On Pilot Spoofing Attack in Massive MIMO Systems: Detection and Countermeasure

Weiyang Xu, *Member, IEEE*, Chang Yuan, Shengbo Xu, Hien Quoc Ngo, *Member, IEEE*, Wei Xiang, *Senior Member, IEEE*

***Abstract*—Massive MIMO systems are vulnerable to pilot spoofing attacks (PSAs) since the estimated channel state information can be contaminated by the eavesdropping link, thus incurring severe information leakage in downlink transmission. To safeguard legitimate communications, this paper proposes a PSA detection method which relies on pilot manipulation. Specifically, users randomly partition pilot sequences into two parts, where the first part remains unchanged and the second one is multiplied with a diagonal matrix. Although a malicious node may follow the same way to send pilots, this makes it more likely to be detected. According to the principle of the likelihood-ratio test, the proposed detector is designed based on a decision metric that does not include the legitimate channel. This feature differentiates our scheme from existing ones and remarkably improves the detection accuracy. Besides, the possibility of performance enhancement by joint detection is discussed. Furthermore, based on pilot manipulation, a jamming-resistant receiver is designed. The key of this receiver is a new channel estimator that is robust to the PSA. Finally, extensive simulations are carried out to validate our proposed algorithms.**

***Index Terms*—Massive MIMO, physical layer security, channel estimation, pilot spoofing attack, secrecy rate.**

I. Introduction

To meet a worldwide growing throughput demand is a major challenge of contemporary wireless communication systems. An effective strategy for increasing spectral efficiency is to deploy a large number of antennas at the base stations (BSs), while sharing the same time-frequency resources [1], [2]. Such systems, referred to as massive multiple-input multiple-output (MIMO), can provide very high spectral and energy efficiency when the number of BS antennas is large, and have recently received a great deal of attention [3].

Wireless communications are vulnerable to eavesdropping due to the broadcasting nature of the wireless medium. Recently, physical layer security has become one of the research hotspots recently in providing secure communications [4]. Rather than high level cryptographic methods, physical layer security employs information-theoretic security and signal processing techniques [5]. In general, passive and active attacks are two major threats to legitimate communications. Massive MIMO systems can dramatically boost security against passive eavesdropping, thanks to its capability to focus the transmission energy in the direction of legitimate users [6]. However, if an eavesdropper launches active attacks, then the achievable secrecy rate will be dramatically reduced [7]. For example, the channel state information (CSI), which is crucial to exploit benefits of massive MIMO, can be estimated by sending pilots ahead of actual data transmission [8]. However, this provides opportunity for a malicious node to launch attack. By sending the same pilots as legitimate users, the eavesdropping link can contaminate the channel estimate, resulting in severe information leakage in downlink transmission [9]. Such mechanism, referred to as pilot spoofing attack (PSA), was first documented in [10] and has received a great deal of attention since then.

To improve the reliability of data transmission, the authors in [11] propose to counteract the effect of active attacks by exploiting the artificial noise. Besides, a jamming detection scheme based on random matrix theory is introduced in [12], where the final decision is made by analyzing the maximum eigenvalue of the sample covariance matrix of the received signal. According to the generalized likelihood-ratio test, a detection method is designed for the uplink of massive MIMO, where unused orthogonal pilots are employed [13], [14]. With the intention of detecting the PSA, the authors of [15] propose a detector that takes advantage of the asymmetry of received signal power levels at the transmitter and legitimate receiver. Moreover, by examining the pilot contamination in the uplink and downlink, a pilot retransmission scheme is designed for jamming detection [16]. Recently, by designating an auxiliary node as a trusted user, an efficient three-phase uplink training method is designed, with which malicious users can be reliably detected [17]. In [18], the transmitter sends pilots to the receiver, then the receiver sends the conjugate of its received signal back to the transmitter, where the final decision on detection is made. More recently, a detection method is designed based on the fact that channel estimation results would be different from its original because of the PSA [19].

In addition to detection algorithms, countermeasures against the PSA are also crucial. In [20], a random training strategy is proposed, where each user is allocated with multiple pilot sequences and then randomly selects one pilot sequence each time to confuse the attacker. Moreover, authors in [21] employ unused pilots to estimate the legitimate and jamming channels simultaneously, and then the estimate of the jamming channel is used to construct linear filters that reject the impact of the jamming signal. Legitimate and eavesdropping channel

W. Y. Xu, C. Yuan and S. B. Xu are with the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, 400044, P. R. China (E-mails: {weiyangxu, 20144098, shengboxu}@cqu.edu.cn).

H. Q. Ngo is with the Institute of Electronics, Communications and Information Technology (ECIT), Queens University Belfast, BT3 9DT, Belfast, U.K., (E-mail: hien.ngo@qub.ac.uk).

W. Xiang is with the School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (E-mail: w.xiang@latrobe.edu.au).

estimation and secure beamforming are discussed in [22] by using the temporal subspace of the pilot signal. More recently, the receive power-to-noise ratio based PSA detection scheme is proposed in [23], and the defense strategy is discussed based on the attackers optimal power allocation. By exploiting the difference in channel estimation in two training phases, a double channel training based scheme is proposed to combat the PSA and uplink jamming simultaneously [24].

In this paper, we propose a new PSA detection scheme relying on pilot manipulation. Then, a jamming-resistant receiver, of which the core component is a channel estimator robust to the PSA, is designed. The main contributions of this paper are summarized as follows.

- We employ pilot manipulation to facilitate PSA detection. Specifically, in the uplink training phase, legitimate users partition pilot sequences into two parts, with the first part kept unchanged and the second part multiplied with a diagonal matrix. The possibility of a malicious node to employ the same manipulation to send pilots is discussed;
- According to the principle of the likelihood-ratio test (LRT), a new PSA detector is designed based on a decision metric that does not include the legitimate channel. This feature enables to remarkably improve the detection accuracy. Performance analyses on the probabilities of false alarm (P_{fa}) and detection (P_d) are carried out. Moreover, joint detection by multiple users is proposed to further enhance system security;
- A jamming-resistant receiver is designed given the aforementioned pilot manipulation. The core component is a new channel estimator robust to eavesdropping. Both analytical and numerical results demonstrate that the proposed receiver is robust to the PSA at the expense of a slightly increased noise power.

The remainder of this paper is organized as follows. The system model and problem formulation are described in Section II. Pilot manipulation is detailed in Section III. The possibility of a malicious node to employ pilot manipulation is analyzed in Section IV. Section V presents the proposed PSA detection algorithm, while Section VI introduces a new jamming-resistant receiver. Numerical results are presented to validate the detection scheme and countermeasure in Section VII. Finally, concluding remarks are drawn in Section VIII.

$Notation$: $\mathbb{C}^{n\times m}$ and $\mathbb{R}^{n\times m}$ denote complex and real matrices of size $n\times m$, respectively. Bold variables represent matrices and vectors. $\mathcal{CN}(\mu,\sigma^2)$ and $\mathcal{N}(\mu,\sigma^2)$ denote complex and real Gaussian distributions of mean μ and variance σ^2, respectively. $\mathbb{E}\{\cdot\}$ and var$\{\cdot\}$ indicate the mean and variance operators, respectively. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ are taken to mean the transpose, conjugate transpose and complex conjugate operators, respectively. $\mathcal{R}\{\cdot\}$ and $\mathcal{I}\{\cdot\}$ refer to the real and imaginary parts of a complex number. Finally, erf$(\cdot)$ represents the error function.

II. System Model and Problem Formulation

In this section, the system model is first described. Next, the impact of the PSA on legitimate communications is discussed.

A. System Model

This paper considers a massive MIMO system, where the BS (*Alice*) employs M antennas to communicate with K single-antenna users (*Bobs*) ($M \gg K$). Denote by $\mathbf{H} = [\sqrt{\beta_{h,1}}\mathbf{h}_1,\dots,\sqrt{\beta_{h,K}}\mathbf{h}_K] \in \mathbb{C}^{M\times K}$ the uplink channel from *Bobs* to *Alice*, $\beta_{h,k}$ and $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ describe the large- and small-scale fading factors related to the k-th *Bob*, respectively. In addition, elements in $\mathbf{h}_k$ are independent and identically distributed (*i.i.d.*) complex Gaussian random variables with zero mean and unit variance, i.e., $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_M)$. Moreover, it is worth noting that the downlink channel can be represented as $\mathbf{H}^T$ due to the channel reciprocity since the time-division duplex is assumed in our study.

In general, *Alice* requires the CSI to *Bobs* to carry out multiuser detection in the uplink and precoding in the downlink. Toward this end, *Bobs* send orthogonal pilot sequences to *Alice* for the purpose of channel estimation, which is referred to as the training phase. However, this unintentionally provides opportunity for malicious nodes to attack legitimate communications. For example, an eavesdropper *Eve* may send identical pilots as *Bobs* to *Alice* if these pilot sequences are publicly known. For simplicity, we assume the number of antennas at *Eve* is K, and our analysis can be easily extended when it is not equal to K. Besides, the k-th antenna sends pilots identical to the k-th *Bob*[1]. The eavesdropping link between *Eve* and *Alice* is denoted by $\mathbf{G} = [\sqrt{\beta_g}\mathbf{g}_1,\dots,\sqrt{\beta_g}\mathbf{g}_K] \in \mathbb{C}^{M\times K}$, with its (m,k)-th element being $g_{m,k} \sim \mathcal{CN}(0,\beta_g)$. Without loss of generality, different channel vectors are supposed to be mutually independent.

B. Impact of the Pilot Spoofing Attack

Because of the limited number of orthogonal sequences and periodic transmission of pilots, *Eve* can learn various pilot assignments to various *Bobs* as well as their transmission time slot [25]. When *Eve* launches a PSA in the training phase, the $M\times\tau$ received signal by *Alice* is

$$\mathbf{Y} = \sqrt{p_B}\mathbf{H}\mathbf{X} + \sqrt{p_E}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X} + \mathbf{N} \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{K\times\tau}$ denotes pilot sequences and τ is the pilot length, with $K \le \tau$; p_B and p_E are the transmit power of *Bobs* and *Eve*, respectively; $\boldsymbol{\alpha} = \text{diag}\,(\alpha_1,\dots,\alpha_K)$ is a diagonal matrix where α_k indicates the power allocation factor of the k-th eavesdropping antenna with $\sum_{k=1}^K \alpha_k = 1$; and $\mathbf{N} \in \mathbb{C}^{M\times\tau}$ denotes the noise matrix with its elements obeying $\mathcal{CN}(0,\beta_n)$. Since $\mathbf{X}\mathbf{X}^H = \tau\mathbf{I}_K$, the channel estimate of $\mathbf{H}$ using the least-square (LS) method is

$$\hat{\mathbf{H}} = \frac{\mathbf{Y}\mathbf{X}^H}{\tau\sqrt{p_B}} = \mathbf{H} + \sqrt{\frac{p_E}{p_B}}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}} + \mathbf{E} \tag{2}$$

where $\mathbf{E} = \frac{\mathbf{N}\mathbf{X}^H}{\tau\sqrt{p_B}}$ is independent of $\mathbf{H}$ and $\mathbf{G}$, with its entries obeying $\mathcal{CN}(0,\beta_n/(p_B\tau))$. Eq. (2) indicates that the channel estimate is contaminated by the eavesdropping link $\mathbf{G}$.

[1] A single-antenna *Eve* is able to attack all users simultaneously by sending a linear combination of different pilot sequences [6]. However, the eavesdropping information, which is a combination of data streams from all users, cannot be further separated. While in our study, a multi-antenna *Eve* could wiretap K independent data streams. Thus from the viewpoint of eavesdropping, a multi-antenna *Eve* is more preferable.

Afterwards, *Alice* broadcasts data streams to *Bobs* by using maximum-ratio transmission precoding, of which the precoding matrix $\mathbf{Q} \in \mathbb{C}^{M \times K}$ is

$$\mathbf{Q} = \frac{\hat{\mathbf{H}}^*}{\sqrt{\mathrm{Tr}(\hat{\mathbf{H}}^T \hat{\mathbf{H}}^*)}} \tag{3}$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace. As a result, part of the beamforming vectors will point to *Eve* since channel estimate is a linear combination of $\mathbf{H}$ and $\mathbf{G}$. Denoting by $\mathbf{d}_{Bobs}$ and $\mathbf{d}_{Eve}$ the received signals by *Bobs* and *Eve*, respectively

$$\begin{aligned} \mathbf{d}_{Bobs} &= \sqrt{p_A}\mathbf{H}^T\mathbf{Q}\mathbf{s} + \mathbf{v}_{Bobs} \\ \mathbf{d}_{Eve} &= \sqrt{p_A}\mathbf{G}^T\mathbf{Q}\mathbf{s} + \mathbf{v}_{Eve} \end{aligned} \tag{4}$$

where p_A denotes the transmit power of *Alice* and $\mathbf{s} \in \mathbb{C}^{K \times 1}$ indicates the information-bearing signal with $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_K$; $\mathbf{v}_{Bobs}$ and $\mathbf{v}_{Eve}$ are noise vectors at *Bobs* and *Eve*, respectively, with $\mathbf{v}_{Bobs} \sim \mathcal{CN}(\mathbf{0}, \beta_b \mathbf{I}_K)$ and $\mathbf{v}_{Eve} \sim \mathcal{CN}(\mathbf{0}, \beta_e \mathbf{I}_K)$. By plugging the precoding matrix $\mathbf{Q}$ into (4) and assuming M is sufficiently large, effective channels behave (nearly) deterministic according to the central limit theorem (CLT). Hence, the sum rate of legitimate users C_{Bobs} and the eavesdropping rate C_{Eve} approximate to

$$\begin{aligned} C_{Bobs} &\approx \sum_{k=1}^{K} \log_2\left(1 + \frac{p_A p_B M \beta_{h,k}^2/\beta_b}{p_B \beta_{h,k} + \alpha_k p_E \beta_g + \beta_n/\tau}\right) \\ C_{Eve} &\approx \sum_{k=1}^{K} \log_2\left(1 + \frac{p_A \alpha_k p_E M \beta_g^2/\beta_e}{p_B \beta_{h,k} + \alpha_k p_E \beta_g + \beta_n/\tau}\right). \end{aligned} \tag{5}$$

Accordingly, the secrecy rate is given by

$$C_{Sec} \approx \left[\sum_{k=1}^{K} \log\left(\frac{p_B \beta_{h,k}^2}{\alpha_k p_E \beta_g^2}\right)\right]^+ \tag{6}$$

where $[x]^+ = \max\{0, x\}$.

Clearly, there exists information leakage to *Eve* due to the PSA, even when the number of antennas at *Alice* is sufficiently large. In addition to reducing the achievable rate of legitimate communications, *Eve* can wiretap messages intended for *Bobs*. If *Eve* enhances its transmit power p_E, the secrecy rate will be tremendously degraded. In the extreme scenario, the secrecy rate reduces to zero if $\alpha_k p_E \beta_g^2 > p_B \beta_{h,k}^2$. Since the estimated channel is contaminated by the PSA, C_{Bobs} and C_{Eve} grow with M at the same speed. Thus deploying more antennas at *Alice* won't help alleviate this leakage problem.

III. Manipulation on Pilot Sequences

In order to detect the PSA, users need to modify their pilot sequences. Specifically, $\mathbf{X}$ is partitioned into $\mathbf{X}_A \in \mathbb{C}^{K \times a\tau}$ and $\mathbf{X}_B \in \mathbb{C}^{K \times b\tau}$, where $a\tau$ and $b\tau$ are integers with $a + b = 1$. We assume $\mathbf{X}_A\mathbf{X}_A^H = a\tau\mathbf{I}_K$ and $\mathbf{X}_B\mathbf{X}_B^H = b\tau\mathbf{I}_K$, which can be achieved by choosing $\mathbf{X}$ as a Hadamard matrix[2]. Then $\mathbf{X}_B$ is multiplied with $\mathbf{W} = \mathrm{diag}\left(w_1, \ldots, w_K\right) \in \mathbb{C}^{K \times K}$, in which w_k is a random number in the interval of $(0, 1)$. As a result, the modified pilot sequences are denoted by

$$\tilde{\mathbf{X}} = \boldsymbol{\Xi}\hat{\mathbf{X}} \tag{7}$$

where $\hat{\mathbf{X}} = [\mathbf{X}_A, \mathbf{W}\mathbf{X}_B]$, and $\boldsymbol{\Xi} \in \mathbb{R}^{K \times K}$ is a diagonal matrix with its k-th diagonal element being $\xi_k = \sqrt{\frac{1}{a + w_k^2 b}}$. Note that the use of $\boldsymbol{\Xi}$ is for the purpose of power normalization, i.e. $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H = \tau\mathbf{I}_K$. Accordingly, the pilots observed by *Alice* is

$$\begin{aligned} \mathcal{H}_0 :&\quad \mathbf{Y} = \sqrt{p_B}\mathbf{H}\tilde{\mathbf{X}} + \mathbf{N} \\ \mathcal{H}_1 :&\quad \mathbf{Y} = \sqrt{p_B}\mathbf{H}\tilde{\mathbf{X}} + \sqrt{p_E}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X} + \mathbf{N} \end{aligned} \tag{8}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ are hypotheses of absence and presence of *Eve*[3]. Moreover, the received signals corresponding to $\mathbf{X}_A$ and $\mathbf{X}_B$ are denoted by $\mathbf{Y}_A \in \mathbb{C}^{M \times a\tau}$ and $\mathbf{Y}_B \in \mathbb{C}^{M \times b\tau}$, respectively, with $\mathbf{Y} = [\mathbf{Y}_A \; \mathbf{Y}_B]$.

Note that the manipulation mentioned above is only applied to pilot sequences but not to information-bearing symbols. In the interest of fairness, *Alice* is unaware of a (or b) and $\mathbf{W}$. Thus, before estimating the CSI, *Alice* needs to estimate these parameters.

A. Estimation of a (or b)

As the first step, it is necessary to differentiate $\mathbf{Y}_A$ and $\mathbf{Y}_B$. Therefore, *Alice* computes the normalized covariance matrix $\mathbf{Y}^H\mathbf{Y}/M$, and the vector of its diagonal elements is $\mathbf{t} = [t_{A,1}..., t_{A,a\tau}, t_{B,a\tau+1}, ..., t_{B,\tau}]$[4]. Since M is a large number, each element in $\mathbf{t}$ approximates to a Gaussian variable according to the CLT. Hence, the expectation of $t_{A,i}$ is

$$\begin{aligned} \mathcal{H}_0 : \mathbb{E}\left\{t_{A,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left|\xi_k\right|^2 + \beta_n \\ \mathcal{H}_1 : \mathbb{E}\left\{t_{A,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left|\xi_k\right|^2 + p_E\beta_g + \beta_n \end{aligned} \tag{9}$$

the detailed derivations can be found in Appendix A. On the other hand, if $a\tau + 1 \leq i \leq \tau$, the expectation changes to

$$\begin{aligned} \mathcal{H}_0 : \mathbb{E}\left\{t_{B,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left|w_k\xi_k\right|^2 + \beta_n \\ \mathcal{H}_1 : \mathbb{E}\left\{t_{B,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left|w_k\xi_k\right|^2 + p_E\beta_g + \beta_n. \end{aligned} \tag{10}$$

By subtracting (10) from (9), we have

$$\begin{aligned} \mathcal{H}_0 : \mathbb{E}\left\{t_{A,i}\right\} - \mathbb{E}\left\{t_{B,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left(1 - w_k^2\right) \xi_k^2 \\ \mathcal{H}_1 : \mathbb{E}\left\{t_{A,i}\right\} - \mathbb{E}\left\{t_{B,i}\right\} &= p_B \sum_{k=1}^{K} \beta_{h,k} \left(1 - w_k^2\right) \xi_k^2. \end{aligned} \tag{11}$$

Note that $\mathbb{E}\{t_{A,i}\} - \mathbb{E}\{t_{B,i}\}$ is identical under $\mathcal{H}_0$ and $\mathcal{H}_1$, which means there is one and only one jump in $\mathbb{E}\{\mathbf{t}\}$. With out

[2] Note that this assumption holds only if $K \leq a\tau$ and $K \leq b\tau$, which means $2K \leq \tau$. Therefore, hereafter, we assume $\tau \geq 2K$. Theoretically, to guarantee the mutual orthogonality among pilot sequences of K users, the minimum pilot length is $\tau = K$. However, pilots are reused among different cells in cellular networks, which increases the minimum pilot length to the product of K and the pilot reuse factor [2]. Therefore, our assumption that $\tau \geq 2K$ is reasonable.

[3] Here, we assume *Eve* is unaware of the manipulation in (7). However, the scenario that *Eve* knows (7) will be discussed in Section IV.

[4] For ease of exposition, we use $t_{A,i}$ to indicate the first $a\tau$ elements in $\mathbf{t}$, while $t_{B,i}$ denotes the remaining ones.

loss of generality, let's consider the i-th element t_i. According to the definition, t_i is expanded as

$$t_i = \frac{\mathbf{y}_i^H \mathbf{y}_i}{M} = \sum_{m=1}^{M} \frac{y_{i,m}^* y_{i,m}}{M} \tag{12}$$

where $\mathbf{y}_i$ is the i-th row of $\mathbf{Y}$, and $y_{i,m}$ is the m-th entry of $\mathbf{y}_i$. Eq. (12) indicates t_i is the sample mean of $y_{i,m}^* y_{i,m}$, where the sample size is M. On the other hand, the statistical mean of t_i can be computed as

$$\mathbb{E}\{t_i\} = \sum_{m=1}^{M} \mathbb{E}\left\{\frac{y_{i,m}^* y_{i,m}}{M}\right\} = \mathbb{E}\{y_{i,m}^* y_{i,m}\}. \tag{13}$$

Note that (13) is built upon the fact all $y_{i,m}^* y_{i,m}$ are *i.i.d.* random variables. It is known that as the number of samples increases, the sample mean of a random variable will gradually converges to the statistical mean. Thus according to (12) and (13), t_i can be considered as a good approximation of $\mathbb{E}\{t_i\}$ if M is large enough, which is easily satisfied in the scenario of massive MIMO.

Given the above analysis, one can expect there is only one significant jump in the vector $\mathbf{t}$. In other words, the boundary between $\mathbf{Y}_A$ and $\mathbf{Y}_B$ can be found by searching this jump, i.e., finding the maximum of the difference between neighboring components in $\mathbf{t}$. In the following analysis, we assume *Alice* can successfully differentiate $\mathbf{Y}_A$ and $\mathbf{Y}_B$, i.e.,

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{Y}_A = \sqrt{p_B}\mathbf{H}\mathbf{\Xi}\mathbf{X}_A + \mathbf{N}_A \\ \mathcal{H}_1 &: \mathbf{Y}_A = \sqrt{p_B}\mathbf{H}\mathbf{\Xi}\mathbf{X}_A + \sqrt{p_E}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X}_A + \mathbf{N}_A \end{aligned} \tag{14}$$

where $\mathbf{N}_A \in \mathbb{C}^{M \times a\tau}$ denotes the noise associated with the transmission of $\mathbf{X}_A$. Similarly, $\mathbf{Y}_B$ can be expressed as

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{Y}_B = \sqrt{p_B}\mathbf{H}\mathbf{\Xi}\mathbf{W}\mathbf{X}_B + \mathbf{N}_B \\ \mathcal{H}_1 &: \mathbf{Y}_B = \sqrt{p_B}\mathbf{H}\mathbf{\Xi}\mathbf{W}\mathbf{X}_B + \sqrt{p_E}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X}_B + \mathbf{N}_B \end{aligned} \tag{15}$$

where $\mathbf{N}_B \in \mathbb{C}^{M \times b\tau}$ is the noise similar to $\mathbf{N}_A$.

B. The Estimation of $\mathbf{W}$

In this part, we will show how to estimate $\mathbf{W}$ with $\mathbf{Y}_A$ and $\mathbf{Y}_B$. In the absence of *Eve*, by multiplying $\mathbf{Y}_A$ and $\mathbf{Y}_B$ with $\mathbf{X}_A^H$ and $\mathbf{X}_B^H$ respectively, one arrives at

$$\mathcal{H}_0 : \begin{aligned} \mathbf{Z}_A &= \frac{\mathbf{Y}_A\mathbf{X}_A^H}{a\tau\sqrt{p_B}} = \mathbf{H}\mathbf{\Xi} + \frac{\mathbf{N}_A\mathbf{X}_A^H}{a\tau\sqrt{p_B}} \\ \mathbf{Z}_B &= \frac{\mathbf{Y}_B\mathbf{X}_B^H}{b\tau\sqrt{p_B}} = \mathbf{H}\mathbf{\Xi}\mathbf{W} + \frac{\mathbf{N}_B\mathbf{X}_B^H}{b\tau\sqrt{p_B}} \end{aligned}. \tag{16}$$

In the presence of *Eve*, it can be obtained

$$\mathcal{H}_1 : \begin{aligned} \mathbf{Z}_A &= \frac{\mathbf{Y}_A\mathbf{X}_A^H}{a\tau\sqrt{p_B}} = \mathbf{H}\mathbf{\Xi} + \sqrt{\frac{p_E}{p_B}}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}} + \frac{\mathbf{N}_A\mathbf{X}_A^H}{a\tau\sqrt{p_B}} \\ \mathbf{Z}_B &= \frac{\mathbf{Y}_B\mathbf{X}_B^H}{b\tau\sqrt{p_B}} = \mathbf{H}\mathbf{\Xi}\mathbf{W} + \sqrt{\frac{p_E}{p_B}}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}} + \frac{\mathbf{N}_B\mathbf{X}_B^H}{b\tau\sqrt{p_B}} \end{aligned} \tag{17}$$

where $\mathbf{Z}_A, \mathbf{Z}_B \in \mathbb{C}^{M \times K}$. Through defining $\mathbf{\Phi} \triangleq \mathbf{Z}_A - \mathbf{Z}_B$, it follows from (16) and (17) that

$$\mathbf{\Phi} = \mathbf{H}\mathbf{\Xi}\left(\mathbf{I}_K - \mathbf{W}\right) + \frac{1}{\tau\sqrt{p_B}}\left(\frac{\mathbf{N}_A\mathbf{X}_A^H}{a} - \frac{\mathbf{N}_B\mathbf{X}_B^H}{b}\right). \tag{18}$$

Note that $\mathbf{\Phi}$ is identical under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ because $\sqrt{p_E/p_B}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}$ is canceled out by subtraction. As a result, the presence of *Eve* has no impact on estimating $\mathbf{W}$.

To estimate $\mathbf{W}$, we first compute $\mathbf{\Theta} = \mathbf{\Phi}^H\mathbf{\Phi}$, i.e.,

$$\begin{aligned} \mathbf{\Theta} = & \left(\mathbf{I}_K - \mathbf{W}\right)\mathbf{\Xi}\mathbf{H}^H\mathbf{H}\mathbf{\Xi}\left(\mathbf{I}_K - \mathbf{W}\right) \\ & + \frac{1}{\tau^2 p_B}\left(\frac{\mathbf{X}_A\mathbf{N}_A^H\mathbf{N}_A\mathbf{X}_A^H}{a^2} + \frac{\mathbf{X}_B\mathbf{N}_B^H\mathbf{N}_B\mathbf{X}_B^H}{b^2}\right) \\ & - \frac{1}{\tau^2 p_B}\left(\frac{\mathbf{X}_A\mathbf{N}_A^H\mathbf{N}_B\mathbf{X}_B^H}{ab} + \frac{\mathbf{X}_B\mathbf{N}_B^H\mathbf{N}_A\mathbf{X}_A^H}{ab}\right) \\ & + \left(\mathbf{I}_K - \mathbf{W}\right)\mathbf{\Xi}\mathbf{H}^H\frac{1}{\tau\sqrt{p_B}}\left(\frac{\mathbf{N}_A\mathbf{X}_A^H}{a} - \frac{\mathbf{N}_B\mathbf{X}_B^H}{b}\right) \\ & + \frac{1}{\tau\sqrt{p_B}}\left(\frac{\mathbf{X}_A\mathbf{N}_A^H}{a} - \frac{\mathbf{X}_B\mathbf{N}_B^H}{b}\right)\mathbf{H}\mathbf{\Xi}\left(\mathbf{I}_K - \mathbf{W}\right). \end{aligned} \tag{19}$$

According to the CLT, as M grows large, $\mathbf{\Theta}$ approximates to a $K \times K$ diagonal matrix whose k-th diagonal element is given by

$$\theta_k \approx (1 - w_k)^2 \xi_k^2 M\beta_{h,k} + \left(\frac{1}{a} + \frac{1}{b}\right)\frac{M\beta_n}{\tau p_B}. \tag{20}$$

As a consequence, w_k can be estimated as

$$\tilde{w}_k = \frac{2M\beta_{h,k} + \sqrt{4M\beta_{h,k}\left(a+b\right)\left(\theta_k - \tilde{n}\right) - 4ab(\theta_k - \tilde{n})^2}}{2M\beta_{h,k} + 2b\tilde{n} - 2b\theta_k} \tag{21}$$

where

$$\tilde{n} = \left(\frac{1}{a} + \frac{1}{b}\right)\frac{M\beta_n}{\tau p_B}.$$

In this way, $\tilde{\mathbf{W}} = \mathrm{diag}(\tilde{w}_1, \ldots, \tilde{w}_K)$ is obtained. However, $\tilde{w}_k$ can never be identical to w_k because of the approximation in (20). To improve the estimation accuracy, we propose the following method. Although w_k can be randomly selected in the range of $(0, 1)$ in theory, w_k is restricted to a finite set $\mathcal{P}_{\mathbf{w}}$ in practice. Then after obtaining $\tilde{w}_k$, one can undertake a line search to find $\tilde{w}_k'$ that is closest to $\tilde{w}_k$, i.e.,

$$\tilde{w}_k' = \arg\min_{w_k \in \mathcal{P}_{\mathbf{w}}} \left|\tilde{w}_k - w_k\right|. \tag{22}$$

The estimation accuracy decreases with the size of $\mathcal{P}_{\mathbf{w}}$, but the randomness increases with it. Here, the randomness refers to the the possible values of w_k. The more random w_k is, the more difficult for *Eve* to know $\mathbf{W}$. Thus, the selection of the size of $\mathcal{P}_{\mathbf{w}}$ needs to balance between security level and estimation accuracy.

So far, we assume *Eve* does not know the pilot manipulation in (7), and thus it transmits the original $\mathbf{X}$. However, it is possible for *Eve* to follow what *Bobs* do. In the next, we will discuss the scenario where Eve knows how the pilot sequences are manipulated.

IV. Pilot Manipulation Known by Eve

The modified pilot sequences at *Eve* are supposed to be

$$\tilde{\mathbf{X}}_E = \mathbf{\Xi}_E\hat{\mathbf{X}}_E = [\mathbf{X}_P, \mathbf{W}_E\mathbf{X}_Q] \tag{23}$$

where $\mathbf{W}_E = \mathrm{diag}(w_{E,1}, \ldots, w_{E,K}) \in \mathbb{C}^{K \times K}$, $\mathbf{X}_P \in {}^{K \times p\tau}$ and $\mathbf{X}_Q \in \mathbb{C}^{K \times q\tau}$ are two parts of $\mathbf{X}$ with $p + q = 1$, and

$\mathbf{\Xi}_E \in \mathbb{C}^{K\times K}$ is a diagonal matrix with its k-th diagonal entry being $\xi_{E,k} = \sqrt{\frac{1}{p+w_{E,k}^2 q}}$. Depending on how much *a priori* information *Eve* knows, there exist several possibilities:

1) Case 1: *Eve* is ignorant of the manipulation, namely $p = 1$, $q = 0$ and $\tilde{\mathbf{X}}_E = \mathbf{X}$, and this is the case discussed before;
2) Case 2: *Eve* partitions pilots into two parts. Due to the lack of knowledge of a and b, it is assumed $p \neq a$ and $q \neq b$;
3) Case 3: *Eve* partitions pilots into two parts of lengths a and b. Due to the lack of knowledge of $\mathbf{W}$, it is assumed $\mathbf{W}_E \neq \mathbf{W}$;
4) Case 4: *Eve* partitions pilots into two parts of lengths a and b, and it knows $\mathbf{W}$ and then sets $\mathbf{W}_E = \mathbf{W}$. In this case, *Eve* is completely synchronous with *Bobs*.

From Case 1 to 4, *Eve* behaves more actively and also needs more *a priori* information. Interestingly, from the viewpoint of eavesdropping, it is not necessarily beneficial for *Eve* to be more active, as will be shown next.

A. Detection of Eve in Case 2

It has been shown that $\mathbb{E}\{\mathbf{t}\}$ experiences one and only one jump in Case 1. On the other hand, if *Eve* sends pilots which are partitioned into two parts with lengths $p < a$ and $q > b$, then $\mathbb{E}\{\mathbf{t}\}$ becomes[5]

$$\begin{cases} p_B \sum_{k=1}^{K} \beta_{h,k}|\xi_k|^2 + \alpha_k p_E \sum_{k=1}^{K} \beta_g |\xi_{E,k}|^2 + \beta_n, \\ \qquad\qquad\qquad 1 \le i \le p\tau \\ p_B \sum_{k=1}^{K} \beta_{h,k}|\xi_k|^2 + \alpha_k p_E \sum_{k=1}^{K} \beta_g |w_{E,k}\xi_{E,k}|^2 + \beta_n, \\ \qquad\qquad\qquad p\tau+1 \le i \le a\tau \\ p_B \sum_{k=1}^{K} \beta_{h,k}|w_k\xi_k|^2 + \alpha_k p_E \sum_{k=1}^{K} \beta_g |w_{E,k}\xi_{E,k}|^2 + \beta_n, \\ \qquad\qquad\qquad a\tau+1 \le i \le \tau. \end{cases} \tag{24}$$

For Case 3, $\mathbb{E}\{\mathbf{t}\}$ changes to

$$\begin{cases} p_B \sum_{k=1}^{K} \beta_{h,k}|\xi_k|^2 + \alpha_k p_E \sum_{k=1}^{K} \beta_g |\xi_{E,k}|^2 + \beta_n, \\ \qquad\qquad\qquad 1 \le i \le a\tau \\ p_B \sum_{k=1}^{K} \beta_{h,k}|w_k\xi_k|^2 + \alpha_k p_E \sum_{k=1}^{K} \beta_g |w_{E,k}\xi_{E,k}|^2 + \beta_n, \\ \qquad\qquad\qquad a\tau+1 \le i \le \tau. \end{cases} \tag{25}$$

Finally, in Case 4, $\mathbb{E}\{\mathbf{t}\}$ is derived as

$$\begin{cases} \sum_{k=1}^{K} \left(p_B\beta_{h,k} + \alpha_k p_E \beta_g\right)|\xi_k|^2 + \beta_n, \\ \qquad\qquad\qquad 1 \le i \le a\tau \\ \sum_{k=1}^{K} \left(p_B\beta_{h,k} + \alpha_k p_E \beta_g\right)|w_k\xi_k|^2 + \beta_n, \\ \qquad\qquad\qquad a\tau+1 \le i \le \tau. \end{cases} \tag{26}$$

The derivations of (24), (25) and (26) are straightforward and thus omitted. To verify our results, Fig. 1 draws one sample of $\mathbf{t}$ in the four cases. It has been shown in Section III-A that $\mathbf{t}$ matches well with $\mathbb{E}\{\mathbf{t}\}$ when M is sufficiently large. Thus the results would be similar if $\mathbb{E}\{\mathbf{t}\}$ is drawn. One can find that in Case 2, if *Eve* recklessly divides pilot sequences without knowing a and b, there will be two jumps in $\mathbf{t}$. While in the other cases, only one jump is observed. Therefore, *Alice* could readily detect the presence of *Eve* in Case 2 by inspecting the number of jumps.

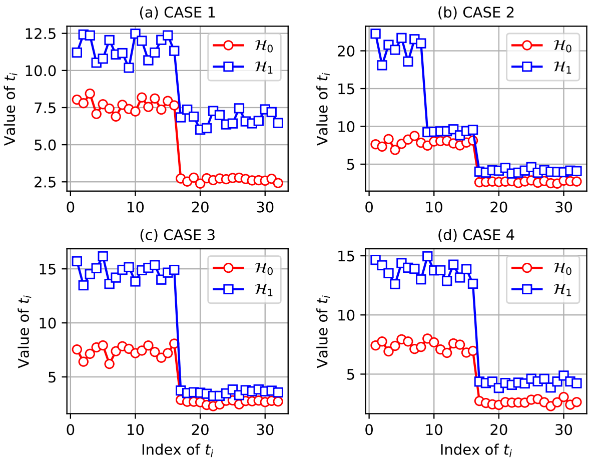

Fig. 1. Diagonal elements of the covariance matrix of the received signal at *Alice* in four cases.

B. Detection of Eve in Cases 3 and 4

Note that in (18), the eavesdropping link $\mathbf{G}$ is canceled out by computing $\mathbf{\Phi} = \mathbf{Z}_A - \mathbf{Z}_B$. Hence *Eve* would not affect the estimation of w_k in (21). However in Cases 3 and 4, the estimation of $\mathbf{W}$ will be impacted whether or not *Eve* knows $\mathbf{W}$. As a result, $\mathbf{\Phi}$ changes to

$$\mathbf{\Phi} = \mathbf{H\Xi}\left(\mathbf{I}_K - \mathbf{W}\right) + \frac{1}{\tau\sqrt{p_E}}\left(\frac{\mathbf{N}_A\mathbf{X}_A^H}{a} - \frac{\mathbf{N}_B\mathbf{X}_B^H}{b}\right) + \sqrt{\frac{p_E}{p_B}}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{\Xi}_E\left(\mathbf{I}_K - \mathbf{W}_E\right). \tag{27}$$

It is observed that the eavesdropping channel still exists, which leads to inaccuracy in estimating $\mathbf{W}$.

In Cases 3 and 4, *Eve* tries to be as synchronous with *Bobs* as possible, with the objective of interfering with legitimate communications more effectively and ensuring its own hidden ability simultaneously. However, we will show this could be counterproductive. Before that, it is worth noting that the first $a\tau$ elements in $\mathbf{t}$ follow a Gaussian distribution with $\mathbb{E}\{t_{A,i}\}$, while the last $b\tau$ elements follow another Gaussian distribution with mean $\mathbb{E}\{t_{B,i}\}$. Furthermore, the difference between the average of the first $a\tau$ elements in $\mathbf{t}$ and that of the last $b\tau$ elements is

$$u_{\text{jump}} = \frac{1}{a\tau}\sum_{i=1}^{a\tau} t_{A,i} - \frac{1}{b\tau}\sum_{i=a\tau+1}^{\tau} t_{B,i}. \tag{28}$$

Clearly, u_{jump} is a good approximation to $\mathbb{E}\{t_{A,i}\} - \mathbb{E}\{t_{B,i}\}$ if M is sufficiently large. On the other hand, $\mathbb{E}\{t_{A,i}\} - \mathbb{E}\{t_{B,i}\}$ is a function of w_k according to (11). Thus by substituting w_k

[5] It is worth noting that our analysis can be easily extended if two parts are of lengths $p > a$ and $q < b$.

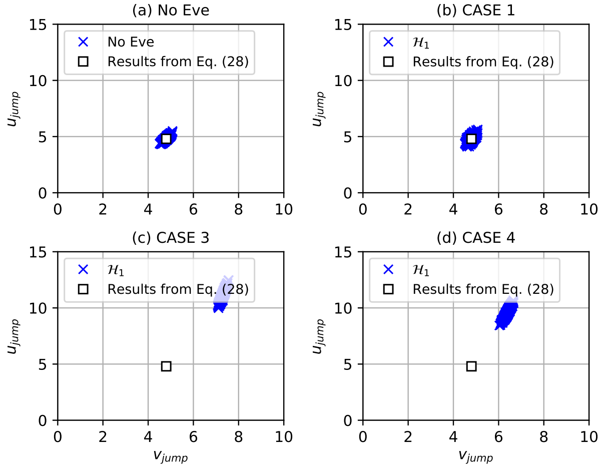

Fig. 2. u_{jump} versus v_{jump} in four cases.

with $\tilde{w}_k$, one can obtain another method to estimate $\mathbb{E}\{t_{A,i}\} - \mathbb{E}\{t_{B,i}\}$, which is termed v_{jump}, i.e.,

$$v_{\text{jump}} = p_B \sum_{k=1}^{K} \beta_{h,k} \frac{1-\tilde{w}_k^2}{a+\tilde{w}_k^2 b} \approx p_B \sum_{k=1}^{K} \beta_{h,k} \left(1-w_k^2\right) \xi_k^2. \tag{29}$$

According to (11), (25) and (26), one can compute u_{jump} as follows

$$u_{\text{jump}} \approx \begin{cases} p_B \sum\limits_{k=1}^{K} \beta_{h,k} \left(1-w_k^2\right)\xi_k^2 & \text{no } Eve \\ p_B \sum\limits_{k=1}^{K} \beta_{h,k} \left(1-w_k^2\right)\xi_k^2 & \text{Case 1} \\ p_E \sum\limits_{k=1}^{K} \beta_g \left(\left|\xi_{E,k}\right|^2 - \left|w_{E,k}\xi_{E,k}\right|^2\right) + & \\ \qquad p_B \sum\limits_{k=1}^{K} \beta_{h,k} \left(1-w_k^2\right)\xi_k^2 & \text{Case 3} \\ \sum\limits_{k=1}^{K} \left(p_B\beta_{h,k} + p_E\beta_g\right)\left(1-w_k^2\right)\xi_k^2 & \text{Case 4} \end{cases} \tag{30}$$

Eq. (30) clearly indicates if *Eve* is unaware of pilot modification, u_{jump} is approximately the same as v_{jump}. Meanwhile, it is also shown these two quantities are different in Cases 3 and 4. Fig. 2 displays u_{jump} versus v_{jump} in the four cases. Results in this figure validate the effectiveness of our analysis. Hence, *Eve* would expose itself in Case 3 or 4, since *Alice* can detect *Eve* via comparing u_{jump} to v_{jump}.

To conclude, a more active *Eve* requires more *a priori* information to launch attacks. However, the approach doesn't pay since its hidden ability will be severely harmed. Therefore, it is reasonable to assume that *Eve* makes no change on pilot sequences in the following sections.

V. Proposed PSA Detection Scheme

In this part, a new PSA detection algorithm is designed for a single user. After that, improving detection accuracy through joint detection among multiple users is investigated.

A. Design of the Decision Metric

Given the estimate of $\mathbf{W}$, one can design the decision metric of the proposed PSA detection. Toward this end, we first define the following

$$\boldsymbol{\Omega} \triangleq \left(\mathbf{Z}_B - \mathbf{Z}_A\tilde{\mathbf{W}}\right)\left(\mathbf{I}_K - \tilde{\mathbf{W}}\right)^{-1} \tag{31}$$

where $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are defined in (16) and (17). To facilitate the analysis, we assume the estimated $\mathbf{W}$ is accurate such that $\tilde{\mathbf{W}} = \mathbf{W}$[6], and thus we have

$$\begin{aligned} \mathcal{H}_0: \quad & \boldsymbol{\Omega} = \mathbf{V} \\ \mathcal{H}_1: \quad & \boldsymbol{\Omega} = \sqrt{\frac{p_E}{p_B}}\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}} + \mathbf{V} \end{aligned} \tag{32}$$

where

$$\mathbf{V} = \left(\frac{\mathbf{N}_B\mathbf{X}_B^H}{b\tau\sqrt{p_B}} - \frac{\mathbf{N}_A\mathbf{X}_A^H\mathbf{W}}{a\tau\sqrt{p_B}}\right)\left(\mathbf{I}_K - \mathbf{W}\right)^{-1}$$

and $\boldsymbol{\Omega} = [\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K] \in \mathbb{C}^{M\times K}$ with $\boldsymbol{\zeta}_k \in \mathbb{C}^{M\times 1}$ being its k-th column, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K] \in \mathbb{C}^{M\times K}$. Specifically, the k-th column of $\mathbf{V}$ follows

$$\mathbf{v}_k \sim \mathcal{CN}\left(\mathbf{0}, \frac{\lambda_k\beta_n}{p_B\tau}\mathbf{I}_M\right)$$

where $\lambda_k = \left(\frac{1}{b} + \frac{w_k^2}{a}\right) / \left(1-w_k\right)^2$. In most existing PSA detection methods, the decision metric under hypothesis $\mathcal{H}_0$ includes the legitimate channel, while it under $\mathcal{H}_1$ includes both the legitimate and eavesdropping channels. Interestingly, (32) indicates the legitimate link $\mathbf{H}$ is canceled out under both $\mathcal{H}_0$ and $\mathcal{H}_1$. Then from the detection point of view, the original problem of detecting *Eve* from the legitimate signal and noise translates to detecting *Eve* simply from noise. This could bring about remarkable performance improvements, as will be displayed later.

In practice, *Eve* can either attack all users or part of them. Hence, one needs to consider PSA detection for the single-user case. In our scheme, whether or not the k-th *Bob* is under attack is only determined by the k-th column of $\boldsymbol{\Omega}$, which is expressed by

$$\begin{aligned} \mathcal{H}_0: \quad & \boldsymbol{\zeta}_k = \mathbf{v}_k \\ \mathcal{H}_1: \quad & \boldsymbol{\zeta}_k = \sqrt{\frac{\alpha_k p_E}{p_B}}\mathbf{g}_k + \mathbf{v}_k \end{aligned} \tag{33}$$

where $\mathbf{g}_k$ indicates the k-th column of the eavesdropping link $\mathbf{G}$. In addition, the distribution of $\boldsymbol{\zeta}_k$ is derived as

$$\begin{aligned} \mathcal{H}_0: \quad & \boldsymbol{\zeta}_k \sim \mathcal{CN}\left(\mathbf{0}, \frac{\lambda_k\beta_n}{p_B\tau}\mathbf{I}_M\right) \\ \mathcal{H}_1: \quad & \boldsymbol{\zeta}_k \sim \mathcal{CN}\left(\mathbf{0}, \left(\frac{\alpha_k p_E}{p_B}\beta_g + \frac{\lambda_k\beta_n}{p_B\tau}\right)\mathbf{I}_M\right). \end{aligned} \tag{34}$$

By cascading the real and imaginary parts of $\boldsymbol{\zeta}_k$, one can obtain $\boldsymbol{\psi}_k = \left[\mathcal{R}(\boldsymbol{\zeta}_k^T), \mathcal{I}(\boldsymbol{\zeta}_k^T)\right]^T \in \mathbb{R}^{2M\times 1}$. Furthermore, the distribution of this new vector is

$$\begin{aligned} \mathcal{H}_0: \quad & \boldsymbol{\psi}_k \sim \mathcal{N}\left(\mathbf{0}, \sigma_0^2\mathbf{I}_{2M}\right) \\ \mathcal{H}_1: \quad & \boldsymbol{\psi}_k \sim \mathcal{N}\left(\mathbf{0}, \sigma_1^2\mathbf{I}_{2M}\right) \end{aligned} \tag{35}$$

[6] However, the estimate $\tilde{\mathbf{W}}$ is used in the simulation verification.

where

$$\sigma_0^2=\frac{\lambda_k\beta_n}{2p_B\tau},\quad \sigma_1^2=\frac{\alpha_k p_E}{2p_B}\beta_g+\frac{\lambda_k\beta_n}{2p_B\tau}.$$

According to the Neyman-Pearson theorem, the LRT principle is exploited to decide which hypothesis is true. Specifically, the likelihood-ratio test is given as follows

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\psi}_k\right)&=\frac{f\left(\boldsymbol{\psi}_k;\mathcal{H}_1\right)}{f\left(\boldsymbol{\psi}_k;\mathcal{H}_0\right)}\\&=\frac{\frac{1}{\left(2\pi\sigma_1^2\right)^M}\exp\left(-\frac{1}{2\sigma_1^2}\sum\limits_{i=1}^{2M}\psi_{k,i}^2\right)}{\frac{1}{\left(2\pi\sigma_0^2\right)^M}\exp\left(-\frac{1}{2\sigma_0^2}\sum\limits_{i=1}^{2M}\psi_{k,i}^2\right)}\\&=\left(\frac{\sigma_0^2}{\sigma_1^2}\right)^M\exp\left(\frac{\sigma_1^2-\sigma_0^2}{2\sigma_0^2\sigma_1^2}\sum_{i=1}^{2M}\psi_{k,i}^2\right)\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}}\eta_k\end{aligned}\tag{36}$$

where $\psi_{k,i}$ is the i-th component of $\boldsymbol{\psi}_k$ and η_k is the threshold for the k-th user. Applying the logarithmic operation to both sides of (36) yields

$$\ln\left(\mathcal{L}\left(\boldsymbol{\psi}_k\right)\right)=\ln\left(\left(\frac{\sigma_0^2}{\sigma_1^2}\right)^M\right)+\frac{\sigma_1^2-\sigma_0^2}{2\sigma_0^2\sigma_1^2}\sum_{i=1}^{2M}\psi_{k,i}^2\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}}\ln\eta_k.\tag{37}$$

With straightforward mathematical derivations, the final decision is obtained as

$$\phi_k=\sum_{i=1}^{2M}\psi_{k,i}^2\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}}\frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2-\sigma_0^2}\ln\left(\left(\frac{\sigma_1^2}{\sigma_0^2}\right)^M\eta_k\right)=\eta_k'.\tag{38}$$

Note that η_k' is positive since $\sigma_1^2>\sigma_0^2$. According to (38), the presence of *Eve* is declared if $\phi_k>\eta_k'$, and vise versa.

Next, we consider the complexity of the proposed scheme, which is attributed to two parts. The first part is the detection of boundary between $\mathbf{Y}_A$ and $\mathbf{Y}_B$ and estimation of $\mathbf{W}$, while the second part relates to the PSA detection. To be specific, the computational burden of these two parts are

$$\begin{aligned}&\texttt{Part\_1}: && \left(K+2\tau\right)\left(MC_{mul}+\left(M-1\right)C_{add}\right)+\tau C_{div}\\&\texttt{Part\_2}: && K\left(\left(4M-1\right)C_{add}+MC_{div}+3MC_{mul}\right)\end{aligned}\tag{39}$$

where C_{add}, C_{div} and C_{mul} denote the complex addition, division and multiplication, respectively. The proposed scheme incurs extra complexity due to the additional operations in part 1. However, from the detection perspective, it will be shown later that our scheme performs much better than traditional ones, for example [19].

B. Probabilities of False Alarm and Detection

Lemma 1: Suppose x_i are n i.i.d. samples drawn from a standard normal distribution, it then comes to $\sum_{i=1}^n x_i^2\sim\chi_n^2$.

Lemma 2: Suppose D is a χ^2 random variable with v degrees of freedom. If c is a positive constant, then it turns out that $cD\sim\Gamma(k=v/2,\theta=2c)$ is a Gamma random variable with shape parameter $v/2$ and rate parameter $2c$.

First, the decision metric can be converted to

$$\begin{aligned}\mathcal{H}_0:&\quad\phi_k=\sum_{i=1}^{2M}\psi_{k,i}^2=\sigma_0^2\sum_{i=1}^{2M}\left(\frac{\psi_{k,i}}{\sigma_0}\right)^2\\\mathcal{H}_1:&\quad\phi_k=\sum_{i=1}^{2M}\psi_{k,i}^2=\sigma_1^2\sum_{i=1}^{2M}\left(\frac{\psi_{k,i}}{\sigma_1}\right)^2.\end{aligned}\tag{40}$$

Since components in $\boldsymbol{\psi}_k$ are *i.i.d.* Gaussian random variables, it follows from Lemma 1 that

$$\begin{aligned}\mathcal{H}_0:&\quad\sum_{i=1}^{2M}\left(\frac{\psi_{k,i}}{\sigma_0}\right)^2\sim\chi_{2M}^2\\\mathcal{H}_1:&\quad\sum_{i=1}^{2M}\left(\frac{\psi_{k,i}}{\sigma_1}\right)^2\sim\chi_{2M}^2.\end{aligned}\tag{41}$$

With Lemma 2, (40) and (41), the distribution of ϕ_k can be shown as

$$\begin{aligned}\mathcal{H}_0:&\quad\phi_k\sim\Gamma\left(M,\frac{\lambda_k\beta_n}{p_B\tau}\right)\\\mathcal{H}_1:&\quad\phi_k\sim\Gamma\left(M,\frac{\alpha_k p_E}{p_B}\beta_g+\frac{\lambda_k\beta_n}{p_B\tau}\right).\end{aligned}\tag{42}$$

As can be seen from (42), ϕ_k is a Gamma random variable under both hypotheses, with the same shape parameters but different scale parameters. The probability distribution function of ϕ_k is

$$\begin{aligned}f\left(\phi_k;\mathcal{H}_0\right)&=\frac{\exp\left(-\frac{\phi_k p_B\tau}{\lambda_k\beta_n}\right)}{\Gamma\left(M\right)\left(\frac{\lambda_k\beta_n}{p_B\tau}\right)^M}\phi_k^{M-1}\\f\left(\phi_k;\mathcal{H}_1\right)&=\frac{\exp\left(-\frac{\phi_k}{\frac{\alpha_k p_E}{p_B}\beta_g+\frac{\lambda_k\beta_n}{p_B\tau}}\right)}{\Gamma\left(M\right)\left(\frac{\alpha_k p_E}{p_B}\beta_g+\frac{\lambda_k\beta_n}{p_B\tau}\right)^M}\phi_k^{M-1}\end{aligned}\tag{43}$$

where $\Gamma(\cdot)$ is the gamma function.

In detection theory, false alarm denotes the event that the presence of *Eve* is falsely declared when it is actually absent. In our scheme, the probability of false alarm for the k-th user is

$$\begin{aligned}P_{fa}^k&=1-\Pr\left\{\phi_k<\eta_k';\mathcal{H}_0\right\}\\&=1-\int_0^{\eta_k'}f\left(\phi_k;\mathcal{H}_0\right)\mathrm{d}\,\phi_k\\&=1-\frac{1}{\Gamma\left(M\right)}\gamma\left(M,\frac{\eta_k' p_B\tau}{\lambda_k\beta_n}\right)\end{aligned}\tag{44}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function. Given a predefined value of P_{fa}^k, threshold η_k' is computed according to (44), and then one can make the final decision through comparing the decision metric with η_k'. Moreover, the probability of detection for the k-th user is

$$\begin{aligned}P_d^k&=1-\Pr\left\{\phi_k<\eta_k';\mathcal{H}_1\right\}\\&=1-\int_0^{\eta_k'}f\left(\phi_k;\mathcal{H}_1\right)\mathrm{d}\,\phi_k\\&=1-\frac{1}{\Gamma\left(M\right)}\gamma\left(M,\frac{\eta_k'}{\frac{\alpha_k p_E}{p_B}\beta_g+\frac{\lambda_k\beta_n}{p_B\tau}}\right).\end{aligned}\tag{45}$$

In practice, it is difficult to obtain closed-form expressions of η_k' and P_d^k, and thus numerical methods can be resorted to using commercial software packages such as MATLAB.

C. Impact of Key Parameters on the Probability of Detection

Lemma 3: Suppose that G is a Gamma random variable with shape and scale parameters k and θ, then its mean and variance are

$$\begin{aligned}\mathbb{E}\left\{G\right\} &= k\theta \\ \mathrm{var}\left\{G\right\} &= k\theta^2.\end{aligned} \tag{46}$$

According to Lemma 3, the mean and variance of ϕ_k under hypothesis $\mathcal{H}_0$ are

$$\begin{aligned}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\} &= \frac{M\lambda_k\beta_n}{p_B\tau} \\ \mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\} &= M\left(\frac{\lambda_k\beta_n}{p_B\tau}\right)^2.\end{aligned} \tag{47}$$

Similarly, those results under hypothesis $\mathcal{H}_1$ are

$$\begin{aligned}\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\} &= \frac{\alpha_k p_E}{p_B}M\beta_g + \frac{\lambda_k\beta_n}{p_B\tau} \\ \mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\} &= M\left(\frac{\alpha_k p_E}{p_B}\beta_g + \frac{\lambda_k\beta_n}{p_B\tau}\right)^2.\end{aligned} \tag{48}$$

Besides, ϕ_k approximates to a Gaussian variable if M is large enough. Hence, we have $\phi_k \sim \mathcal{N}\left(\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}, \mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}\right)$ under $\mathcal{H}_0$, and $\phi_k \sim \mathcal{N}\left(\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}, \mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}\right)$ under $\mathcal{H}_1$.

Afterwards, P_{fa}^k approximates to

$$P_{fa}^k \approx 1 - \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\eta_k' - \mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}}{\sqrt{2\,\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}}\right)\right). \tag{49}$$

Then the threshold is recomputed as

$$\eta_k' \approx \mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\} + \sqrt{2\,\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}\mathrm{erf}^{-1}\left(1 - 2P_{fa}^k\right) \tag{50}$$

and the probability of detection is given by

$$P_d^k \approx 1 - \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\eta_k' - \mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}}{\sqrt{2\,\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}}}\right)\right). \tag{51}$$

Through substituting $\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}$ and $\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}$ into (51), the probability of detection is rewritten as

$$P_d^k = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\left(\mathrm{erf}^{-1}\left(1 - 2P_{fa}^k\right) + \sqrt{\frac{M}{2}}\right)\frac{\sigma_0^2}{\sigma_1^2} - \sqrt{\frac{M}{2}}\right). \tag{52}$$

Eq. (52) suggests that *Eve* is more likely to be detected if it raises its transmit power since σ_1^2 grows with P_E. On the other hand, the eavesdropping rate increases with a larger P_E according to (5). Therefore, *Eve* needs to tunes its transmit power to balance between eavesdropping and hidden abilities. Moreover, as $\mathrm{erf}(\cdot)$ is a monotonically increasing function, increasing M or pilot length τ is effective in improving the detection accuracy.

Meanwhile, w_k also influences the detection performance. To show its impact, P_d^k in (51) is first rewritten as

$$P_d^k = 1 - \frac{1}{2}\left(1 + \mathrm{erf}\left(g\left(x\right)\right)\right) \tag{53}$$

where

$$\begin{aligned}g\left(x\right) &= \frac{c_1x - c_2}{x + c_3}, \quad c_1 = \mathrm{erf}^{-1}\left(1 - 2P_{fa}^k\right) \\ c_2 &= \sqrt{\frac{M}{2}}\frac{\alpha_k p_E}{p_B}\beta_g, \quad c_3 = \frac{\alpha_k p_E}{p_B}\beta_g, \quad x = \frac{\lambda_k\beta_n}{p_B\tau}.\end{aligned} \tag{54}$$

The first-order derivative of $g(x)$ with respect to x is

$$g'\left(x\right) = \frac{c_1c_3 + c_2}{\left(x + c_3\right)^2}. \tag{55}$$

Since $g'(x)$ are positive, $g(x)$ is a monotonically increasing function of x. Besides, it is easy to show that x grows with w_k when $0 < w_k < 1$. As a result, $g(x)$ becomes smaller when w_k decreases, which in turn gives rise to a higher probability of detection. However, this does not mean it is always beneficial to reduce w_k, because the estimation accuracy of w_k in (21) cannot be guaranteed if w_k is too small.

D. Improving Performance Through Joint Detection

Our detection is built upon the observations in (33), where legitimate link $\mathbf{h}_k$ is absent. Note that this feature differentiates the proposed scheme from existing ones. For example, in [19], the observation is based on the estimated channel results, i.e.,

$$\begin{aligned}\mathcal{H}_0: &\quad \tilde{\boldsymbol{\zeta}}_k = \mathbf{h}_k + \mathbf{v}_k \\ \mathcal{H}_1: &\quad \tilde{\boldsymbol{\zeta}}_k = \mathbf{h}_k + \sqrt{\frac{\alpha_k p_E}{p_B}}\mathbf{g}_k + \mathbf{v}_k.\end{aligned} \tag{56}$$

Given (56), one can still follow (34)-(38) to design an appropriate decision metric. However, due to the presence of $\mathbf{h}_k$, the variance of the new decision metric becomes larger under both $\mathcal{H}_0$ and $\mathcal{H}_1$. Specifically, this means both σ_0^2 and σ_1^2 increase by the same amount. Then according to (52), one can derive the probability of detection will reduce as a result. Therefore, the proposed detection scheme outperforms existing ones, in which the legitimate link is part of the decision metric. Moreover, to further enhance accuracy, joint detection among multiple users is a strategy worth pursuing.

It is known that if *Eve* raises its jamming power towards the k-th *Bob*, the eavesdropping rate grows. However, this could make *Eve* more likely to be detected. Thus, a better strategy for *Eve* is to attack more users simultaneously, while reducing the jamming power towards any single user. In this way, although the eavesdropping rate related to a single user drops, the overall eavesdropping rate is guaranteed. More importantly, the hidden ability of *Eve* is enhanced by doing so. To tackle this eavesdropping strategy, we consider the joint detection by multiple users.

The decision metric of the proposed joint detection is $\tilde{\phi} = \sum_{k=1}^{K_1}\phi_k$, where K_1 is the number of users participating in joint detection. As before, we employ Gaussian approximation to analyze $\tilde{\phi}$. Since all ϕ_k are mutually independent, $\tilde{\phi}$ approximates to a Gaussian random variable, i.e.,

$$\begin{aligned}\mathcal{H}_0: &\quad \tilde{\phi} \sim \mathcal{N}\left(\sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}, \sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}\right) \\ \mathcal{H}_1: &\quad \tilde{\phi} \sim \mathcal{N}\left(\sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}, \sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}\right)\end{aligned} \tag{57}$$

where

$$\begin{aligned}\sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}&=\sum_{k=1}^{K_1}\frac{M\lambda_k\beta_n}{p_B\tau}\\ \sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}&=\sum_{k=1}^{K_1}M\left(\frac{\lambda_k\beta_n}{p_B\tau}\right)^2\\ \sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}&=\sum_{k=1}^{K_1}\left(\frac{\alpha_k p_E}{p_B}M\beta_g+\frac{M\lambda_k\beta_n}{p_B\tau}\right)\\ \sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}&=\sum_{k=1}^{K_1}M\left(\frac{\alpha_k p_E}{p_B}\beta_g+\frac{\lambda_k\beta_n}{p_B\tau}\right)^2.\end{aligned}$$

Note that in joint detection, $\mathcal{H}_0$ indicates no user is under attack, while $\mathcal{H}_1$ means all users are under attack.

It follows from (57) that the probability of false alarm for the joint detection is

$$\tilde{P}_{fa}=1-\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\tilde{\eta}-\sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}}{\sqrt{2\sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}}\right)\right). \quad (58)$$

Then the threshold is computed by

$$\tilde{\eta}=\sum_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}+\sqrt{2\sum_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}\mathrm{erf}^{-1}\left(1-2\tilde{P}_{fa}\right). \quad (59)$$

What's more, the probability of detection is given by

$$\tilde{P}_d=\frac{1}{2}-\frac{1}{2}\mathrm{erf}\left(\frac{\mathrm{erf}^{-1}\left(1-2\tilde{P}_{fa}\right)\sqrt{\sum\limits_{k=1}^{K_1}q_k^2}-\sqrt{\frac{M}{2}}\sum\limits_{k=1}^{K_1}r_k}{\sqrt{\sum\limits_{k=1}^{K_1}\left(r_k+q_k\right)^2}}\right) \quad (60)$$

where

$$q_k=\frac{\lambda_k\beta_n}{p_B\tau},\quad r_k=\frac{\alpha_k p_E}{p_B}\beta_g.$$

The probability of detection for a single user in (52) is a special case of (60) when $K_1=1$.

Finally, we discuss whether the joint detection helps improve performance. For comparative purposes, we convert the decision metric ϕ_k to a Gaussian variable of unit variance. As a consequence, the difference between expectations of decision metrics (normalized) under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ is

$$d_k=\left|\frac{\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}}{\sqrt{\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}}-\frac{\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}}{\sqrt{\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}}}\right|. \quad (61)$$

In the same way, after converting $\tilde{\phi}$ to another Gaussian variable with unit variance, this difference is given by

$$\tilde{d}=\left|\frac{\sum\limits_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}}{\sqrt{\sum\limits_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}}-\frac{\sum\limits_{k=1}^{K_1}\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}}{\sqrt{\sum\limits_{k=1}^{K_1}\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}}}\right|. \quad (62)$$

Since the normalized decision metric is of unit variance, joint detection outperforms single-user detection if $\tilde{d}>d_k$. For example, as it is difficult for *Eve* to know large-scale fading factors of *Bobs*, a straightforward way is equal power allocation among antennas, then (62) simplifies to

$$\begin{aligned}\tilde{d}&=\left|\frac{K_1\mathbb{E}\left\{\phi_k;\mathcal{H}_0\right\}}{\sqrt{K_1\mathrm{var}\left\{\phi_k;\mathcal{H}_0\right\}}}-\frac{K_1\mathbb{E}\left\{\phi_k;\mathcal{H}_1\right\}}{\sqrt{K_1\mathrm{var}\left\{\phi_k;\mathcal{H}_1\right\}}}\right|\\ &=\sqrt{K_1}d_k.\end{aligned} \quad (63)$$

Thus the detection accuracy is improved. On the other hand, *Eve* may allocate much more power to a certain antenna than the others. In the event of this, the performance of joint detection is not necessarily better than single-user detection. However, as shown before, this strategy is unwise for *Eve*.

VI. Proposed PSA-resistant Receiver

Pilot modification not only enhances the detection performance, but also helps design a PSA-resistant receiver. The root cause of the PSA is that channel estimates are corrupted by the eavesdropping link. Hence, instead of traditional channel estimation methods, we utilize $\mathbf{\Phi}$ to construct a new channel estimator. In (18), it has already been known that the eavesdropping link $\mathbf{G}$ is removed from $\mathbf{\Phi}$. Hence, the channel can be estimated as

$$\begin{aligned}\tilde{\mathbf{H}}&=\mathbf{\Phi}(\mathbf{I}_K-\mathbf{w})^{-1}\mathbf{\Xi}^{-1}\\ &=\mathbf{H}+\mathbf{E}_1\end{aligned} \quad (64)$$

where

$$\mathbf{E}_1=\frac{1}{\tau\sqrt{p_B}}\left(\frac{\mathbf{N}_A\mathbf{X}_A^H}{a}-\frac{\mathbf{N}_B\mathbf{X}_B^H}{b}\right)\left(\mathbf{I}_K-\mathbf{W}\right)^{-1}\mathbf{\Xi}^{-1}.$$

It is evident the result in (64) is free of the eavesdropping link. Hence, the downlink beamforming vector will not point to *Eve*, so that information leakage is avoided. Meanwhile, note that the new channel estimator slightly enhances the noise power, which means it cannot achieve the same secrecy rate as the PSA-free receiver.

For ease of comparison, the achievable rate of downlink transmission is considered. After precoding by the new channel estimator, the signal received by *Bobs* is

$$\tilde{\mathbf{d}}_{Bobs}=\sqrt{p_A}\mathbf{H}^T\tilde{\mathbf{Q}}\mathbf{s}+\mathbf{v}_{Bobs} \quad (65)$$

where

$$\tilde{\mathbf{Q}}=\frac{\tilde{\mathbf{H}}^*}{\sqrt{\mathrm{Tr}(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}^*)}}$$

is the precoding matrix that depends on $\tilde{\mathbf{H}}$. At the same time, the signal received by *Eve* is

$$\tilde{\mathbf{d}}_{Eve}=\sqrt{p_A}\mathbf{G}^T\tilde{\mathbf{W}}\mathbf{s}+\mathbf{v}_{Eve}. \quad (66)$$

Without loss of generality, a unit transmit power is assumed, i.e., $p_A=1$. Besides, we assume the noise generated at *Bobs* and *Eve* follows the same distribution of $\mathcal{CN}(\mathbf{0},\mathbf{I}_K)$. Thus given (64) and (65), if M is sufficiently large, the achievable sum rate of users is given approximately as

$$C'_{Bobs}\approx\sum_{k=1}^{K}\log_2\left(1+\frac{M\beta_{h,k}^2}{\beta_{h,k}+\left(\frac{1}{a}+\frac{1}{b}\right)\frac{a+w_k^2b}{\left(1-w_k\right)^2}\frac{\beta_n}{\tau p_B}}\right). \quad (67)$$

TABLE I
THE SIMULATION PARAMETERS

Number of antennas (M)	100
Transmit power of *Bobs* (P_B)	0 dB
$\beta_{h,k}$ and β_g	1
Pilot length (τ)	32
Noise variance (β_n)	0 dB
a	1/2
w_k	1/2

While according to (66), the eavesdropping rate approximates to

$$C'_{Eve} \approx \sum_{k=1}^{K} \log_2 (1+\beta_g). \tag{68}$$

It is important to note that C'_{Eve} is independent of M, which contrasts the result in (5). Consequently, the secrecy rate of the proposed PSA-resistant receiver is computed as

$$C'_{Sec} \approx \left[\sum_{k=1}^{K} \log_2 \left(\frac{M\beta_{h,k}^2}{(1+\beta_g)\left(\beta_{h,k}+\frac{\tilde{n}}{M}\cdot\frac{a+w_k^2 b}{(1-w_k)^2}\right)} \right) \right]^+ . \tag{69}$$

As can be inferred from (69), the secrecy rate could increase indefinitely with M. Our proposed PSA-resistant receiver is capable of improving the secrecy rate through increasing the number of antennas M or the length of pilot τ, which is not the case in (6). Therefore, the proposed receiver is robust against the PSA in massive MIMO systems.

VII. SIMULATION RESULTS

This section presents simulation results for the proposed detection scheme and PSA-resistant receiver. The considered model contains legitimate transmit-receive pairs (i.e., *Alice* and *Bobs*), and a malicious node *Eve*. The channels between the transmit-receive pairs are modeled by independent Rayleigh fading. The signal-to-noise ratio (SNR) used in all simulations is evaluated at *Bobs*. Furthermore, both analytical and numerical results are presented.

A. Single-user Detection

Fig. 3 displays the receiver operating characteristic (ROC) curves of the proposed scheme in the case of single-user detection. The simulation parameters are listed in Table I. As can be observed from the figure, the analytical and numerical results match each other well. In addition, when P_{fa} is fixed, P_d increases with *Eve*'s transmit power. For example, the probability of detection approaches 100% at $P_E = -5$ dB and $P_{fa} = 1\%$. Meanwhile, the eavesdropping rate grows with P_E, thus a trade-off exists.

Fig. 4 shows the comparison between a predefined P_{fa} and the simulated probability of false alarm. Since approximations are used in deriving our detection scheme, some extent of mismatch would exist between P_{fa} and $\Pr\{\phi_k > \eta'_k; \mathcal{H}_0\}$. A large mismatch often signifies that the simulation results are of low degree of confidence. Fortunate enough, the close match between P_{fa} and the simulated $\Pr\{\phi_k > \eta'_k; \mathcal{H}_0\}$ validates the proposed algorithm.

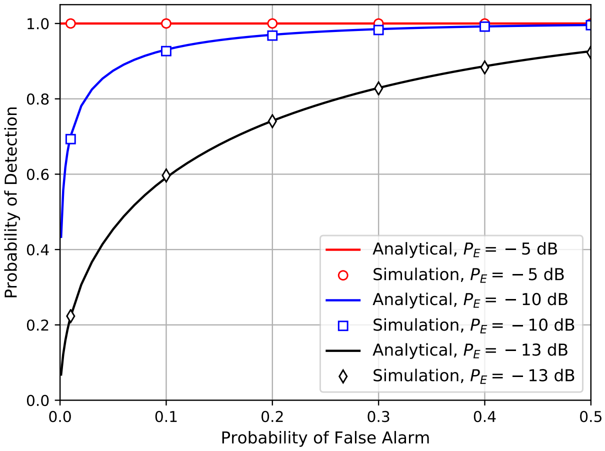

Fig. 3. ROC curves of the proposed single-user detection scheme, where P_E varies.

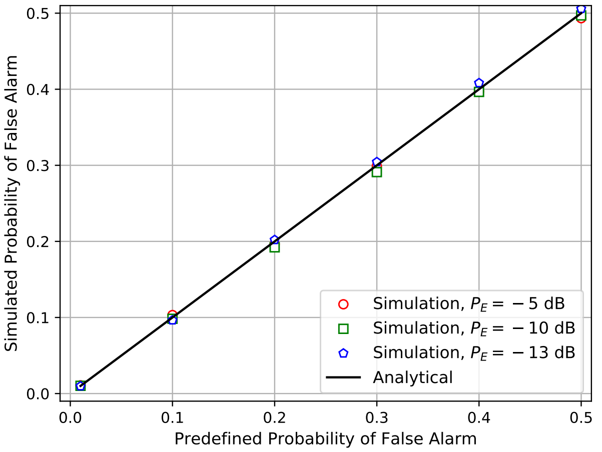

Fig. 4. Comparison between the simulated $\Pr\{\phi_k > \eta'_k; \mathcal{H}_0\}$ and predefined P_{fa}.

Fig. 5 depicts the relationship between P_d and the number of antennas M, where P_E varies and $P_{fa} = 1\%$. As aforementioned, the analytical results match quite well with their numerical counterparts. Besides, it is clear that deploying more antennas is an effective way to improve the detection accuracy. This observation verifies massive MIMO is capable of enhancing the physical layer security. For instance, when $P_{fa} = 1\%$ and $P_E = -8$ dB, P_d still approximately equals 96% when $M = 100$. Besides, note that a large P_E indicates *Eve* has a large transmit power, or *Eve* is physically close to *Alice*.

Fig. 6 draws ROC curves of the proposed detection scheme, where $P_{fa} = 1\%$, $P_E = -10$ dB and τ varies. It is well known that a large τ helps improve channel estimation. From the viewpoint of detection, this figure shows that increasing τ is beneficial to improving the probability of detection. In particular, P_d equals to 60%, 92% and 100% when $\tau = 16$, 32 and 64, respectively. Meanwhile, one needs to be aware of that a large pilot length may make legitimate communications

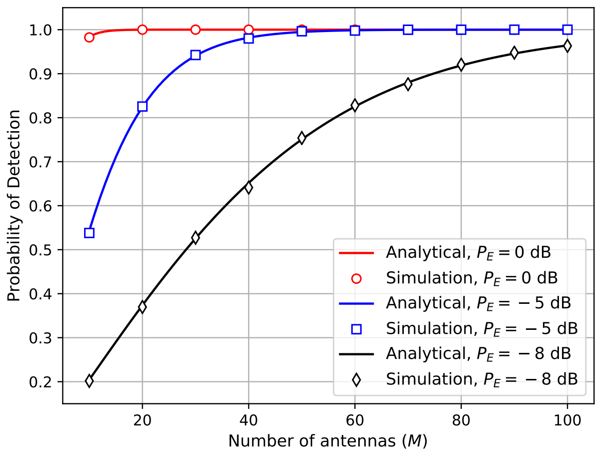

Fig. 5. Relationship between P_d and the number of antennas M at *Alice*, where P_E varies.

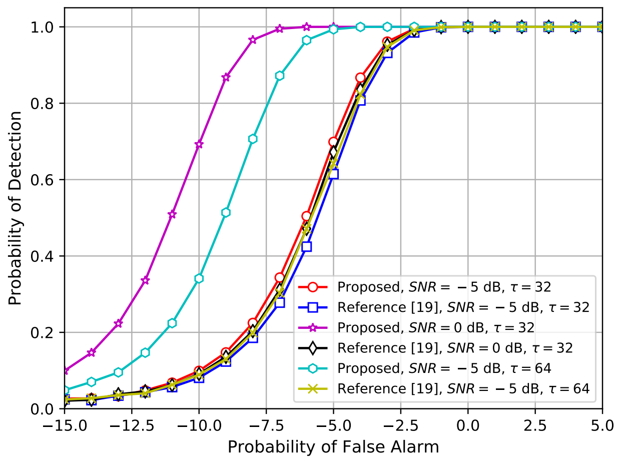

Fig. 7. Performance comparison between the proposed scheme and [19], where $P_{fa} = 1\%$, SNR and τ vary.

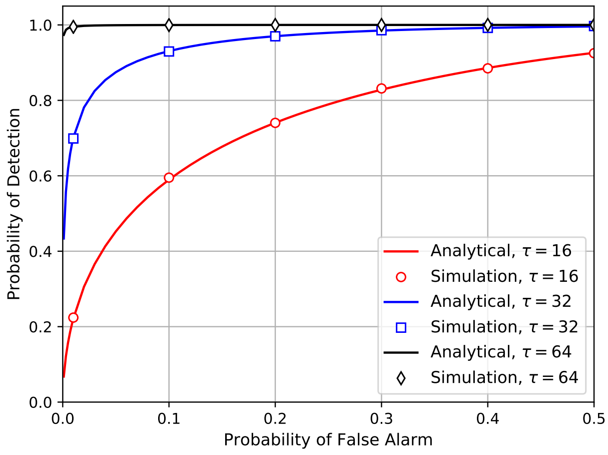

Fig. 6. ROC curves of the proposed single-user detection scheme, where τ varies.

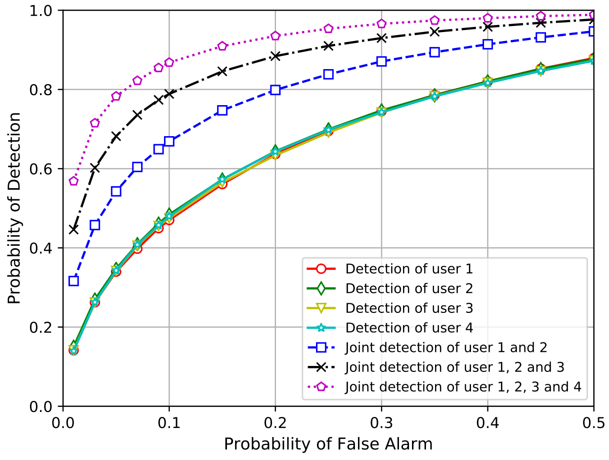

Fig. 8. ROC curves of multi-user joint detection and single-user detection, where $\boldsymbol{\alpha} = \text{diag}\{1/4, 1/4, 1/4, 1/4\}$.

more likely to be subject to eavesdropping.

Fig. 7 compares ROC curves of the algorithms proposed in this paper and [19], where $P_{fa} = 1\%$, and the SNR and τ vary. It is shown when SNR $= -5$ dB and $\tau = 32$, the proposed scheme is slightly better than [19]. As the SNR changes to 0 dB, the performance gap between these two widens. For example, the required P_E for the proposed method to meet 90% probability of detection or higher is 5 dB less than that required by [19]. Moreover, if τ increases to 64, the required jamming power for the proposed method to make P_d higher than 90% is 3 dB less than that required by [19].

B. Multi-user Joint Detection

Fig. 8 draws the ROC curves of joint detection and single-user detection, where $K = 4$, $\boldsymbol{\alpha} = \text{diag}\{1/4, 1/4, 1/4, 1/4\}$ and $\mathbf{W} = \text{diag}\{1/2, 1/2, 1/2, 1/2\}$. Other parameters can be referred to Table I. As can be seen from Fig. 8, single-user detection cannot achieve a decent performance as P_d is less than 50% when $P_{fa} = 10\%$. By contrast, P_d increases remarkably if more users participate in the joint detection. For example, P_d grows from less than 50% to almost 90% when all users are involved. Hence, the proposed joint multi-user detection is quite efficient in the case of equal jamming power allocation.

In addition, it is possible that *Eve* allocates more power to a certain antenna than the others. In this context, Fig. 9 depicts the ROC curves of joint detection with power allocation matrix $\boldsymbol{\alpha} = \text{diag}\{5/8, 1/8, 1/8, 1/8\}$. The other parameters can be referred to Fig. 8. Because the jamming power to user 1 is the largest, P_d under single-user detection by user 1 is the greatest. As for the joint multi-user detection, although the performance degrades compared with single-user detection by user 1, its detection probability is still much better than those of single-user detection by the other users. In conclusion, equal jamming power allocation to all antennas is a wise strategy for *Eve*, as this makes it less likely to be detected while maintaining a decent eavesdropping capability.

C. PSA-resistant Receiver

For the proposed PSA-resistant receiver, Fig. 10 depicts its sum secrecy rate of users, where the parameters can be referred to Fig. 8. The results obtained with conventional receivers are

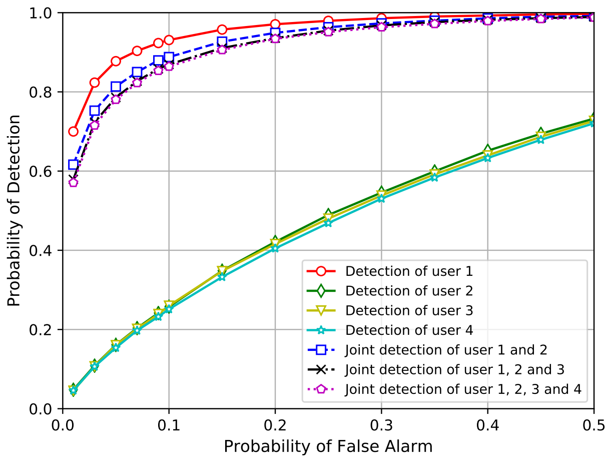

Fig. 9. ROC curves of multi-user joint detection and single-user detection, where $\boldsymbol{\alpha} = \text{diag}\{5/8, 1/8, 1/8, 1/8\}$.

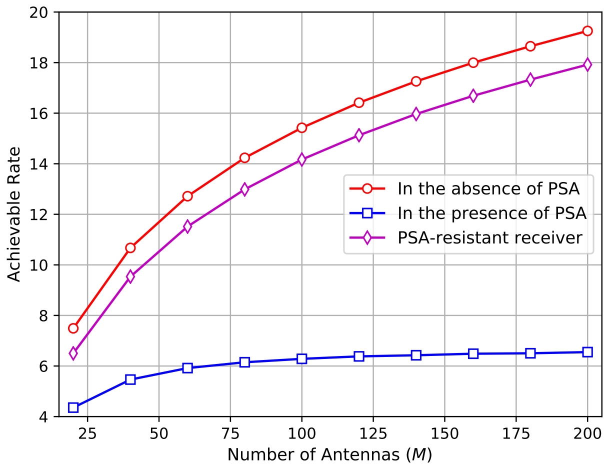

Fig. 10. Secrecy rate of the proposed PSA-resistant receiver.

included as benchmarks, where the two scenarios that legitimate communications are and are not subject to eavesdropping are considered. First of all, the secrecy rate of the conventional receiver cannot be improved by deploying more antennas as converges quickly as M grows. While for our proposed PSA-resistant receiver, the secrecy rate is remarkably enhanced and it grows with M. In addition, the performance gap between the proposed PSA-resistant receiver and eavesdropping-free receivers exists due to noise enhancement in our new channel estimator. Therefore, at the expense of a small rate loss, the proposed PSA-resistant receiver is able to successfully avoid information leakage.

VIII. Conclusion

In this paper, we investigated the PSA detection in massive MIMO systems using pilot manipulation. Our results showed that the best tactics for *Eve* is to ignore pilot manipulation, otherwise its hidden ability will be greatly harmed. According to the LRT principle, a decision metric was designed based on independent observations that do not include the legitimate link. This feature, which differentiates our scheme from existing ones, is the key to remarkably improve detection performance. Moreover, a joint multi-user detection was proved to be effective if *Eve* equally allocates its jamming power among the antennas. Finally, based on pilot manipulation, a new channel estimator was designed such that the estimated CSI is free of the effect of the eavesdropping channel. Numerical results were presented to demonstrate that the proposed detection scheme is effective. In addition, our PSA-resistant receiver is shown to be capable of enhancing security at the expense of a small rate loss attributed to noise enhancement.

Appendix A
Derivation of (9)

Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the normalized covariance matrix of the received signal at *Alice* is computed by

$$\begin{aligned}\mathcal{H}_0 : \mathbf{R} &= \frac{p_B}{M}\tilde{\mathbf{X}}^H\mathbf{H}^H\mathbf{H}\tilde{\mathbf{X}} + \frac{\mathbf{N}^H\mathbf{N}}{M} + \mathbf{S}_0 \\ \mathcal{H}_1 : \mathbf{R} &= \frac{p_B}{M}\tilde{\mathbf{X}}^H\mathbf{H}^H\mathbf{H}\tilde{\mathbf{X}} + \frac{p_E}{M}\mathbf{X}^H\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{G}^H\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X} \\ &\quad + \frac{\mathbf{N}^H\mathbf{N}}{M} + \mathbf{S}_1\end{aligned} \tag{70}$$

where

$$\begin{aligned}\mathbf{S}_0 &= \frac{\sqrt{p_B}}{M}\tilde{\mathbf{X}}^H\mathbf{H}^H\mathbf{N}^H + \frac{\sqrt{p_B}}{M}\mathbf{N}^H\mathbf{H}\tilde{\mathbf{X}} \\ \mathbf{S}_1 &= \mathbf{S}_0 + \frac{\sqrt{p_Bp_E}}{M}\tilde{\mathbf{X}}^H\mathbf{H}^H\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X} + \frac{\sqrt{p_E}}{M}\mathbf{X}^H\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{G}^H\mathbf{N} \\ &\quad + \frac{\sqrt{p_Bp_E}}{M}\mathbf{X}^H\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{G}^H\mathbf{H}\tilde{\mathbf{X}} + \frac{\sqrt{p_E}}{M}\mathbf{N}^H\mathbf{G}\boldsymbol{\alpha}^{\frac{1}{2}}\mathbf{X}\end{aligned} \tag{71}$$

where $\mathbf{H}$, $\mathbf{G}$ and $\mathbf{N}$ are mutually independent. As M grows indefinitely, we have

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{R} \approx p_B\tilde{\mathbf{X}}^H\mathbf{D}\tilde{\mathbf{X}} + \beta_n\mathbf{I}_\tau \\ \mathcal{H}_1 &: \mathbf{R} \approx p_B\tilde{\mathbf{X}}^H\mathbf{D}\tilde{\mathbf{X}} + p_E\beta_g\mathbf{X}^H\boldsymbol{\alpha}\mathbf{X} + \beta_n\mathbf{I}_\tau\end{aligned} \tag{72}$$

where $\mathbf{D} = \text{diag}(\beta_{h,1}, \ldots, \beta_{h,K})$.

For pilot matrix $\mathbf{X}$, its (k,i)-th entry satisfies $\mathbb{E}\{|x_{k,i}|^2\} = 1$. While for the modified $\tilde{\mathbf{X}} = \boldsymbol{\Xi}\hat{\mathbf{X}}$, where $\hat{\mathbf{X}} = [\mathbf{X}_A, \mathbf{W}\mathbf{X}_B]$, its (k,i)-th entry satisfies

$$\mathbb{E}\left\{|\tilde{x}_{k,i}|^2\right\} = \left\{\begin{array}{ll}|\xi_k|^2, & 1 \le i \le a\tau \\ |w_k\xi_k|^2, & a\tau + 1 \le i \le \tau\end{array}\right. . \tag{73}$$

Given (72) and (73), (9) is attainable.

References

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] M. T. L., "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Oct. 2010.

[3] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys & Tuts.*, vol. 21, no. 4, pp. 3109–3132, 2019.

[4] J. Xu, W. Xu, J. Zhu, D. W. K. Ng, and A. L. Swindlehurst, "Secure massive MIMO communication with low-resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 53, no. 6, pp. 3265–3278, May. 2019.

[5] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[6] 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.

[7] M. A. Teeti, “Downlink secrecy rate of one-bit massive MIMO system with active eavesdropping,” *IEEE Access*, vol. 8, pp. 37 821 –37 842, Feb. 2020.

[8] X. Weiyang, X. Shengbo, and L. Bing, “Detection of pilot spoofing attack in massive MIMO systems,” in *Proc. 2019 IEEE Conf. on Commun. (ICC)*, Shanghai, China, May. 2019, pp. 1–6.

[9] D. Darsena, G. Gelli, I. Iudice, and F. Verde, “Design and performance analysis of channel estimators under pilot spoofing attacks in multiple-antenna systems,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3255–3269, Apr. 2020.

[10] X. Zhou, B. Maham, and A. Hjorungnes, “Pilot contamination for active eavesdropping,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, p. 903907, Frb. 2012.

[11] K. Cumanan, H. Xing, P. Xu, G. Zheng, X. Dai, A. Nallanathan, Z. Ding, and G. K. Karagiannidis, “Physical layer security jamming: Theoretical limits and practical designs in wireless networks,” *IEEE Access*, vol. 5, pp. 3603–3611, Dec. 2017.

[12] J. Vinogradova, E. Bjrnson, and E. G. Larsson, “Detection and mitigation of jamming attacks in massive MIMO systems using random matrix theory,” in *Pro. 2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Edinburgh, UK, Jul. 2016, pp. 1–6.

[13] ——, “Jamming massive MIMO using massive MIMO: Asymptotic separability results,” in *Pro. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 3454–3458.

[14] H. Akhlaghpasand, S. M. Razavizadeh, E. Bjrnson, and T. T. Do, “Jamming detection in massive MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 242–245, Nov. 2017.

[15] Q. Xiong, Y.-C. Liang, K. H. Li, and Y. Gong, “An energy-ratio-based approach for detecting pilot spoofing attack in multiple-antenna systems,” *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 932–940, Jan. 2015.

[16] T. T. Do, H. Q. Ngo, T. Q. Duong, T. J. Oechtering, and M. Skoglund, “Massive MIMO pilot retransmission strategies for robustification against jamming,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 58–61, Nov. 2016.

[17] L. Xiaoming, B. Li, H. Chen, Z. Sun, Y.-C. Liang, and C. Zhao, “Detecting pilot spoofing attack in MISO systems with trusted user,” *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 314–317, Feb. 2019.

[18] X. Shengbo, W. Xu, C. Pan, and M. Elkashlan, “Detection of jamming attack in non-coherent massive SIMO systems,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2387–2399, Sep. 2019.

[19] S. Xu, W. Xu, H. Gan, and B. Li, “Detection of pilot spoofing attack in massive MIMO systems based on channel estimation,” *Signal Process.*, vol. 169, pp. 1–9, Apr. 2020.

[20] H.-M. Wang, K.-W. Huang, and T. A. Tsiftsis, “Multiple antennas secure transmission under pilot spoofing and jamming attack,” *IEEE J Sel. Areas Commun.*, vol. 36, no. 4, p. 860876, Apr. 2018.

[21] T. T. Do, E. Bjrnson, E. G. Larsson, and S. M. Razavizadeh, “Jamming-resistant receivers for the massive MIMO uplink,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 210–223, Aug. 2017.

[22] J. K. Tugnait, “Pilot spoofing attack detection and countermeasure,” *IEEE Trans. Commun.*, vol. 66, no. 5, p. 20932106, May 2018.

[23] N. Gao, Z. Qin, and X. Jing, “Pilot contamination attack detection and defense strategy in wireless communications,” *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 938–942, Jun. 2019.

[24] W. Wang, N. Cheng, K. C. Teh, X. Lin, W. Zhuang, and X. Shen, “On countermeasures of pilot spoofing attack in massive MIMO systems: A double channel training based approach,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6697–6708, Jul. 2019.

[25] A. Berk, M. Krunz, and O. O. Koyluoglu, “Vulnerabilities of massive MIMO systems to pilot contamination attacks,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1251–1263, May 2019.