for orthogonal pilots. For clarity, it is noted that the figures Fig. 6(b) and Fig. 6(d) are drawn using the true spatial coordinates of APs. According to formulas (7) and (10), it is known that the large-scale fading coefficients are mainly determined by distance between the two nodes. To some degree, both WGF and IWGF algorithms are regarded as location-based method. Notably, since the performance of WGF approach is constrained by the initial selection, this algorithm can only achieve a local optimal solution. Instead, to avoid stable local convergence, the proposed IWGF algorithm aims to seek the global optimum by adding iterative procedure based on genetic algorithm. Interestingly, these figures can clearly demonstrate that the proposed IWGF approach can effectively avert pilot reuse among nearby users, such as User 6 and User 9, User 16 and User 19, User 11 and User 18, etc.

Pilot contamination would enlarge channel estimation error and eventually degrade the achievable network performance. Thus, it is more appropriate to use the achievable information rate as the metric to appraise the pilot assignment schemes. Subsequently, with the same setup in Fig.6, the corresponding achievable rates of the users are depicted in Fig. 7(a). Note that the network with random pilot assignment method has the worst fairness among the users. Then WGF scheme can obviously enhance the network performance. Apparently, the results show the promising superiority of the proposed IWGF algorithm compared to the other two schemes. Furthermore, Fig. 7(b) draws the cumulative distribution curves of the achievable minimum rate and secrecy rate over 2000 independent realizations. As expected, the performance gaps among the solid curves in Fig. 7(b) also indicate that proposed IWGF scheme outperforms than random scheme and WGF scheme. As depicted in Fig. 7(b), it is seen that the achievable secrecy rates with IWGF and WGF schemes are slightly higher than that of random scheme. In fact, the slight performance improvement can be explained that both the WGF and IWGF algorithms only utilize the information of legitimates users as well as ignore the effect caused by Eve. Interestingly, integration fairness and secrecy in pilot assignment can be left for future work.

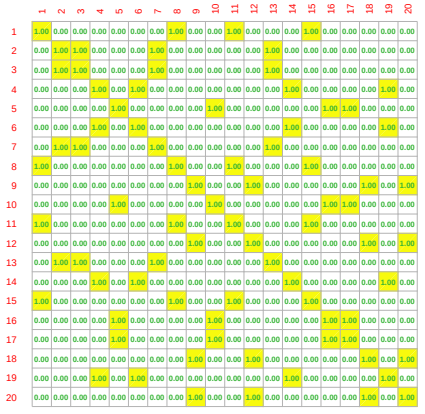
Fig. 8 shows the achievable information rate of users and Eve with and without power control (PC). For clarity, the simulations are performed with following setups: 1) Total number of users is $K = 10$; 2) Indexes U1-U5 stand for the UAVs while indexes U6-U10 represent the TUEs. Also, U1 is assumed to be the target user. 3) The upper-bound constraint Eve's rate is set as $\bar{R}_{E,k}^{T}$=0.75. To make an effective



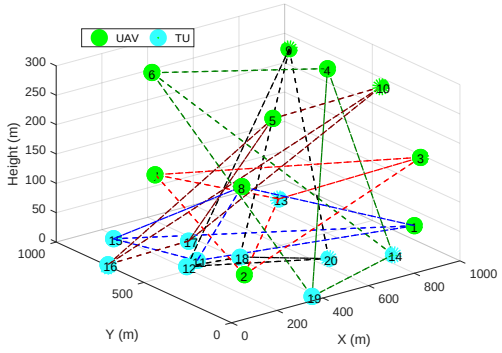
(a) Pilot assignment of WGF algorithm.



(b) Pilot reuse graph of WGF algorithm.

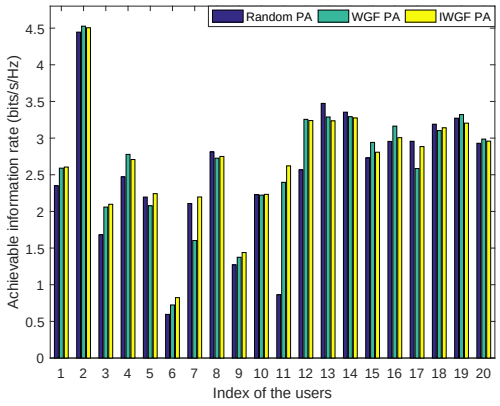


(c) Pilot assignment of IWGF algorithm.

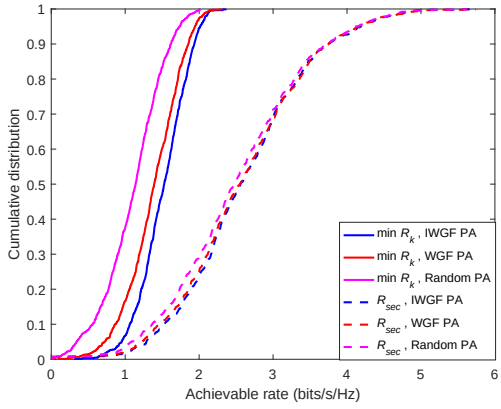


(d) Pilot reuse graph of IWGF algorithm.

Fig. 6. Comparisons of WGF algorithm and the proposed IWGF algorithm.



(a)



(b)

Fig. 7. Comparisons of achievable rates with WGF algorithm and the proposed IWGF algorithm.

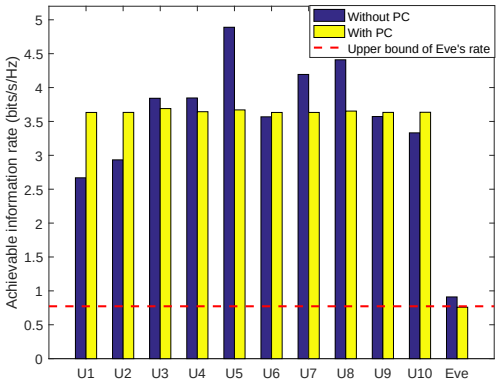


Fig. 8. Comparison of achievable information rates with/without power control.

benchmark, all APs transmit downlink data with full power and utilize proportional power allocation rule. The two cases are numerically compared in Fig. 7, which indicates that the improvement for fairness brought from power control technique is remarkable. The results indicate that using the power control makes the variation of achievable rates obviously becoming smaller. Also, it is observed that the proposed approach can ensure that information leakage to Eve would not exceed the predetermined target threshold.

Moreover, Fig. 9 presents the cumulative distribution of the

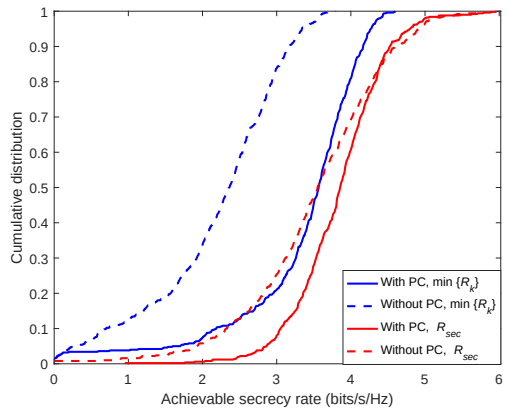


Fig. 9. Cumulative distribution of achievable minimum-rate and secrecy rate with/wihout power control.

achievable minimum rate and secrecy rate with or without power control operation. Remarkably, the curves in Fig. 9 illustrate that proposed power control strategy can effectively improve both the system fairness and security, which show that network can achieve both higher minimum-rate and secrecy rate by employing the proposed power control scheme. Thus, the comparisons between the two cases in Fig. 8 and Fig. 9 can validate the effectiveness and superiority of the provided power control scheme under considered network layout. Taking into consideration that pilot assignment and power control have been respectively verified via numerical results, this section has not performed further verification by combining with these two procedures due to the space limitations.

VII. Conclusion

This paper has considered the secure UAV communication in UC cell-free massive MIMO network with the presence of one UAV acting as active eavesdropper. By exploiting the MMSE estimation and beamforming techniques in the uplink and downlink phases, the closed-form tractable expression of the lower bound for the achievable secrecy rate has been derived to measure the system performance. On the basis of these analytical results, pilot assignment and power control problems have been respectively studied to guarantee the security and network fairness. First, with the weighted graphic framework approach, the genetic algorithm was utilized to iteratively search the globally optimal pilot assignment scheme. Then, a max-min optimization framework has been formulated to optimize the power allocation coefficients. Moreover, the corresponding power allocation issue has been tackled via successive convex approximation and fractional optimization. Finally, correctness of the analysis and effectiveness of the proposed algorithms have been demonstrated through numerical results.

Appendix

A. Proof of Theorem 1

In order to get the deterministic closed-form expression of $SINR_k$, it is necessary to analyze the compute the terms D_k, B_k and $I_{j,k}$, respectively.

To begin with, the term D_k can be reformulated as

$$D_k = \mathbb{E}\left\{\sum_{a\in\mathcal{A}_k}\sqrt{\eta_{k,a}}(\hat{\mathbf{g}}_{k,a}+\tilde{\mathbf{g}}_{k,a})^H\hat{\mathbf{g}}_{k,a}\right\}, \tag{47}$$

where $\tilde{\mathbf{g}}_{k,a} = \mathbf{g}_{k,a}-\hat{\mathbf{g}}_{k,a}$ stands for the estimation error vector.

According to the characteristics of MMSE approach, both estimate vector $\hat{\mathbf{g}}_{k,a}$ and error vector $\tilde{\mathbf{g}}_{k,a}$ are Gaussian distributed and mutually independent. In particular, it is noted that is an zero mean random vector.

Therefore, the first term can be calculated as follows

$$D_k = \mathbb{E}\left\{\sum_{a\in\mathcal{A}_k}\sqrt{\eta_{k,a}}\hat{\mathbf{g}}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right\} = \sum_{a\in A_k}\sqrt{\eta_{k,a}}\gamma_{k,a}. \tag{48}$$

Next, it needs to focus on the computation of term $\mathbb{E}\left\{|B_k|^2\right\}$, which can be denoted as

$$\mathbb{E}\left\{|B_k|^2\right\} = \mathbb{E}\left\{\left|\sum_{a\in\mathcal{A}_k}\sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a} - \mathbb{E}\left\{\sum_{a\in\mathcal{A}_k}\sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right\}\right|^2\right\}. \tag{49}$$

Exploiting the independence among the different channel vectors, the term can be simplified into the sum of several expectations, which can be shown as follows.

$$\mathbb{E}\left\{|B_k|^2\right\} = \sum_{a\in\mathcal{A}_k}\eta_{k,a}\left(\mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\} - \left|\mathbb{E}\left\{\sum_{a\in\mathcal{A}_k}\sqrt{\eta_{k,a}}\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right\}\right|^2\right). \tag{50}$$

Thus, the computation of $\mathbb{E}\left\{|B_k|^2\right\}$ is simplified to evaluating the expectations $\mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\}$, which gives rise to following result.

$$\mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\} = \mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\mathbf{D}_{k,a}\cdot\left(\sum_{i\in\mathcal{K}}\sqrt{\tau_p p_i^u}\mathbf{g}_{i,a}\boldsymbol{\phi}_i^H\boldsymbol{\phi}_k + \sqrt{\tau_p p_E}\mathbf{g}_{E,a}\boldsymbol{\phi}_E^H\boldsymbol{\phi}_k + \mathbf{w}_{a,k}\right)\right|^2\right\}. \tag{51}$$

with $\mathbf{w}_{a,k} = \mathbf{W}_a\boldsymbol{\phi}_k \sim \mathcal{CN}\left(\mathbf{0}, \delta_w^2\mathbf{I}_N\right)$.

Capitalizing on the assumptions that different channel vectors are mutually independent, it is straightforward to obtain the following expression.

$$\mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\mathbf{D}_{k,a}\mathbf{g}_{k,a}\right|^2\right\} = \lambda_{k,a}^{(k)} + \mathrm{tr}\left(\mathbf{D}_{k,a}\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{k,a}\right). \tag{52}$$

Here the definition of notation $\lambda_{k,a}^{(k)}$ is presented in (27).

Furthermore, it is easy to derive the following expression.

$$\mathbb{E}\left\{\left|\mathbf{g}_{k,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\} = \tau_p p_k^u\lambda_{k,a}^{(k)} + \sqrt{\tau_p p_k^u}\mathrm{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{k,a}\right). \tag{53}$$

Then, inserting (52) and (53) into (50) yields

$$\mathbb{E}\left\{|B_k|^2\right\} = \sum_{a\in\mathcal{A}_k}\eta_{k,a}\left(\tau_p p_k^u \lambda_{k,a}^{(k)} + \sqrt{\tau_p p_k^u}\mathrm{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{k,a}\right) - \gamma_{k,a}^2\right). \tag{54}$$

Utilizing the similar approach in [26], the term $\mathbb{E}\left\{|I_{k,j}|^2\right\}$ can be recast as

$$\begin{aligned}\mathbb{E}\left\{|I_{k,j}|^2\right\} &= \tau_p p_k^u \mathbb{E}\left\{\left|\sum_{a\in\mathcal{A}_j}\sqrt{\eta_{j,a}}\mathbf{g}_{k,a}^H\mathbf{D}_{j,a}\mathbf{g}_{k,a}\right|\right\}\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2 \\ &+ \sum_{a\in\mathcal{A}_j}\sum_{i\in\{\mathcal{K}\setminus k,E\}}\eta_{j,a}\tau_p p_i^u \mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{i,a}\mathbf{D}_{j,a}^H\mathbf{G}_{k,a}\right)\left|\boldsymbol{\phi}_i^H\boldsymbol{\phi}_j\right|^2 \\ &+ \delta_w^2\sum_{a\in\mathcal{A}_j}\eta_{j,a}\mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{D}_{j,a}^H\mathbf{G}_{k,a}\right).\end{aligned} \tag{55}$$

Then, it is not hard to calculate the expectation $\mathbb{E}\left\{\left|\sum_{a\in\mathcal{A}_j}\sqrt{\eta_{j,a}}\mathbf{g}_{k,a}^H\mathbf{D}_{j,a}\mathbf{g}_{k,a}\right|\right\}$ via applying the similar procedure as in [26], [29], which is presented as follows.

$$\begin{aligned}&\mathbb{E}\left\{\left|\sum_{a\in\mathcal{A}_j}\sqrt{\eta_{j,a}}\mathbf{g}_{k,a}^H\mathbf{D}_{j,a}\mathbf{g}_{k,a}\right|\right\} = \\ &\sum_{a\in\mathcal{A}_j}\eta_{j,a}\left(\lambda_{k,a}^{(j)} + \mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{k,a}\mathbf{D}_{j,a}^H\mathbf{G}_{k,a}\right)\right) + \\ &\sum_{a\in\mathcal{A}_j}\sum_{b\in\mathcal{A}_j,b\neq a}\sqrt{\eta_{j,a}\eta_{j,b}}\mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{k,a}\right)\mathrm{tr}\left(\mathbf{D}_{j,b}\mathbf{G}_{k,b}\right).\end{aligned} \tag{56}$$

Subsequently, plugging (56) into (55) can obtain the following expression.

$$\begin{aligned}&\mathbb{E}\left\{|I_{k,j}|^2\right\} = \\ &\sum_{a\in\mathcal{A}_j}\eta_{j,a}\sqrt{\tau_p p_j^u}\mathrm{tr}\left(\mathbf{G}_{j,a}\mathbf{D}_{j,a}^H\mathbf{G}_{k,a}\right) + \tau_p p_k^u\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2\cdot\sum_{a\in\mathcal{A}_j} \\ &\left(\eta_{j,a}\lambda_{k,a}^{(j)} + \sum_{b\in\mathcal{A}_j/a}\sqrt{\eta_{j,a}\eta_{j,b}}\mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{k,a}\right)\mathrm{tr}\left(\mathbf{D}_{j,b}\mathbf{G}_{k,b}\right)\right).\end{aligned} \tag{57}$$

Finally, integrating the derived results in (48), (54) and (57), the closed-form expression of can be calculated and presented as (26).

B. Proof of Theorem 2

According to (32), achieving the closed-form expression of $SINR_{E,k}$ is equivalent to derive the deterministic equivalents of $\mathbb{E}\left\{|D_{E,k}|^2\right\}$ and $\mathbb{E}\left\{|I_{E,j}|^2\right\}$, respectively. Most notably, different channel vectors are assumed to be mutually independent. As a result, the term $\mathbb{E}\left\{|D_{E,k}|^2\right\}$ can be presented as follows:

$$\begin{aligned}\mathbb{E}\left\{|D_{E,k}|^2\right\} &= \sum_{a\in\mathcal{A}_k}\eta_{k,a}\mathbb{E}\left\{\left|\mathbf{g}_{E,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\} + \tau_p p_E\cdot \\ &\sum_{a\in\mathcal{A}_k}\sum_{b\in\mathcal{A}_k,b\neq a}\sqrt{\eta_{k,a}\eta_{k,b}}\mathrm{tr}\left(\mathbf{D}_{k,a}\mathbf{G}_{E,a}\right)\mathrm{tr}\left(\mathbf{D}_{k,b}^H\mathbf{G}_{E,b}\right)\end{aligned} \tag{58}$$

Next, by using the similar manipulations in Theorem 1, it is straightforward to rewrite the term $\mathbb{E}\left\{\left|\mathbf{g}_{E,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\}$ as follows.

$$\begin{aligned}&\mathbb{E}\left\{\left|\mathbf{g}_{E,a}^H\hat{\mathbf{g}}_{k,a}\right|^2\right\} = \\ &\mathbb{E}\left\{\left|\mathbf{g}_{E,a}^H\mathbf{D}_{k,a}\left(\sum_{i\in\mathcal{K}}\sqrt{\tau_p p_i^u}\mathbf{g}_{i,a}\boldsymbol{\phi}_i^H\boldsymbol{\phi}_k + \sqrt{\tau_p p_E}\mathbf{g}_{E,a} + \mathbf{w}_{a,k}\right)\right|^2\right\} \\ &= \tau_p p_E\lambda_{E,a}^{(k)} + \sqrt{\tau_p p_k^u}\mathrm{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{E,a}\right).\end{aligned} \tag{59}$$

By inserting (59) into (58), the corresponding closed-form expression of $\mathbb{E}\left\{|D_{E,k}|^2\right\}$ can be presented as follows.

$$\begin{aligned}&\mathbb{E}\left\{|D_{E,k}|^2\right\} = \\ &\sum_{a\in\mathcal{A}_k}\eta_{k,a}\left(\tau_p p_E\lambda_{E,a}^{(k)} + \sqrt{\tau_p p_k^u}\mathrm{tr}\left(\mathbf{G}_{k,a}\mathbf{D}_{k,a}^H\mathbf{G}_{E,a}\right)\right) + \\ &\tau_p p_E\sum_{a\in\mathcal{A}_k}\sum_{b\in\mathcal{A}_k,b\neq a}\sqrt{\eta_{k,a}\eta_{k,b}}\mathrm{tr}\left(\mathbf{D}_{k,a}\mathbf{G}_{E,a}\right)\mathrm{tr}\left(\mathbf{D}_{k,b}^H\mathbf{G}_{E,b}\right).\end{aligned} \tag{60}$$

Subsequently, the interference caused by jth user ($j\neq k$) can be computed as

$$\begin{aligned}&\mathbb{E}\left\{|I_{E,j}|^2\right\} = \sum_{a\in\mathcal{A}_j}\eta_{j,a}\sqrt{\tau_p p_j^u}\mathrm{tr}\left(\mathbf{G}_{j,a}\mathbf{D}_{j,a}^H\mathbf{G}_{E,a}\right) + \\ &\tau_p p_E\left|\boldsymbol{\phi}_k^H\boldsymbol{\phi}_j\right|^2\cdot\sum_{a\in\mathcal{A}_j}\left\{\eta_{j,a}\lambda_{E,a}^{(j)} + \right. \\ &\left.\sum_{b\in\mathcal{A}_j,b\neq a}\sqrt{\eta_{j,a}\eta_{j,b}}\mathrm{tr}\left(\mathbf{D}_{j,a}\mathbf{G}_{E,a}\right)\mathrm{tr}\left(\mathbf{D}_{j,b}\mathbf{G}_{E,b}\right)\right\}.\end{aligned} \tag{61}$$

Finally, plugging (60) and (61) into (32) can yield the deterministic equivalents as (33), which completes the proof.

REFERENCES

[1] H. Q. Ngo, L. N. Tran, T. Q. Duong, et al., "On the Total Energy Efficiency of Cell-Free Massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25-39, Mar. 2018.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, et al., "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834-1850, Jan. 2017.

[3] A. Papazafeiropoulos, P. Kourtessis, M. D. Renzo, et al., "Performance Analysis of Cell-Free Massive MIMO Systems: A Stochastic Geometry Approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3523-3537, Apr. 2020.

[4] G. Femenias and F. Riera-Palou, "Wideband Cell-Free mmWave Massive MIMO-OFDM: Beam Squint-Aware Channel Covariance-Based Hybrid Beamforming," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4695-4710, Jul. 2022.

[5] T. K. Nguyen, H. H. Nguyen and H. D. Tuan, "Max-Min QoS Power Control in Generalized Cell-Free Massive MIMO-NOMA With Optimal Backhaul Combining," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 10949-10964, Oct. 2020.

[6] A. A. Polegre, L. Sanguinetti and A. G. Armada, "Pilot Decontamination Processing in Cell-Free Massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3990-3994, Dec. 2021.

[7] H. A. Ammar, R. Adve, S. Shahbazpanahi, et al., "User-Centric Cell-Free Massive MIMO Networks: A Survey of Opportunities, Challenges and Solutions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 611-652, 1st Quart. 2022.

[8] Y. Ming, Z. Sha, Y. Dong, et al., "Downlink Resource Allocation With Pilot Length Optimization for User-Centric Cell-Free MIMO Networks," *IEEE Commun. Lett.*, vol. 26, no. 11, pp. 2705-2709, Nov. 2022.

[9] M. Attarifar, A. Abbasfar and A. Lozano, "Random vs Structured Pilot Assignment in Cell-Free Massive MIMO Wireless Networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Kansas City, MO, USA, May 2018, pp. 1-6.

[10] Y. Zhang, H. Cao, P. Zhong, et al., "Location-Based Greedy Pilot Assignment for Cell-Free Massive MIMO Systems," in *Proc. IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2018, pp. 392-396.

[11] S. Buzzi, C. D'Andrea, M. Fresia, et al., "Pilot Assignment in Cell-Free Massive MIMO Based on the Hungarian Algorithm," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 34-37, Jan. 2021.

[12] H. Liu, J. Zhang, S. Jin and B. Ai, "Graph Coloring Based Pilot Assignment for Cell-Free Massive MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9180-9184, Aug. 2020.

[13] W. Zeng, Y. He, B. Li and S. Wang, "Pilot Assignment for Cell Free Massive MIMO Systems Using a Weighted Graphic Framework," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6190-6194, Jun. 2021.

[14] Y. Zeng, J. Lyu and R. Zhang, "Cellular-Connected UAV: Potential, Challenges, and Promising Technologies," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 120-127, Feb. 2019.

[15] M. Ghamari, P. Rangel, M. Mehrubeoglu, et al., "Unmanned Aerial Vehicle Communications for Civil Applications: A Review," *IEEE Access*, vol. 10, pp. 102492-102531, Sept. 2022.

[16] A. Fotouhi, H. Qiang, M. Ding, et al., "Survey on UAV Cellular Communications: Practical Aspects, Standardization Advancements, Regulation, and Security Challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417-3442, 4th Quart. 2019.

[17] W. Mei, Q. Wu and R. Zhang, "Cellular-Connected UAV: Uplink Association, Power Control and Interference Coordination," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5380-5393, Nov. 2019.

[18] N. Yang, L. Wang, G. Geraci, et al., "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20-27, Apr. 2015.

[19] O. A. Topal, M. O. Demir, Z. Liang, et al., "A Physical Layer Security Framework for Cognitive Cyber-Physical Systems," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 32-39, Aug. 2020.

[20] D. Kapetanovic, G. Zheng and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21-27, Jun. 2015.

[21] J. Tang, H. Wen, K. Zeng, et al., "Light-Weight Physical Layer Enhanced Security Schemes for 5G Wireless Networks," *IEEE Netw.*, vol. 33, no. 5, pp. 126-133, Sept.-Oct. 2019.

[22] Y. Wu, R. Schober, D. W. K. Ng, et al., "Secure Massive MIMO Transmission With an Active Eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3880-3900, Jul. 2016.

[23] X. Zhang, D. Guo, K. An, et al., "Secure Transmission in Multi-Pair AF Relaying Massive MIMO Networks Against Active Pilot Spoofing Attacks," *IEEE Access*, vol. 7, pp. 3547-3560, Jan. 2019.

[24] J. Wang, J. Lee, F. Wang, et al., "Jamming-Aided Secure Communication in Massive MIMO Rician Channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6854-6868, Dec. 2015.

[25] T. M. Hoang, H. Q. Ngo, T. Q. Duong, et al., "Cell-Free Massive MIMO Networks: Optimal Power Control Against Active Eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724-4737, Oct. 2018.

[26] X. Zhang, D. Guo, K. An, et al., "Secrecy Analysis and Active Pilot Spoofing Attack Detection for Multigroup Multicasting Cell-Free Massive MIMO Systems," *IEEE Access*, vol. 7, pp. 57332-57340, May 2019.

[27] X. Zhang, D. Guo, K. An and B. Zhang, "Secure Communications Over Cell-Free Massive MIMO Networks With Hardware Impairments," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1909-1920, Jun. 2020.

[28] X. Zhang, T. Liang, K. An, et al., "Secure Transmission in Cell-Free Massive MIMO With RF Impairments and Low-Resolution ADCs/DACs," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8937-8949, Sept. 2021.

[29] X. Zhang, T. Liang and K. An, "Secrecy performance analysis of cell-free massive MIMO in the presence of active eavesdropper with low resolution ADCs," *Wireless Netw.*, vol. 27, pp. 4839-4852, Sept. 2021.

[30] X. Zhang, T. Liang, K. An, et al., "Secure Transmission in RIS-Assisted Cell-free Massive MIMO system with Low Resolution ADCs/DACs," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop*, Austin, TX, USA, Apr. 2022, pp. 339-344.

[31] S. Elhoushy, M. Ibrahim and W. Hamouda, "Exploiting RIS for Limiting Information Leakage to Active Eavesdropper in Cell-Free Massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 443-447, Mar. 2022.

[32] S. J. Maeng, Y. Yapıcı, İ. Güvenç, et al., "Precoder Design for Physical-Layer Security and Authentication in Massive MIMO UAV Communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2949-2964, Mar. 2022.

[33] X. Zhang, Y. Chen, Y. Zhang, et al., "Analysis of Secure User-Centric Cell-Free Massive MIMO Network for Multi-UAV Communications," in *2022 IEEE International Conference on Unmanned Systems (ICUS)*, Guangzhou, China, Oct. 2022, pp. 744-750.

[34] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706-709, Dec. 2017.

[35] V. Tentu, E. Sharma, D. N. Amudala, et al., "UAV-Enabled Hardware-Impaired Spatially Correlated Cell-Free Massive MIMO Systems: Analysis and Energy Efficiency Optimization," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2722-2741, Apr. 2022.

[36] C. D' Andrea, A. Garcia-Rodriguez, G. Geraci, et al., "Analysis of UAV Communications in Cell-Free Massive MIMO Systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 133-147, Jan. 2020.

[37] L. Jiacheng and L. Lei, "A Hybrid Genetic Algorithm Based on Information Entropy and Game Theory," *IEEE Access*, vol. 8, pp. 36602-36611, Mar. 2020.

[38] A. Dockhorn and S. Lucas, "Choosing Representation, Mutation, and Crossover in Genetic Algorithms," *IEEE Comput. Intell. Mag.*, vol. 17, no. 4, pp. 52-53, 1 Nov. 2022.