Pilot Contamination Attack Detection for NOMA in 5G Mm-Wave Massive MIMO Networks

Ning Wang, Long Jiao, Amir Alipour-Fanid, Monireh Dabaghchian and Kai Zeng, *Member, IEEE*

***Abstract*—Power non-orthogonal multiple access (NOMA) has been considered as a new enabling technology in 5G communication. In this paper, we introduce the problem of pilot contamination attack (PCA) on NOMA in millimeter wave (mmWave) and massive MIMO 5G communication. Due to the new characteristics of NOMA such as superposed signals with multi-users, PCA detection faces new challenges. By harnessing the sparseness and statistics of mm-Wave and massive MIMO virtual channel, we propose two effective PCA detection schemes for NOMA tackling static and dynamic environments, respectively. For the static environment, the problem of PCA detection is formulated as a binary hypothesis test of the virtual channel sparsity. For the dynamic environment, the statistic of the peaks in the virtual channel is leveraged to distinguish the contamination state from the normal state. A peak estimation algorithm and a machine learning based detection framework are proposed to achieve high detection performance. To further optimize the proposed scheme, a feature selection algorithm and an optimization model considering the detection accuracy and detection delay are presented. Simulation results evaluate and confirm the effectiveness of the proposed detection schemes. The detection rate can approach 100% with 10^{-3} false alarm rate in the static environment and above 95% in the dynamic environment under various system parameters.**

I. Introduction

Non-orthogonal multiple access (NOMA) has been recognized as one of the key enabling technologies for fifth-generation mobile networks (5G) [1]. Via non-orthogonal resource allocation, NOMA can improve spectral efficiency and achieve massive connectivity with low transmission latency and signaling cost. Specifically, power NOMA is capable of bringing a new power dimension to perform multiplexing in the existing access domain, where it can be regarded as superposing multiple signals into one orthogonal resource. NOMA transmission with mulitple-input and mulitple-output (MIMO) and millimeter Wave (mmWave) can achieve even higher spectral efficiency.

Despite its great potential in 5G communications, NOMA is vulnerable to pilot contamination attack (PCA), in which an attacker can send the same pilot signals as that of legitimate users (LUs) [2]. In the channel training phase, the PCA attacker can inject the same pilot as the legitimate user's to control the channel estimation result and affect the precoding process. If there are not any countermeasures, the attacker not only can degrade the signal reception at the legitimate receiver, but also can cause confidential information leakage in the communication system [3].

Ning Wang, Long Jiao, Amir Alipour-Fanid, Monireh Dabaghchian and Kai Zeng are with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, US 22030 e-mail: ({nwang5, ljiao, aalipour, mdabaghc, kzeng2}@gmu.edu).

Manuscript received November 11, 2018; revised XX XX, XXXX.

Existing countermeasures to PCA do not apply to NOMA scenario. Most existing PCA detection methods assume that the pilot signals of multiple users are assigned into different orthogonal resource blocks, where they can be distinguished by the orthogonality or time division between two pilot signals [3], [4]. Anytime more than one transmitter get detected by these detection techniques, they raise alarms. For instance, random phase-shift keying (PSK) pilot [5] and special beamformer [6] are based on the hypothesis that each legitimate user should be assigned to a single orthogonal resource block. The existence of more than one transmitter in an orthogonal resource block leads to a different PSK or beamformer signal, which will be detected as a PCA in the detection process.

PCA in NOMA has different characteristics compared to the PCA in existing communication networks. Based on NOMA, two or more LUs are assigned to one time-division-duplex (TDD) or orthogonal resource block [7]. More importantly, superimposed pilot schemes have been proposed for channel estimation in NOMA up-link [8]–[10]. More than one LU could transmit their pilot signals to the base station (BS) or access point (AP) at the same single orthogonal resource block for channel estimation. Under this case, the false alarm rate of most existing pilot-contamination attack detection methods will significantly increase.

For this issue, in this paper, we propose to exploit a unique characteristic in 5G wireless communication to counter PCA in NOMA. In 5G communication networks, mmWave possesses high propagation loss and directionality which is highly sensitive to legitimate users' positions [11]. Furthermore, supported by massive MIMO, mmWave virtual channel, a representation of geometry channel model [12], will exhibit sparsity that can be used to distinguish PCA from normal states in NOMA.

Based on the virtual channel model, we present two effective PCA detection schemes tackling static and dynamic environments, respectively. Based on channel sparseness, PCA detection is transformed into a binary hypothesis test for the static environment, where the number of multipaths is constant. For the dynamic environment where the number of multipaths is a random variable, we propose a peak estimation and machine learning based framework for PCA detection. Moreover, we present a feature selection algorithm and an optimization model jointly considering detection accuracy and detection delay to optimize the proposed scheme. Simulation results show that our proposed detection scheme can reach a high detection accuracy. The detection rate reaches 100% with 10^{-3} false alarm rate in the static environment and above 95% in the dynamic environment.

The contributions of this work lie in three aspects:

- To the best of our knowledge, this is the first work to

TABLE I: Summary of abbreviations.

Abbreviations	Notations
NOMA	Non-orthogonal multiple access
PCA	Pilot contamination attacks
MIMO	Multiple input multiple output antenna
Mm-Wave	Millimeter wave communication
CSI	Channel state information
SIC	Successive interference cancellation
BS	Base station
AP	Access point
LU	Legitimate user
SC	Superimposition coding
PSK	Phase shift keying
CCD	Connected component districts
ROC	Receiver operating characteristic curve
MLM	Machine learning model
SVM	Support vector machine
ELM	Extreme learning machine
LOG	Logistics regression

investigate the PCA detection in NOMA. Existing PCA detection techniques mainly focus on massive MIMO scenarios, while the risk of PCA in NOMA is not yet widely recognized.

- We utilize a novel characteristic of the massive MIMO mmWave channel to achieve PCA detection. Virtual channel model is a geometry channel model for mmWave. With massive MIMO, this channel model will show an integrated characteristic, called sparsity, which can be utilized to distinguish PCA from the normal case.
- We propose two effective PCA detection schemes for static and dynamic environments, respectively, where the detection strategies rely solely on signal processing in BS (or AP) without any modification requirements on the users. In addition, we present a feature selection algorithm and an optimization model to optimize the delay (the number of channel measurements) with a constraint on the detection accuracy.

This paper is the extended version of our conference paper [13], which has developed a basic mathematical model and solution to address the PCA detection problem in NOMA based on the virtual channel. Comparing with our earlier work, this paper (a) presents a detailed analysis of the PCA detection problem in NOMA, (b) introduces a features selection algorithm to choose the best features from a group of statistical properties, (c) provides an optimization model jointly considering detection accuracy and detection delay, and (d) provides more simulations of several popular machine learning algorithms for comparison.

The rest of this paper is organized as follows. Section II introduces related works. Section III describes the system model, challenges and motivation. In Section IV, the analysis of the sparsity of virtual channel is provided. PCA detection schemes for static and dynamic environments are proposed in Section V and VI, respectively. Section VII provides the simulation and evaluation results. Section VIII concludes this paper.

II. Related work

In this section, we introduce the technical background of this study, including NOMA, PCA detection schemes, and virtual channels.

A. NOMA

NOMA can be divided into power domain and code-domain NOMA. Here, we focus on power domain NOMA [1], which can serve multiple users in the same time slot, and multiple access is realized by allocating different power levels to different users. NOMA can transmit superposed signals in one OFDMA subcarrier or spread code to improve spectral efficiency. In power NOMA, a transmitter can use superimposition coding (SC) to form the superposed signal, and the corresponding receiver can extract the corresponding signals based on successive inference cancellation (SIC) technology [7]. Moreover, in order to trade off between complexity and efficiency, users are usually grouped in pairs (i.e., two users in one group) [14]. Therefore, a notable characteristic of NOMA communication is that there are at least two legitimate users' messages in a superimposed signal at the same OFDM resource block.

B. PCA detection

PCA was first proposed in [2]. Later on, an overview on PCA for massive MIMO was given in [3], [4]. According to [3], two main methods can be used to detect PCA: random pilots and special beamforming.

Random phase-shift keying (PSK) pilot was first proposed in [5], and an improved variation which considers the case of three or more observations is provided in [15]. Basically, there are two steps in this scheme. Firstly, the legitimate user (LU) transmits a sequence of random PSK pilot symbols, which are chosen independently from an N-PSK constellation. Secondly, after obtaining these PSK symbols, BS detects the phase change between two symbols and estimates whether there are an active attackers over this period of time. If the received signals have valid PSK symbols, there is no PCA, otherwise, a PCA is detected. The other random pilot scheme is based on random frequency shifts [16]. LU deliberately introduces multiple random frequency shifts when transmitting the pilot sequence. Then BS can detect whether there is a PCA by calculating the autocorrelation matrix for the training phase. In the absence of the attack, the dimension of the signal subspace of the autocorrelation matrix should be one. While in the case of attacking, the dimension should equal to two. Moreover, BS can use the minimum description length algorithm to discern this phenomenon [16].

A special beamformer has been designed to detect the activity of PCA [6]. In the first phase, LU transmits a training signal to the BS which uses this training signal to estimate the channel. In the second phase, BS establishes a specialized beamformer that ensures an agreed value of the received signal at LU close to 1, and transmits the signals to LU by using the same beamformer. If the PCA is active, the channel estimation in the second phase will inevitably be affected, i.e., the agreed value at LU will sharply drop under the case of PCA.

Unfortunately, these existing PCA schemes cannot be directly applied to the NOMA scenario in 5G networks. More details

on the challenges detecting PCA in NOMA are provided in Section III-A.

C. Virtual channel

Virtual channel was first introduced in [12] and has been applied in mm-Wave and MIMO communication systems [17]. It can be acquired by a fast estimation method [18]. Virtual channel model is based on virtual angle-of-arrival (AoA) and virtual angle-of-departure (AoD), and it can be seen as a mapping of the real channel in a special space. Furthermore, when the number of antennas is large enough such as in massive MIMO scenarios, the virtual channel matrix presents a sparsity characteristic. Fortunately, supported by mmWave and massive MIMO 5G communication, the tiny wavelengths allow for dozens to hundreds of antenna elements to be placed in an array on a relatively small physical platform [19]. The integration of massive MIMO and mmWave can present some unique properties, such as sparsity and being time-varying in virtual channel, which are exploited to achieve efficient PCA detection in this paper.

III. Challenges, Motivation, and System Model

In this section, we will introduce the challenges and motivation of this paper, and illustrate the system and attack model of this study.

A. Challenges

To achieve a desirable PCA detection in NOMA, we have to tackle the following two issues:

1) **Existing methods are not suitable for NOMA.** Most of the existing PCA detection schemes assume that there is only one LU sending pilot sequence at the channel estimation phase. Hence, they achieve the PCA detection based on the abnormal case that there are multiple transmitters. For example, for the random pilot scheme [5], [16], [20], LU sends a sequence of random PSK pilot symbols. If there are other transmitters who are sending pilot sequence at the same time, BS will receive a number of invalid PSK symbols that will be recognized as a PCA case. However, for NOMA communications, there have been more than one LU using the same one orthogonal resource at the same time, and that will influence the PSK pilot symbols for PCA detection. As a result, the false alarm rate in the detection schemes will be increased significantly. A similar situation occurs in the special beamforming scheme [6]. During the PCA detection process, BS receives the special beam which may have been degraded by other LUs' messages due to the existence of multiple LUs in one NOMA signal. In this situation, BS is unable to accurately distinguish PCA case from the normal case based on the special beamforming.
2) **PCA detection should not burden LUs and normal communication.** Existing PCA detection methods do not sufficiently consider the low complexity profile of LUs in 5G communication, where low-latency and high efficiency are required. Under the conditions of low complexity and severe energy constraints on many LUs such as low-cost IoT devices, any extra signal processing requirements for LUs other than normal communication should be considered an overload. On the other hand, we could allow certain extra signal processing or computation at AP or BS, which usually has more resources and power. Moreover, a desirable physical layer security solution should not require any changes to LUs or introduce any extra communication overhead. Therefore, mechanisms in existing methods, such as complex pilot design [5], [16], [20] and retransmission [6], are not preferable for resource-constrained devices.

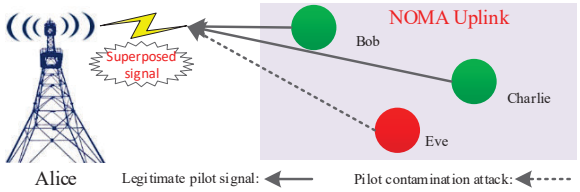


Fig. 1: PCA in a NOMA communication pair.

In this work, for the first time, we propose techniques for the PCA detection in NOMA communications based on 5G technology characteristics. This technique do not introduce any computation overhead at LUs nor any change in the transmission signals or communication protocol.

B. Motivation of our methods

We have the following three subtle observations on the characteristics of mmWave channel, wavelengths, and sparsity.

- A representation of the geometry channel for mmWave, i.e., virtual channel, exhibits a sparsity characteristic when there is a large enough number of antennas (as shown in IV-A);
- The tiny wavelength in mmWave allows for hundreds of antenna elements to be placed in an array on a relatively small physical platform [19];
- Different numbers of transmitters result in different levels of the sparsity in the virtual channel (as shown in IV-A).

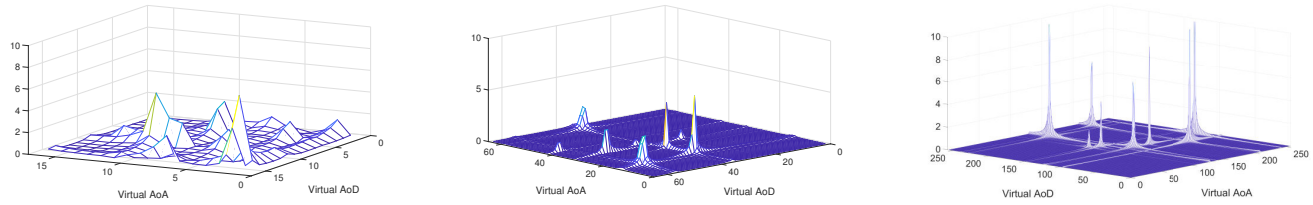
Inspired by these observations, to achieve a desirable PCA detection in NOMA, we present two detection schemes tackling the static and dynamic scenarios, respectively.

C. System model

Here, we focus on uplink NOMA 5G communication, and Fig. 1 illustrates the communication system model. Without loss of generality, Alice-Bob-Charlie-Eve model (i.e., a base station Alice, two LUs Bob and Charlie, and Eve who launches PCA) is used to represent the NOMA communication. Bob and Charlie equipped with N_t antennas compose a NOMA pair. While, Alice with N_r antennas observes the received training signal during a window of length T. According to [8] [9] [10], the superimposed pilot signal $\mathbf{Y} \in C^{N_r \times T}$ is given by

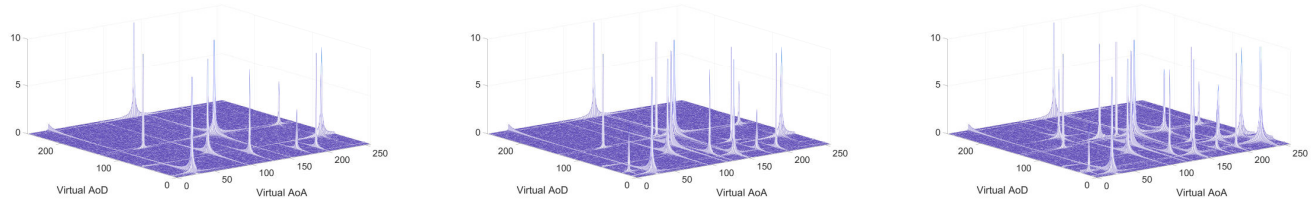
$$\mathbf{Y} = \mathbf{H}_B \mathbf{X}_B + \mathbf{H}_C \mathbf{X}_C + \mathbf{W}, \tag{1}$$

where $\mathbf{H}_B \in \mathbb{C}^{N_r \times N_t}$ and $\mathbf{H}_C \in \mathbb{C}^{N_r \times N_t}$ represent the mmWave MIMO channel for Alice-Bob and Alice-Charlie,



(a) Antennas are 16×16. (b) Antennas are 64×64. (c) Antennas are 256×256

Fig. 2: Virtual channel performance with respect to an increasing number of antennas. (The number of multipaths is 8.)



(a) Two transmitters. (b) Three transmitters. (c) Four transmitters.

Fig. 3: Virtual channel under different transmitters when the antenna is 256×256 and the number of multi-path is 5.

respectively. $\mathbf{X}_B \in \mathbb{C}^{N_t \times T}$ and $\mathbf{X}_C \in \mathbb{C}^{N_t \times T}$ are the transmitted pilot signals, where N_t rows of each transmitted pilot constitute of the length-T training sequences transmitted from N_t antennas, and $\mathbf{W} \in \mathbb{C}^{N_r \times T}$ is the additive Gaussian noise independent of $\mathbf{H}$ and $\mathbf{X}$.

D. Attacking model

After obtaining $\mathbf{Y}$, Alice estimates the corresponding Alice-Bob and Alice-Charlie channels based on channel state information (CSI). Using CSI, Alice can establish the corresponding beamform for Bob and Charlie, or extract the corresponding message from Bob and Charlie' transmitted pilot signal.

If there is a pilot contamination attack, attacker Eve using the same number of antennas as Bob and Charlie can send the same pilot sequence. The received signal at Alice becomes

$$\mathbf{Y} = \mathbf{H}_B\mathbf{X}_B + \mathbf{H}_C\mathbf{X}_C + \mathbf{H}_E\mathbf{X}_E + \mathbf{W}, \tag{2}$$

where $\mathbf{H}_E \in \mathbb{C}^{N_r \times N_t}$ is the mmWave MIMO channel for Alice-Eve, and $\mathbf{X}_E \in \{\mathbf{X}_B, \mathbf{X}_C\}$ is the imitated pilot sequence by Eve. Notice that Eve may have more than one transmitter. Under the multiple transmitters case, $\mathbf{X}_E$ is substituted with $\mathbf{X}_E = \{\mathbf{X}_{E,1}, \mathbf{X}_{E,2}, ..., \mathbf{X}_{E,n}\}$ where n indicates the number of transmitters at Eve.

With the same pilot sequence to LUs, Eve can easily influence the channel estimation at Alice. If there are no security measures, normal communications will be affected by the attacker. For instance, due to the injection of the pilot sequence by Eve, Alice may point its beam to Eve when beamforming is used. Thus, Eve will obtain an illegitimate advantage for eavesdropping. Even worse, Eve may have an opportunity to replace Bob (or Charlie) in a NOMA pairing, which leads to a high risk of information leakage. In addition, it should be noted that time and frequency synchronization is an important issue for multiuser communications [21]. If the attacker cannot synchronize with the legitimate system, the attacker may not be able to replace LUs but may still gain some advantage in eavesdropping by affecting the channel estimation. The attacker may even result in a DoS attack by disrupting the channel estimation. However, it is not the goal of the pilot contamination attack studied in this paper. In this paper, the attacker's goal is to affect the channel estimation result through pilot contamination in order to gain eavesdropping advantages in the following communication. Therefore, we consider the worst-case scenario where a powerful attacker has the capability of synchronizing with the legitimate communication system. This kind of attacker could be a compromised user in the system.

IV. Sparsity Characteristic of Virtual Channel

In this section, we will introduce the virtual channel model and its sparsity under massive MIMO.

Formally, a sparse multipath structure of mm-Wave can be represented by a geometry channel model with scatters formed by ray tracing [12], [18]. The channel matrix is

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{\rho}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_r(\phi_{r,l})\mathbf{a}_t^*(\phi_{t,l}), \tag{3}$$

where N_t and N_r are the number of transmitter and receiver antennas, respectively. ρ denotes the average path-loss. L denotes the number of scatters and α_l is the corresponding fading coefficients with zero mean complex Gaussian distribution. $\phi_{r,l} \in (0, 2\pi]$ and $\phi_{t,l} \in (0, 2\pi]$ denote the physical AoD and AoA angles at the transmit and receive sides. Vectors $\mathbf{a}_r(\phi_{r,l}) \in {}^{N_r \times 1}$, and $\mathbf{a}_t(\phi_{t,l}) \in {}^{N_t \times 1}$ are the antenna array responses.

The virtual channel representation characterizing the mmWave massive MIMO channel can be written as

$$\mathbf{H} = \mathbf{U}_r\mathbf{H}_V\mathbf{U}_t^H, \tag{4}$$

where $\mathbf{U}_r \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{U}_t \in \mathbb{C}^{N_t \times N_t}$ are unitary discrete Fourier transform matrices. Thus, $\mathbf{H}_V \in \mathbb{C}^{N_r \times N_t}$ is the virtual

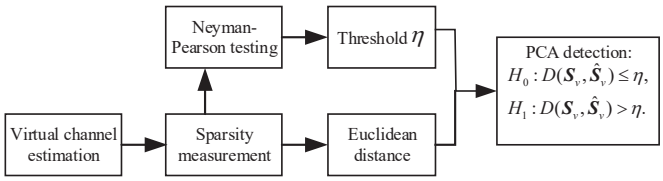


Fig. 4: Detection strategy under static environment.

channel matrix, where the entry is the corresponding path gain, and $\mathbf{H}$ and $\mathbf{H_V}$ are unitarily equivalent.

A. The characteristics of the sparsity of virtual channel

According to [12], for the discrete model, the number of paths that lie in each spatial bin in the virtual channel affects the statistics of $\mathbf{H}_V$. Consider massive MIMO mmWave channel, there will be two characteristics of the sparsity for the virtual channel as shown below.

1) **The virtual channel is sparse when the number of antennas is large enough.** As the number of antennas increases, fewer paths would contribute to each spatial bin in the virtual channel. An example of the virtual channel under different antennas is shown in Fig. 2. We can see that the power of the virtual channel is more and more concentrated as the number of antennas increases. In other words, the virtual channel becomes more and more sparse.
2) **The sparsity of virtual channel can reflect the number of transmitters.** It is easy to understand that as there are multiple transmitters, the virtual channel will show more peaks than that of only one transmitter, since the number of multipaths will be increased under the case of multiple transmitters. Fig. 3 illustrates this phenomenon. It is obvious that as the number of transmitters increases, the number of the virtual channel peaks significantly increases.

The aforementioned characteristics can be used to distinguish normal and PCA cases, even in the presence of a superposed signal under NOMA scenarios. Next, we propose two detection schemes based on the characteristics of the virtual channel.

V. Pilot contamination detection under static environment

In this section, based on the sparsity of the virtual channel, we introduce the PCA detection scheme for a static environment, where the position of LUs is fixed and the communication environment is stable. An example of such an application is a smart home scenario, where LUs are various monitors. Since the position of these monitors is fixed, for the channel between AP and a LU, the number of multipaths approximates to a constant. Here, we directly consider the sparsity of the virtual channel as a detection metric and establish a binary hypothesis test to achieve the PCA detection, as illustrated in Fig. 4.

A. Detect PCA under static environment

After obtaining the virtual channels, we first use a filter with a threshold δ_T to filter the values in the virtual channel. This threshold δ_T adopts empiric values i.e., $\delta_T \in [0.5, 1]$, where elements with amplitudes smaller than δ_T in the virtual channel matrix are set as zero and other elements are remained. Then, the sparsity of the obtained virtual channel can be illustrated by a simple metric of matrix, i.e.,

$$S_i = \|\mathbf{H}_{V,i}\|_0, \tag{5}$$

where $\|\cdot\|_0$ means ℓ_0-norm, and $\mathbf{H}_{V,i}$ denotes the normalization of the measurement of the i-th virtual channel.

Let $\mathbf{S}_v = [S_1, S_2, ..., S_v]$, where v is the number of independent measurements of the virtual channel during a window of period T. Euclidean distance is introduced to discriminate PCA from the normal communication. Euclidean distance can be given by

$$D(\mathbf{S}_v, \hat{\mathbf{S}}_v) = \left\|\mathbf{S}_v - \hat{\mathbf{S}}_v\right\|_2, \tag{6}$$

where $\|\cdot\|_2$ means ℓ_2-norm, $\mathbf{S}_v$ denotes the vector of the sparsity involving the channels under the standard situation, i.e., normal communication case, and $\hat{\mathbf{S}_v} = [\hat{S}_1, \hat{S}_2, ..., \hat{S}_v]$ indicates the vector of sparsity involving the current channels which is being tested.

If the current communication is normal, the sparsity of the current virtual channel is the same as that of normal virtual channel. Thus, S_i and $\hat{S}_i$ have the same statistical properties. Hence, the Euclidean distance between $\mathbf{S}_v$ and $\hat{\mathbf{S}_v}$ is small. In contrast, if there is a PCA, the sparsity of the current virtual channel is different from that of the normal virtual channel. The Euclidean distance between $\mathbf{S}_v$ and $\hat{\mathbf{S}_v}$ would be larger. In this way, we adopt a directly strategy which chooses an appropriate threshold to detect PCA, i.e., hypothesis testing.

B. Hypothesis testing

Typically, the threshold is chosen according to different criteria for performance. Here, we apply the Neyman-Pearson testing to choose the threshold η, where the detection minimizes the miss rate subject to a maximum tolerable constraint on a given false alarm rate.

Based on the analysis shown in section V-A, we define the following hypothesis:

1) $H_0 : D\left(\mathbf{S}_v, \hat{\mathbf{S}}_v\right) \le \eta$,
2) $H_1 : D\left(\mathbf{S}_v, \hat{\mathbf{S}}_v\right) > \eta$.

The null hypothesis H_0 represents that the current superposed signals are under the normal communication, while the alternative hypothesis H_1 indicates that the superposed signals include PCA.

By observing the virtual channel under the normal case and PCA case, we find that the value of S_i and $\hat{S}_i$ can be approximated as a random variable following a normal distribution. Fig. 5 shows the probability distribution of S_i and $\hat{S}_i$ and the fitting curves with normal distributions. We can see that normal distributions can fit the data under the

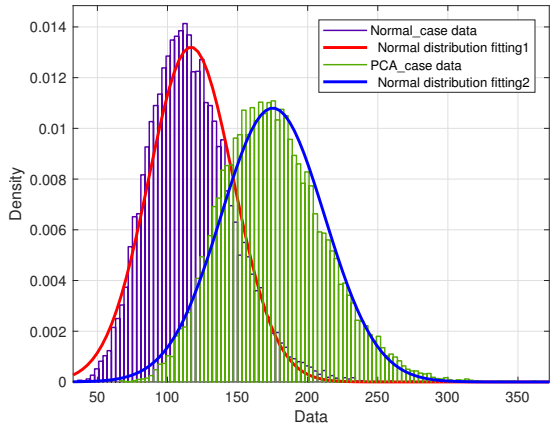


Fig. 5: The results of normal distribution fitting for the normal and PCA case (20,000 times Monte Carlo simulation).

normal case and PCA case well. Thus, an effective detection scheme can be given by using Neyman-Pearson testing based on this observation. For simplification of analysis, let $S_i \sim CN(\mu_a, \sigma_a^2)$ and $\hat{S}_i \sim CN(\mu_k, \sigma_k^2)$, $k \in \{a, b\}$, where a indicates normal case and b indicates PCA case. Thus, $S_i - \hat{S}_i$ follows a bivariate random normal distribution. Consider a hypothesis test statistic $\xi = D(\mathbf{S}_v, \hat{\mathbf{S}}_v)^2$, the detection based on Neyman-Pearson testing can be given as follows.

Under the null hypothesis H_0, since the distribution of S_i and $\hat{S}_i$ are consistent, the Euclidean distance follows a chi distribution with v degrees of freedom, and the hypothesis test statistic $\xi = D\left(\mathbf{S}_v, \hat{\mathbf{S}}_v\right)^2$ is a chi-square random variable with v degrees of freedom.

Thus, the corresponding probability distribution function (PDF) is given by

$$P_{H_0} = \frac{\xi^{\frac{v}{2}-1} exp(-\frac{\xi}{2})}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}, \tag{7}$$

where $\Gamma(x) = \int_0^\infty u e^{x-1} du$ is Gamma function.

The false alarm rate $P_{FA} = P\{\xi > \eta \,|H_0\}$ can be given by

$$P_{FA} = \int_\eta^\infty P_{H_0}(\xi) dW = 1 - Q_{\chi_v^2}(\eta), \tag{8}$$

where $Q_{\chi_v^2}(\cdot)$ is the cumulative distribution function of the central chi-square distribution.

Next, based on the given false alarm rate P'_{FA} we can obtain the corresponding threshold η', i.e.,

$$\eta' = Q_{\chi_v^2}^{-1}(P'_{FA}), \tag{9}$$

where $Q_{\chi_v^2}^{-1}(\cdot)$ is the inverse function of $Q_{\chi_v^2}(\cdot)$.

Accordingly, under hypothesis H_1, the test statistic ξ follows a non-center chi-square distribution with v degrees of freedom, and the non-centrality parameter is

$$\lambda_{H_1} = \sum_{j=1}^{v} \frac{\mu_{ab,j}^2}{\sigma^2}, \tag{10}$$

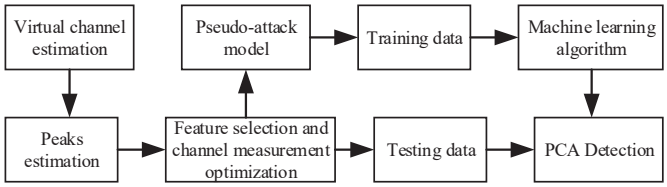


Fig. 6: Detection strategy under dynamic environment.

where $\mu_{ab} = \mu_b - \mu_a$. Thus, the corresponding PDF is

$$P_{H_1} = \frac{1}{2}\left(\frac{\xi}{\lambda_{H_1}}\right)^{\frac{v-2}{4}} exp\left(-\frac{1}{2}(\xi + \lambda_{H_1})\right) I_{\frac{v}{2}-1}(\sqrt{\xi\lambda_{H_1}}), \tag{11}$$

where $I_v(\cdot)$ is the first kind Bessel function of order v, with the definition of

$$I_r(\theta) = \sum_{k=0}^{\infty} \frac{\left(\theta/2\right)^{2k+r}}{k!\Gamma(r+k+1)}. \tag{12}$$

Then, miss detection rate $P_{MD} = P\{\xi \le \eta \,|H_1\}$ can be described as

$$P_{MD} = P\{\xi > \eta \,|H_1\} = 1 - Q_{\chi^2_{2M}(\lambda_{H_1})}(\eta), \tag{13}$$

where $Q_{\chi_v^2(\lambda_{H_1})}(\cdot)$ is the cumulative distribution function of the non-central chi-square distribution, i.e.,

$$Q_{\chi_v^2(\lambda_{H_1})}(\eta) = \int_{-\infty}^{x} P_{H_1}(t) dt, \tag{14}$$

and the detection rate is $P_D = 1 - P_{MD}$.

VI. Pilot Contamination Detection under Dynamic Environment

In this section, we consider a dynamic environment. Due to the mobility of the users, the number of multipaths in the wireless channel is variable. For such a scenario, we propose to use the number of peaks in the virtual channel matrix for detection, then utilize a machine learning based technique to solve the detection problem. Furthermore, we present a method for appropriate feature selection and establish an optimization model jointly considering the detection accuracy and detection delay. The strategy of this detection is shown in Fig. 6. It is worthy to note that this detection scheme for dynamic scenario can be used in the static scenario as well.

A. Peak estimation

1) Analysis of dynamic environment: Although the number of multipaths in a channel is unsteady under dynamic environments, they have a certain regularity. Based on the mmWave channel model, multipaths present a state of clusters, where there are several clusters and each cluster is composed of several multipaths. Moreover, the number of cluster and the number of cluster subpaths are approximated by random normal distribution [22] [23]. Table II shows the statistical properties of the number of clusters and cluster sub-paths under mmWave channel model [22].

From Table II, we can see that the number of the cluster and the number of the cluster sub-paths are following a certain

TABLE II: Statistical properties of the number of cluster and cluster sub-paths based on mm-Wave channel model [22].

Type of statistic	Quantity	Measured (μ, σ)	Simulated (μ, σ)
28 GHz wideband channel, NLOS	Number of clusters	(3.4,2.1)	(3.2,2.1)
urban outdoor, 20dB lobe threshold	Number of cluster sub-paths	(2.1,1.6)	(2.2,1.7)
28 GHz wideband channel, NLOS	Number of clusters	(4.1,2.3)	(4.0,2.4)
urban outdoor, 10dB lobe threshold	Number of cluster sub-paths	(2.0,1.7)	(2.1,1.6)

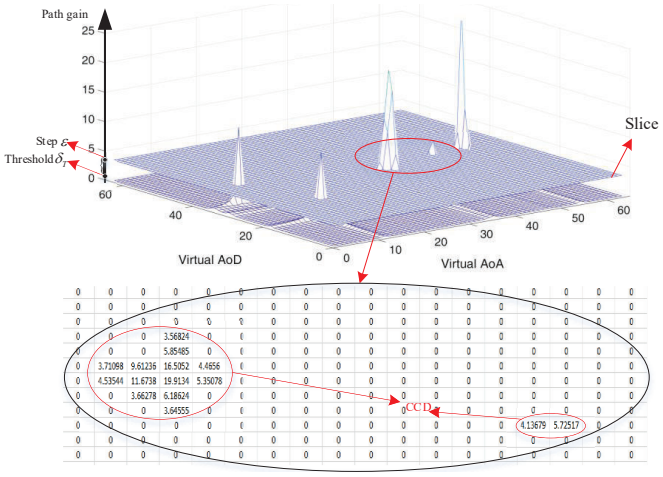


Fig. 7: Illustration of peak estimation.

statistical distribution corresponding to the environment. Here, we use the peaks in the virtual channel represent the multipaths in mmWave channel model. The definition about the peak of virtual channels is presented in *Definition 1*.

Definition 1: The peak of virtual channels: The peak of virtual channels is defined as the representation of multiplaths in mmWave channel model. The number of the peaks equals to the product of the number of the clusters and the number of the cluster sub-paths.

It is worthy to notice that we do not need a prior knowledge about the statistical properties of the channel for the proposed machine learning-based detection, since the training data in machine learning is used to adapt the environment. In addition, we find that 1) these peaks may overlap each other, 2) the values which compose the peaks are not always smooth, and 3) there are no regular methods that can be used to estimate the number of the peaks. Therefore, we present a feasible scheme to estimate the number of the peaks in virtual channel based on a connected component district (CCD).

Definition 2: Connected component district (CCD): In a matrix, if there is a set of nonzero points which are connected to each other, the region containing these points is defined as a CCD.

2) Peak estimation algorithm: Figure 7 illustrates the strategy to estimate the number of the peaks in virtual channel. The pseudocode of the proposed peak estimation algorithm is shown in Algorithm 1, which can be divided to several steps:

- Mitigate the influence of noise. A filter with a threshold δ_T is used to filter each virtual channel as the same way in Section V. The threshold adopts empiric values, i.e., $\delta_T \in [0.5, 1]$.

Algorithm 1 Peak number estimation algorithm

Initialization: Threshold δ_T, and step size ε.
Input: Virtual channel $\mathbf{H}_V$.
Output: The number of peaks N_{out}.

1) Slice: $K = \text{int}\left(\left(\max(h_1, h_2, ..., h_j)\right)/\varepsilon\right)$.
2) $k = 1, 2, ...K$;
3) **If** $h_{i,j} < \delta_T$, **do** $h_{i,j} \leftarrow 0$;
4) **Else** $h_{i,j} \geq \delta_T$,
5) Extract coordinates: $C_n^{(coordinate)} \leftarrow (i, j)$,
6) **End if**
7) **For** $n = 1, 2, ...N$
8) **If** $n = 1$
9) $a \leftarrow 1$, $b \leftarrow 1$, and $\mathbf{C}_b^{(CCD)}(a) \leftarrow C_1^{(coordinate)}$
10) **Else** $n > 1$
11) **For** all a and b
12) $D_{dis} = \text{Manhattan}(C_b^{(CCD)}(a), C_n^{(coordinate)})$
13) **End for**
14) **if any** $D_{dis} < 2$,
15) Add element: $C_a^{(CCD)}(b+1) \leftarrow C_n^{(coordinate)}$
16) **End if**
17) **If all** $D_{dis} \geq 2$
18) Add element: $b \leftarrow 1$ and $C_{a+1}^{(CCD)}(b) \leftarrow C_n^{(coordinate)}$
19) **End if**
20) **End if**
21) **End for**
22) Undate: $\delta_T = \delta_T + \varepsilon$
23) Update: $N_k^{CCD} = a$
24) Extract the largest peak number: $N_{out} \leftarrow \max(N^{CCD})$

- Slice the virtual channel. Slice the virtual channel by step ε. Thus, the matrix of the virtual channel on the slice is called as slice matrix. The total number of slices, denoted as K, can be given by

$$K = \text{int}\left(\max(h_{1,2}, ..., h_{i,j})/\varepsilon\right) \tag{15}$$

where $h_{1,1}, ..., h_{i,j}$ are the value of the nonzero elements in the virtual channel $\mathbf{H}_V$, and i and j are the location of the nonzero elements in the virtual channel matrix.

- Update CCD. Here, a Manhattan distance, denoted as D_{dis}, to determine whether two points are within the same CCD. The Manhattan distance is given by

$$\text{Manhattan}(O_1, O_2) = |i_{O_1} - i_{O_2}| + |j_{O_1} - j_{O_2}|, \tag{16}$$

where O_1 and O_2 denote two points within $\mathbf{H}_V$; i and j correspond to the coordinate of the value in the virtual channel matrix, denoted as $C_n^{(coordinate)}$, $n = 1, 2, ...N$ where N is the number of non-zero values in the slice matrix. $\mathbf{C}^{(coordinate)}$ is the set of the coordinate. In addition, $C_a^{(CCD)}(b)$ denotes the b-th point in the a-th CCD, and $\mathbf{C}^{(CCD)}$ is the set of the coordinate for all CCD. By comparing the Manhattan distance, all non-zero

TABLE III: List of statistical features.

Features	Description
f_1: Euclidean distance	$\sqrt{\sum_{\varphi=1}^{B_L}(p_{\alpha,\varphi}-p_{\beta,\varphi})^2}$
f_2: Total power	$\sum_{\varphi=1}^{B_L} p_{\alpha,\varphi}-\sum_{\varphi=1}^{B_L} p_{\beta,\varphi}$
f_3: Mean	$\frac{1}{B_L}\sum_{\varphi=1}^{B_L} p_{\phi,\varphi}$
f_4: Median	$\{(\varphi+1)/2\}th$
f_5: Absolute deviation	$median(\lvert p_{\phi,\varphi}-median(p_{\phi,\varphi})\rvert)$
f_6: Standard deviation	$\sqrt{\frac{1}{\varphi-1}\sum_{\varphi=1}^{B_L}(p_{\phi,\varphi}-mean(p_{\phi,\varphi}))^2}$
f_7: Skewness	$\frac{1}{B_L}\sum_{\varphi=1}^{B_L}\left(p_{\phi,\varphi}-mean(p_{\phi,\varphi})/\ f_6\right)^3$
f_8: Kurtosis	$\frac{1}{B_L}\sum_{\varphi=1}^{B_L}\left(p_{\phi,\varphi}-mean(p_{\phi,\varphi})/\ f_6\right)^4$
f_9: Mean square	$\frac{1}{B_L}\sum_{\varphi=1}^{B_L}\left(p_{\phi,\varphi}\right)^2$
f_{10}: Root mean square	$\sqrt{\frac{1}{B_L}\sum_{\varphi=1}^{B_L}\left(p_{\phi,\varphi}\right)^2}$
f_{11}: Pearson's Skewness	$3\cdot\left(mean(p_{\phi,\varphi})-median(p_{\phi,\varphi})\right)/f_6$
f_{12}: MAX	$Max(p_{\phi,\varphi})$
f_{13}: MIN	$Min(p_{\phi,\varphi})$

values $\mathbf{C}_n^{(coordinate)}$in the slice matrix are mapped into $C_a^{(CCD)}(b)$.

- Extract the largest number of CCD. After a successive measurement with a step ε for a virtual channel, the largest number of CCD can be extracted.

B. Machine learning based PCA Detection framework

Here, we introduce a machine learning based detection framework to detect PCA, including a feature selection, a pseudo-attacker model and an optimization model for the detection delay and detection accuracy. The present detection framework can be illustrated in Algorithm 2.

Algorithm 2 Machine learning based PCA detection

Initialization: Parameter λ

1) Repeat (for each episode)
2) Obtain the training data based on Pseudo-attack model.
3) Calculate the minimum of total cost function J.
4) Obtain the optimal size of B_L based on Eq. (21).
5) Train MLM based on the training data and B_L.
6) **For** $t=1,2,...B_L$
7) Select features based on selection algorithm.
8) Calculate the predictive value based on MLM.
9) **If** $q(\mathbf{x})=1$
10) Accept this message.
11) **Else**
12) Raise alarm.
13) **End If**
14) **End For**
15) **End** Repeat

1) Features and feature selection: Suppose that the number of successive independent channel measurement is B_L. Let $\mathbf{p}_\phi=\{p_{\phi,1},p_{\phi,2},...,p_{\phi,\varphi},...,p_{\phi,B_L}\}$, $\phi\in\{\alpha,\beta\}$ denote the number of peaks in the virtual channel, and subscript α and β denote the reference and current signals, respectively. Table III gives a group of popular statistical features [24].

Generally, for a classification problem, how to find the best features is an elemental question. Too many invalid features to extract will occupy unnecessary resource. Here, we introduce a feature selection algorithm to select the appropriate features for the PCA detection.

Define set $S=\{(X_i,y_i)\}_{i\in\{1,...,m\}}$, where the pair of ($X_i$,$y_i$) denotes the ith example and its corresponding label in set S, respectively. When training a predictive model for PCA detection, different features have different significance. We introduce vector $W=\{W_1,W_2,\cdots,W_l\}\in\mathbb{R}^l$ to denote the weight vector for all features. For instance, $W_i=0$ means the ith feature does not contribute useful information on prediction/identification and can be discarded. Therefore, given a design matrix $\mathbf{X}\in R^{m*l}$ and label vector $y\in\mathcal{Y}$, where $\mathcal{Y}=\{1,2\}$ is the class (label) set, our goal is to learn a predictive mapping $h(X_i,W)\rightarrow y_i$ for $i=1,2,\cdots,m$ such that the number of non-zeros in W, and the disagreement between $h(X_i,W)$ and y_i are minimized. To address this problem, we introduce the sparsity-inducing regularization terms to the conventional empirical loss function as follows [25]:

$$\min_W \mathcal{L}(y,h(\mathbf{X},W))+\lambda\Omega(W), \tag{17}$$

where $\mathcal{L}(\cdot)$ is the empirical loss function between the labels $y\in\mathcal{Y}$ and the predictions $h(\mathbf{X},W)$.

We choose $h(\cdot)$ to be one-vs-all logistic regression function [26] as we find that for our dataset this method performs efficiently with high performance evaluation results. In Eq. (17), $\Omega(W)$ is the sparsity-inducing regularization term that enforces a number of zeros in the weight vector W. Therefore, the optimization objective jointly minimizes the prediction error by $\mathcal{L}(\cdot)$ and the number of features by $\Omega(W)$. The non-negative parameter λ is called the *regularization parameter* which balances the trade-off between these two terms. The larger the parameter λ, the more we emphasize on minimizing the number of features, and vice versa. In our problem, we propose to utilize the well-known ℓ_1-norm [25] as the sparsity inducing term; namely we let $\Omega(W)=\|W\|_1$, and thus our generic objective in Eq. (17) is instantiated as

$$\min_W\ \mathcal{L}(y,h(\mathbf{X},W))+\lambda\|W\|_1, \tag{18}$$

where $||W\|_1=\sum_i^k|W_i|$ is the summation of the absolute values of all the elements in W.

Minimizing the above equation will enforce some of the elements of W to be zeros, of which the corresponding features will thus be discarded. In this way, the target of feature selection is achieved.

2) Machine learning based on a pseudo-attack model: In general, the positive sets corresponding to the normal case can be obtained since the LUs are relatively easy to be controlled and predicted. However, because the attackers are unknown and unpredictable, it is difficult for AP or BS to accurately obtain the negative sets corresponding to the PCA case. According to the characteristics of the virtual channel as shown in section IV-A, we propose a pseudo-attack model to transform the problem of PCA detection into a binary classification problem.

The definition of the pseudo-attack model is presented by Definition 2, followed by Theorem 1 and Lemma 1.

Definition 3: (Pseudo-attack model): We define a pseudo-attacker model who has one more transmitter than that of the normal communication in NOMA communication in the process of sending the pilot sequence.

Lemma 1: Based on the positive sets from LUs, we can obtain the negative sets according to pseudo-attack model. Thus, we are able to establish a binary classifier based on the positive and negative sets, and this classifier can be used to distinguish the real PCA case from the normal case.

Proof 1: See Appendix.

Lemma 2: The trained classifier which is based on the pseudo-attack model is able to apply to the scenario of multiple attackers.

Proof 2: See Appendix.

Based on Lemma 1 and Lemma 2, we establish a binary classifier to detect PCA in NOMA. Let $TR_{XY} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_l, y_l)\}$ be a set of training vectors, $\mathbf{x}_i \in \mathbb{R}^n$, and the corresponding hidden states $y \in \Psi = \{1, 2\}$. The binary classifier uses the discriminant function $f : \chi \subseteq \mathbb{R}^n \to \mathbb{R}$ to train the corresponding machine learning model (MLM). If there is a set of testing vectors, $TE_X = \{\widehat{\mathbf{x}}_1, ..., \widehat{\mathbf{x}}_k\}$, based on the MLM, we can give the predictive value, i.e., $\{\widehat{y}_1, .., \widehat{y}_k, .., \widehat{y}_K\}$. Then the binary classification rule $q : \chi \to \Psi = \{1, 2\}$ can be defined as

$$q(\widehat{\mathbf{x}}_k) = \begin{cases} \widehat{y}_k = 1 & \text{for} \quad Normal \\ \widehat{y}_k = 2 & \text{for} \quad PCA \end{cases} \tag{19}$$

3) Optimization between detection accuracy and detection delay: As mentioned in the feature selection algorithm, the number of successive independent channel measurements is B_L. Based on the aforementioned binary classifier, the PCA detection can be completed as soon as B_L virtual channel data are collected. In other word, it has to wait B_L length virtual channels to be able to do a PCA detection. Apparently, the more channel information is obtained, the higher detection accuracy will be. Larger value of B_L results in a better performance for PCA detection. However, the benefit of an increasing B_L is gradually decreasing in the detection scheme. At the same time, a detection scheme with a low latency is desired. In addition, in many 5G applications, such as Unmanned Aerial Vehicle (UAV) communication, it is required to obtain the detection result as quickly as possible before the high-speed UAV leaves the current BS [27]. Therefore, we are interested to apply a technique that not only results in a high detection accuracy but also short detection delay.

To obtain the optimal B_L, we establish an optimization model, where the detection accuracy and detection delay are jointly considered. Let $\mathcal{H} = \{h_\gamma\}$ be a binary-class classifier to classify the incoming virtual channel $\tilde{\mathbf{x}}(B_L)$. A misclassification cost function is used to illustrate the performance of detection accuracy. Test misclassification cost function is defined as $C_1(\hat{y}, \tilde{y})$, where $\hat{y} = h_\gamma(\tilde{\mathbf{x}}(B_L))$ is the predicted class label, while the true class label of the incoming channel estimation is $\tilde{y}$. At the same time, let $C_2(\tilde{x}) \in \mathbb{N}^*$ be the time cost function which indicates the time cost value if PCA detection is postponed to collect B_L channel. In that case, we have the estimated total cost function $J(\,\cdot\,)$ as follows:

$$J(\tilde{\mathbf{x}}(B_L)) = C_1(h_\gamma(\tilde{\mathbf{x}}(B_L), \tilde{y})) + C_2(B_L). \tag{20}$$

We want to minimize the cost function $J(\,\cdot\,)$ such that we can find the optimal value for B_L for the detection accuracy and detection delay. Here, we exploit the loss function of the machine learning algorithm as the misclassification cost function and use the size of B_L as the detection time cost. Thus, Eq. (20) becomes

$$\min(J(\tilde{\mathbf{x}}(B_L))) = \min_{B_L} \mathcal{K}(y, h_\gamma(\tilde{x}, B_L)) + \lambda B_L, \tag{21}$$

where $\mathcal{K}(\,\cdot\,)$ is the empirical loss function between the labels y and the predictions $h_\gamma(\tilde{x}, B_L)$. We choose $h_\gamma(\,\cdot\,)$ to be one-vs-all logistic regression function [28], [29]. The nonnegative parameter λ is called the regularization parameter which balances these two terms. The smaller the parameter λ, the more we emphasize on high detection accuracy, and vice versa. It implies that we can fix the detection accuracy by adjusting the parameter λ. Then the corresponding optimal B_L can be given by Eq. (21).

The optimization objective in Eq. (21) jointly minimizes the misclassification error by $\mathcal{K}(\,\cdot\,)$ and the size of B_L. According to [28], [29], the term $\mathcal{K}(y, h_\gamma(\tilde{x}, B_L))$ is a decremental function under an increasing B_L. Hence, Eq. (21) is a quadratic function. As a result, we can use the derivative of Eq. (21) to directly solve the equation Eq. (21) to obtain the optimal B_L.

4) Complexity analysis: In this subsection, compared with the state-of-the-art scheme [16], we analyze the computational complexity of the proposed method. For the proposed PCA detection process, the core detection algorithm is algorithm 1: peak number estimation algorithm, in which the main calculation step is Manhattan distance, i.e., Equ. 16. According to the pseudocode of the algorithm 1, the computational complexity of the proposed detection scheme is $O(n^3)$. Uncoordinated frequency shift (UFS) detection algorithm [16] is the latest research about PCA detection recently. In this detection method, minimum description length (MDL) is the core calculation, and the computational complexity of the detection way can approximate $O(n\log(n))$. Thus, from the computational complexity of the algorithm, both of the detection schemes belong to polynomial algorithms.

However, on the other hand, the implementation of our proposed detection scheme is different from that of the existing detection method. For the UFS detection method, it needs the transmitter to train pilot symbols in advance. Instead, our proposed scheme only needs signal processing at BS or AP. There is no need to change the existing communication protocol, which is a very appealing feature for compatability.

VII. Simulation results

In this section, we provide numerical results for the proposed detection schemes. The performance of PCA detection under static environment and dynamic environment are illustrated. For the presented machine learning based detection framework for dynamic environment, peak estimation evaluation, feature selection evaluation, the performance of optimization for the detection accuracy and delay, comparison methods and multiple attackers evaluation are discussed in turn.

A. PCA detection under static environment

Without loss of generality, we set the physical AoD/AoA, i.e., $\phi_{t,l}$ and $\phi_{r,l}$ to be randomly generated as uniform distribution in $(0, 2\pi)$. Moreover, we mainly consider three key factors which may influence the performance of detection scheme: (1)

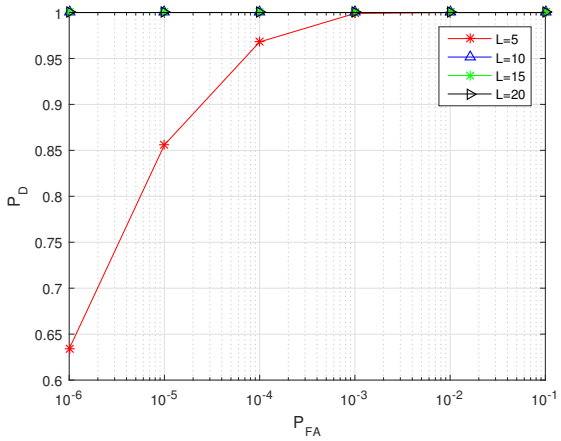


Fig. 8: Simulated false alarm rate P_{FA} vs. P_D with different multipath conditions $L = \{5, 10, 15, 20\}$. In this example, the signal to noise ratio is SNR=10dB; carrier frequency is 28GHz; the antenna number of transmitters and receivers are $N_t = N_r = 128$; the length of channel is $v = 50$, and the number of Monte Carto simulation is 10,000.

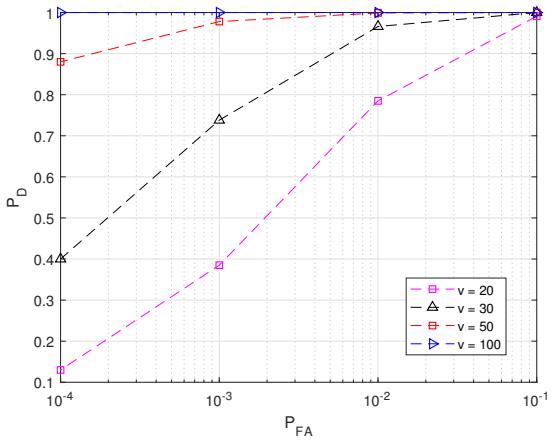


Fig. 9: Simulated false alarm rate P_{FA} vs. P_D with different lengths of channel $v = \{20, 30, 50, 100\}$. In this example, the signal to noise ratio is SNR=10dB; carrier frequency is 28GHz; the antenna number of transmitters and receivers are $N_t = N_r = 256$; the multipath condition is $L = 5$, and the number of Monte Carto simulation is 10,000.

The number of multipaths $L = \{5, 10, 15, 20\}$, (2) the length of the estimated virtual channel, i.e., $v = \{20, 30, 50, 100\}$ ($v \leq T$, refer to III-C), and (3) the signal to noise ratios $SNRs = \{-5, 0, 5, 10, 15, 20\}$ dB. In Fig. 8, the detection performance of the proposed scheme is presented by receiver operating characteristic curve (ROC), where P_D denotes the detection rate and P_{FA} is the false alarm rate. We can see that the proposed scheme shows a better performance in the rich multipath environment. For example, when the multipath number $L = 5$, the detection rate P_D reaches above 95% when the false alarm rate is 10^{-4}, and it can reach 100% as $P_{FA} = 10^{-3}$. While the detection performance is even better when the multipath number increases under the same conditions, such as $L = 10$, $L = 15$ and $L = 20$, where the detection rate is 100%.

Figure 9 shows the performance of ROC with different v. It is more obvious that the lengths of the collected virtual channel influences the performance of the proposed detection scheme. As the length of the virtual channel increases, the performance of detection improves. For instance for $v = 50$, the performance of the detection rate approximates to 90% for $P_{FA} = 10^{-4}$, while it reaches to 100% when $P_{FA} = 10^{-2}$. Under the same conditions, the detection accuracy of the proposed scheme can reach 100% for $v = 100$ even if $P_{FA} = 10^{-4}$.

Figure 10 illustrates the fluctuation of detection performance under different SNRs. It is obvious that higher SNRs lead to a better detection performance. Moreover, the detection performance presents an acceptable performance even for a lower SNR. For instance, when SNR=-5dB or SNR=0dB, the detection rate surpass 85% as the false alarm rate is 10^{-4}, and it approximates 100% when $P_{FA} = 10^{-2}$.

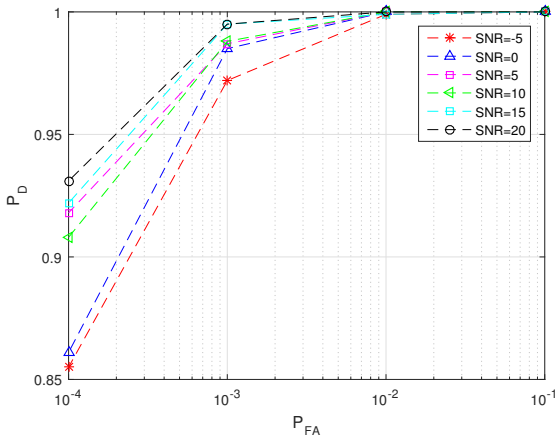


Fig. 10: Simulated false alarm rate P_{FA} vs. P_D with different SNRs. In this example, the carrier frequency is 28GHz; the antenna number of transmitter and receiver is $N_t = N_r = 256$; the multipath condition is $L = 5$; the length of channel is $v = 50$, and the number of Monte Carto simulation is 10,000.

B. PCA detection under dynamic environment.

1) Evaluate the peak estimation: In the proposed detection scheme, the number of peaks in the virtual channels is a fundamental and important detection parameter. The imprecise estimation of this parameter may impact the performance of PCA detection. Here, we consider the peak estimation accuracy which is the probability of correctly estimating the number of virtual channel peaks under different channel noises. Figure 11 (a) illustrates the fluctuation of estimation performance under different SNRs. It is obvious that a higher SNR leads to a better estimation accuracy. Even with the lower SNRs, the proposed peak estimation algorithm still achieves an impressive performance. For instance, when SNR = 0dB, the estimation accuracy surpasses 90%, and it approximates 100% as SNR exceeds 10dB. Figure. 11 (b) shows the PCA detection performance under the imprecise estimation of peak information. We can see that a lower SNR reduces the peak es-

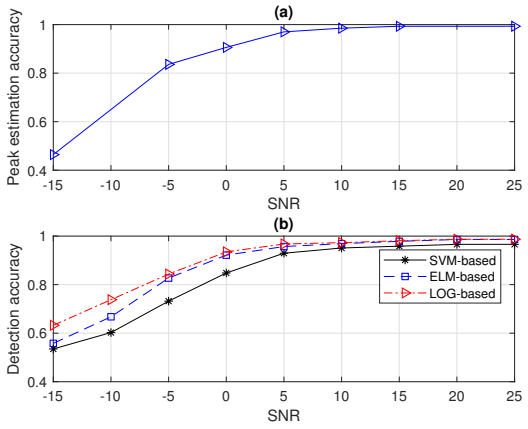


Fig. 11: Peak estimation accuracy vs. SNR.

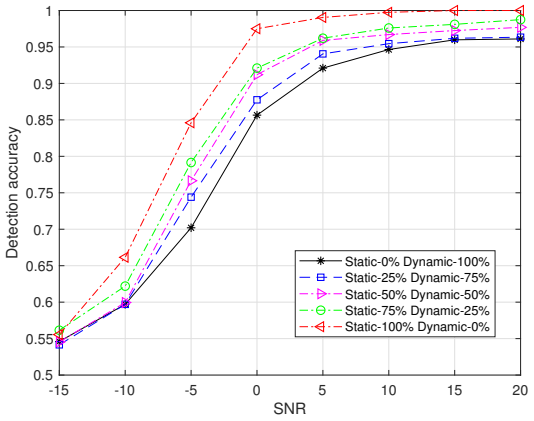


Fig. 12: The effectiveness of the machine learning based detection under intermediate scenarios.

timation accuracy. For example, when SNR = -10dB, the peak estimation accuracy is below 70% and the PCA detections are under 80%. The impact of the imprecise estimation on the detection accuracy will be weakened when SNR is higher than 5dB. For instance, the detection performance under the case of SNR=10dB is very close to that of SNR=15dB. It implies that the proposed scheme possesses a good anti-noise capability. In fact, it is not necessary to obtain a very high accuracy for the peak estimation. When the peak estimation accuracy is acceptable, e.g., higher than 90%, the impact of the imprecise peak estimation will be negligible to the detection accuracy.

2) Effectiveness under intermediate scenarios: Figure. 12 demonstrates the effectiveness of the machine learning based detection scheme under intermediate scenarios. Here, the considered data are comprised of the data from the static and dynamic scenarios, where the percentages of the data from the static scenario are $[0\%, 25\%, 50\%, 75\%, 100\%]$, and the rest of the data is from dynamic scenario. The number of successive independent channel measurement B_L is 15, and the size of the total training and testing data are 1,000. From Fig. 12, we can see that the performance of "Static-100%" is higher than other cases. Moreover, the higher the proportion of the data from static scenario, the better the detection performance.

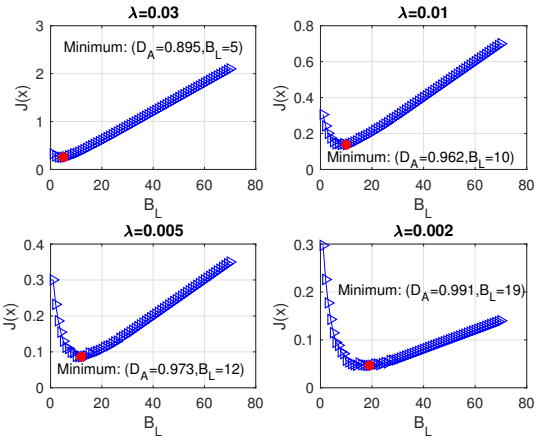


Fig. 13: The optimization of detection accuracy and detection delay under different regularization parameters.

The reason of this phenomenon lies in that the number of multipaths will approximate to a constant under the static scenario. Thus, the data from static scenario is more stable than that from dynamic scenario, and a stable environment is more conducive to distinguish the difference between the PCA case and the normal case. Furthermore, these results imply that the proposed machine learning-based detection under dynamic scenarios can tackle the intermediate cases where static and dynamic mix together as long as we can obtain enough training samples.

3) Optimization of detection accuracy and delay: Figure 13 illustrates the performance of the total cost function $J(\tilde{\mathbf{x}}(B_L))$ with respect to different λ, where red point denotes the minimum value. λ reflects the different consideration between detection accuracy and detection delay. When λ is larger, such as $\lambda = 0.03$, the system weighs on smaller detection delay. Hence, $B_L = 5$ is selected and the detection accuracy of the classifier is 89.5%. As the λ increases, the system gradually switches into working towards a higher detection accuracy. For instance, when $\lambda = 0.002$, the detection accuracy reaches 99.1%, at the same time $B_L = 19$. That means the system needs to collect more channel estimation information to complete a detection in return for higher detection accuracy of the classifier.

4) Feature selection performance evaluation: In this subsection, we evaluate the feature selection methodology introduced in Section 14. The one-vs-other logistic regression hypothesis is trained according to the optimization problem in Eq. (18) on training set with various sizes (from 1,000 to 8,000) and $B_L = 10$. The regularization parameter λ is tuned through 10-fold cross-validation to avoid model overfitting/underfitting, and achieve high recognition accuracy and minimum number of selected features. Fig. 14(a) shows the selected features (in yellow) from all the available features, versus the training set size. The features shown in blue are the discarded features ($W_i = 0$). According to Fig. 14(a), feature selection method selects the more valuable features and reduces the number of the features on average in the range of 84.6%, i.e., from 13 to 2. This will result in reduction of training and prediction running time. Fig. 14(b) and 14(c)

(a) Selected features with various training sizes. (b) Test accuracy with selected features. (c) Test accuracy without selected features

Fig. 14: Experiment results of features selection.

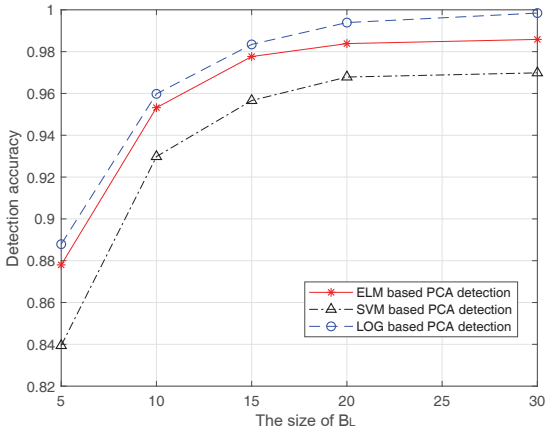


Fig. 15: Detection accuracy vs. the size of B_L under different machine learning algorithm. In this example, the training data size and the testing data size are 1,000; the carrier frequency is 28GHz; the antenna number of transmitter and receiver is $N_t = N_r = 256$ and SNR=10dB.

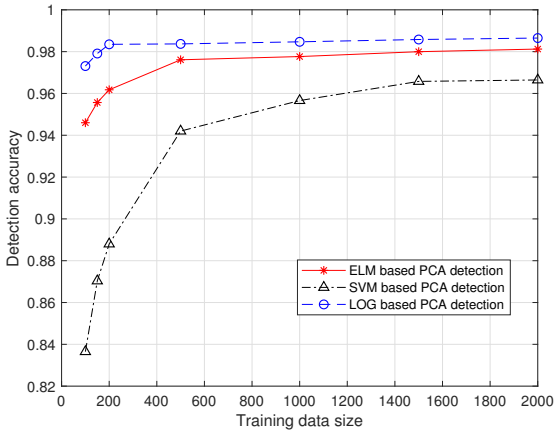


Fig. 16: Detection accuracy vs. the size of training data under different machine learning algorithm. In this example, the size of B_L is 15; testing data size is the same as the corresponding training data size; the carrier frequency is 28GHz; the antenna number of transmitter and receiver is $N_t = N_r = 256$ and SNR=10dB.

show the PCA detection accuracy on the test dataset with and without selected features, respectively. We can see that in both cases high precision of correct PCA detection is achieved (higher than 95%). Therefore, we conclude that the introduced feature selection method reduces model training and prediction overhead without sacrificing the performance.

5) Analysis of different machine learning algorithms: In this subsection, we consider the performance of PCA detection scheme under different machine learning algorithms. It is worth noting that we do not focus on how to optimize these machine learning algorithms. We are more interested in which machine learning algorithm is more suitable for the proposed PCA detection scheme. Herein, we consider three popular machine learning algorithms, including support vector machine (SVM), extreme learning machine (ELM) and logistics regression (LOG). Moreover, for consistency, we consider the minimum Bayes risk as a performance criterion, given by

$$P_e^{(min)} = \min(P_{MD} + P_{FA}), \tag{22}$$

where P_{MD} denotes the miss detection rate and P_{FA} is the false alarm rate. To simplify the analysis, we denote detection accuracy as $1 - P_e^{(min)}$ for various detections, and the selected features are Euclidean distance (f_1) and Root mean square (f_{10}) according to the features selection algorithm.

Figure 15 illustrates the influence of the size of B_L on the detection performance under different machine learning algorithms with the same training and testing data size (1,000). As the size of B_L increases, the detection accuracy is improved. However, when the size of B_L exceeds 20, the trend of rising curves is not significant. In other words, the size of B_L is the restriction on the detection performance when the used B_L is very small, such as less than 10. Furthermore, in Fig. 15 we can see that LOG based PCA detection scheme reaches a higher detection accuracy than that of other machine learning algorithms. For example, when the size of B_L is 15, the detection accuracy of LOG based PCA detection is higher than 98%, while for other schemes, it is less than 98% under the same conditions.

Figure 16 shows the detection performance under different training data with the same size of B_L. We can see that all of the curves are slowly rising when the training length is

TABLE IV: Training time under different sizes of B_L and data set size.

LOG ELM SVM Training time (s)		Training data set size								
		100			200			500		
The size of B_L	5	0.0023	0.0078	0.0169	0.0033	0.0101	0.0125	0.0059	0.0218	0.0242
	10	0.0094	0.0125	0.0213	0.0156	0.017	0.0216	0.0504	0.0625	0.0625
	15	0.0142	0.0313	0.0313	0.0504	0.0625	0.0625	0.109	0.1719	0.0938
	20	0.0156	0.0659	0.0334	0.0625	0.0859	0.065	0.1894	0.2438	0.6844
	30	0.125	0.0725	0.0491	0.1875	0.0885	0.1016	0.2425	0.3266	0.9797
LOG ELM SVM Training time (s)		Training data set size								
		1000			1500			2000		
The size of B_L	5	0.0114	0.2203	1.196	0.0150	0.4766	20.70	0.1406	0.7422	33.3
	10	0.0469	0.3063	1.375	0.1806	0.6250	23.54	0.2344	0.75	43.54
	15	0.1894	0.3438	1.98	0.314	0.6459	31.90	0.3344	0.8059	48.05
	20	0.1938	0.3906	2.117	0.346	0.6884	45.15	0.4175	0.8359	59.01
	30	0.2969	0.4141	9.968	0.3594	0.7094	54.69	0.4312	0.9047	90.8

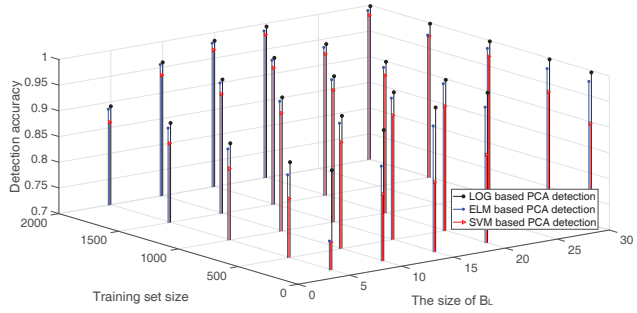


Fig. 17: Detection accuracy under different training length with the different sizes of B_L.

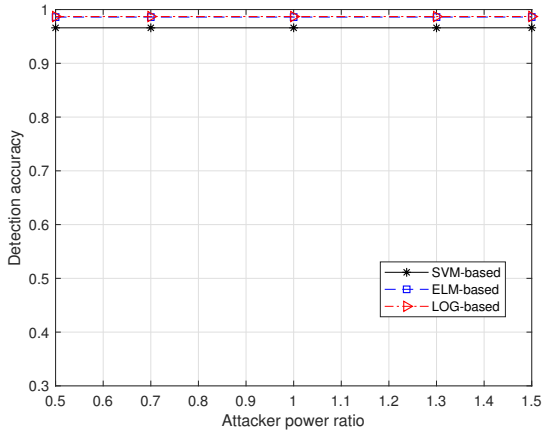


Fig. 18: PCA detection under different powers of attackers.

above 500. While, when the training data is less than 500, the lack of training data becomes the bottleneck for the detection performance. When the training data size is larger than 1,000, the detection performance of ELM based PCA detection is close to that of LOG based PCA detection scheme, and the performance of SVM based PCA detection is at the bottom. We also notice that when the training data length is 1,000, the detection performance is close to that of the length 2,000. It implies that the proposed scheme doesn't need massive training data.

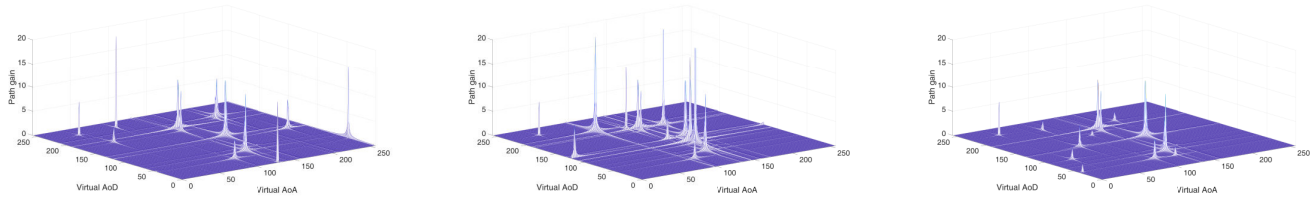
Figure 17 shows the detection performance under a larger scale, and Table IV shows the training time under different B_L and data set size. On one hand, we can see that the bigger the size of B_L and larger training data can reach a higher detection accuracy. On the other hand, the larger size of the training data and B_L will require more training time. Furthermore, under the same conditions i.e., the same training data and B_L, the detection accuracy of LOG based PCA detection always is higher than that of the other two schemes. Meanwhile, we can see that LOG based PCA detection requires a shorter time to train model than that of SVM and ELM based schemes under the same conditions. Therefore, LOG algorithm is more suitable for the proposed PCA detection framework than SVM and ELM.

6) PCA detection evaluation under various powers of attackers: In this subsection, the proposed scheme under different transmit powers of the attacker is evaluated. For consistency, we consider an attacker-power ratio as a criterion to illustrate the transmitting power of the attacker, given by

$$R_{AP} = \frac{P_{attacker}}{P_{user}} \tag{23}$$

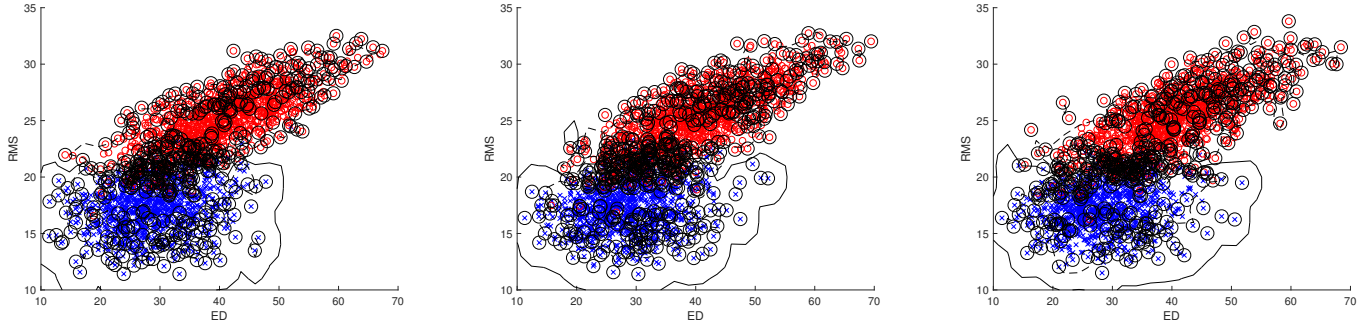
where R_{AP} indicates the attacker-power ratio, $P_{attacker}$ denotes the power of attacker, and P_{user} denotes the power of the legitimate user.

Figure 18 shows the performance of virtual channels under different R_{AP}. We can see that the detection accuracy curves representing different machine learning algorithms are relatively stable under different attacker-power ratios. In other word, simulation results show that the attacker's power does not affect the detection performance of the proposed PCA



(a) Attacker with the same power as the legitimate user. (b) Attacker with 1.5 times power as the legitimate user. (c) Attacker with 1.5 times power as the legitimate user.

Fig. 19: The performance of virtual channel under different powers of attackers.



(a) Only one attacker. (b) Two attackers. (c) Three attackers.

Fig. 20: Characteristic space under different attackers.

TABLE V: Detection performance under multi-attackers.

Scenarios	**Training dataset** (1,000)	**Testing dataset** (1,000)	**Detection accuracy**(%) (SVM, ELM, LOG)
One attacker	Two users and one attacker	Two users and one attacker	(92.1, 95.65, 96.1)
Two attackers	Two users and one attacker	Two users and two attackers	(96.15, 97.85, 98.21)
Three attackers	Two users and one attacker	Two users and three attackers	(96.85, 97.94, 98.34)

detection. The performance of virtual channel under different attacker-power ratios is presented in Fig. 19. Figure. 19(a) shows the scenario based on which the transmitting power of the attacker is the same as the legitimate user's power. Figure. 19(b) and Fig. 19(c) display the phenomenon for $R_{AP} = 1.5$ and $R_{AP} = 0.5$, respectively. We can see that the number of paths with a high gain in 19(b) is more than that of Fig. 19(a). That means that the increase in the attacker's power is manifested by an increase in the number of the paths with a high gain. An inverse phenomenon appears in Fig. 19(c), where the attacker is with 0.5 times transmitting power of the legitimate user, and the number of the paths with a small gain increases. However, the number of peaks in virtual channels doesn't change, i.e., the sparsity of the virtual channel is not affected by the attacker's power level. Considering that, for the proposed detection scheme, the peak number and sparsity of virtual channel are the key metrics. The transmitting power of the attacker does not impact the detection performance of the proposed method.

7) PCA detection evaluation under multiple attackers: Fig. 20 shows the characteristic space under one, two and three attackers, respectively. We note that there is no significant difference between Fig. 20(a), Fig. 20(b) and Fig. 20(c). It indicates that although the training data is based on one attacker scenario, the characteristic space for more attackers still has a good performance. Table V gives a more detailed description of the detection performance under multiple attackers. The detection performance can be improved in the multiple attacker scenarios. For instance, detection accuracy is 92.1% under SVM based PCA detection when there is only one attacker, while it reaches 96.8% as the number of attackers increases to three under the same conditions. The sparsity of virtual channel helps explain why the detection accuracy under multiple attackers is better than that of one attacker. For the virtual channel under multiple attackers, the number of peaks increases due to the increasing of the number of transmitters, as shown in Fig. 3. This will result in a further reduction in sparsity of virtual channel compared with the case with a single attacker. Thus, the differentiability between PCA case and the case without attackers will be improved in the feature space. Therefore, the PCA detection performance in presence of multiple attackers results in better performance compared to the single attacker case.

VIII. Conclusion and discussion

In this paper, we studied PCA detection in NOMA. The virtual channel under mmWave and massive MIMO was exploited to detect the PCA. Based on the observations of the characteristics of the virtual channel, we proposed a binary hypothesis test and machine learning based detection framework to achieve the PCA detection for static and dynamic applications, respectively. Simulation results demonstrated the effectiveness of the proposed schemes.

The virtual channel model is a promising signal processing technology for the communications using massive MIMO and mmWave [30]. It can be used for channel estimation and the design of beamforming for 5G communication [31]. However, due to the absence of equipment with Massive MIMO and mmWave, there is no real-world dataset used in this paper. Even so, it is obvious that the virtual channel model can provide some new characteristics which can be used to achieve physical layer security for 5G communication. It is worth to further investigate, and that motives our future works.

Appendix

Proof of Lemma 1: Let $\mathbf{p}_{E,i}$ denote the number of successive independent channel measurement corresponding to the pseudo-attack model, while $\mathbf{p}_{A,i}$ denotes the channel sampling of virtual channel corresponding to the normal communication in NOMA. Let a be the number of LUs in the normal NOMA communication. First, the value of $\mathbf{F}_S(\mathbf{p}_{E,i})$ can be determined by $\mathbf{p}_{A,i}$. According to Definition 2, $\mathbf{p}_{E,i}$ has one more transmitter than that of $\mathbf{p}_{A,i}$. Since we know the number of transmitters for $\mathbf{p}_{A,i}$, it is easy to obtain the data for the case of having one more transmitter for $\mathbf{p}_{E,i}$. Let $\mathbf{F_S}$ be the feature space which contains all of the selected features. We then consider the difference between $\mathbf{F}_S(\mathbf{p}_{A,i})$ and $\mathbf{F}_S(\mathbf{p}_{E,i})$. According to the feature selection algorithm, we can select the best appropriate features. For instance, the Euclid distance is selected (In fact, it is one of the best features among the statistical properties according to simulation.). Let $\mathbf{p}_a \subseteq \mathbf{p}_A$, $\mathbf{p}_b \subseteq \mathbf{p}_A$ and $\mathbf{p}_e \subseteq \mathbf{p}_E$. The probability distribution of p_a and p_b can be approximated by two independent normal distributions [22] [23], i.e., p_a and $p_b \sim N_t(\mu_1, {\sigma_1}^2)$. Moreover, since there is a large number of transmitters, PCA scenario possesses more paths than that of normal communication. Thus, $p_e \sim N_t(\mu_2, {\sigma_2}^2)$ For the normal scenario, we have $Z_i = p_{a,i} - p_{b,i}$ where Z_i follows a bivariate random normal distribution, i.e., $\mathbf{Z} \sim N_t(0, 2\sigma^2)$. Thus, the statistic, (i.e., Euclidean distance)

$$Q_E^{(P)} = \sqrt{\sum_{i=1}^{t} \left(\frac{Z_i}{2\sigma_i}\right)^2}, \tag{24}$$

is distributed according to the chi distribution with t degrees of freedom, and the probability density function is

$$f(Z_E^{(P)}; t) = \frac{2^{1-\frac{t}{2}} Z^{t-1} e^{-\frac{Z^2}{2}}}{\Gamma(\frac{t}{2})}, \tag{25}$$

where $\Gamma(*)$ is the Gamma function. Then, the expectation and variance value of Q_B can be calculated accordingly, i.e.,

$$\begin{aligned} E(Q_B) &= \sqrt{2}\frac{\Gamma((t+1)/2)}{\Gamma(t/2)}, \\ \mathrm{var}(Q_E^{(P)}) &= t - E(Q_B)^2. \end{aligned} \tag{26}$$

In contrast, for PCA scenario, if there is only one attacker, the real PCA is the same as the pseudo-attack model. Thus, $Z_i = p_{a,i} - p_{e,i}$ follows a bivariate random normal distribution, i.e., $Z \sim N_t(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Hence, the expectation value of the Euclid distance between $\mathbf{p}_E$ and $\mathbf{p}_A$ can be written as

$$E(Q_E) = \sqrt{\frac{\pi}{2}} L_{1/2}^{(1/t-1)}\left(\frac{-\lambda}{2}\right), \tag{27}$$

where $\lambda = \sum_{i=1}^{t} \left(\frac{\mu_{1i} - \mu_{2i}}{\sigma_{1i} + \sigma_{2i}}\right)^2$. Only if $\mu_1 - \mu_2 = 0$ and $\sigma_1 = \sigma_2$, i.e., when a non-central chi distribution becomes a central chi distribution, we have $E(Q_B) = E(Q_E)$. Otherwise, the expectation value of the non-central chi distribution is greater than that of central chi distribution, i.e., $E(Q_E) > E(Q_B)$.

The analysis of other selected features are equivalent to that of Euclidean distance, hence it is not described in this article. As a result, $\mathbf{F}_S(\mathbf{p}_{A,i})$ and $\mathbf{F}_S(\mathbf{p}_{E,i})$ can be distinguished so long as there is one more transmitter in PCA scenario than that of normal communication. Therefore, the established classifier based on pseudo-attack model can distinguish the PCA case with one attacker from normal case. Thus the proof is completed.

Proof of Lemma 2: When there are multiple attackers, the probability distribution of p_e will still be able to approximate a normal distribution, denoted as $N_t(\mu_3, {\sigma_3}^2)$. Thus, similar to Proof of Lemma1, $Z_i = p_{a,i} - p_{e,i}$ follows a bivariate random normal distribution, i.e., $Z \sim N_t(\mu_1 - \mu_3, \sigma_1^2 + \sigma_3^2)$. Hence, the expectation value of the Euclidean distance between p_a and p_e can be obtained, denoted as $Q_E^{(M)}$. As a result, we have $E(Q_E^{(M)}) \geq E(Q_E) \geq E(Q_A)$, and the equality holds, if and only if, $\mu_1 = \mu_2 = \mu_3$, and $\sigma_1 = \sigma_2 = \sigma_3$. Therefore, the features of multiple attackers in $\mathbf{F}_S$ is closer pseudo-attack model than that of normal communication in NOMA. It implies that the established classifier based on pseudo-attack model can distinguish the PCA case with multiple attackers from normal case. Thus, Lemma 2 is confirmed.

References

[1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.

[2] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell tdd systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, 2011.

[3] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive mimoł5g system," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 905–923, 2016.

[4] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive mimo: An overview on passive eavesdropping and active attacks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 21–27, 2015.

[5] D. Kapetanovic, G. Zheng, K.-K. Wong, and B. Ottersten, "Detection of pilot contamination attack using random training and massive mimo," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*. IEEE, 2013, pp. 13–18.

[6] D. Kapetanovic, A. Al-Nahari, A. Stojanovic, and F. Rusek, "Detection of active eavesdroppers in massive mimo," in *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on*. IEEE, 2014, pp. 585–589.
[7] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5g nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.
[8] J. Ma, C. Liang, C. Xu, and L. Ping, "On orthogonal and superimposed pilot schemes in massive mimo noma systems," *IEEE Journal on Selected Areas in Communications*, 2017.
[9] Y. Chen, J. Schaepperle, and T. Wild, "Comparing idma and noma with superimposed pilots based channel estimation in uplink," in *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on*. IEEE, Conference Proceedings, pp. 89–94.
[10] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 62, no. 12, p. 3111, 2014.
[11] C. Wang and H.-M. Wang, "Physical layer security in millimeter wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5569–5585, 2016.
[12] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.
[13] W. Ning, J. Long, and Z. Kai, "Pilot contamination attack detection for noma in mm-wave and massive mimo 5g communication," in *Proc. of the IEEE CNS Conference*. IEEE, 2018, pp. 1–9.
[14] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5g nonorthogonal multiple-access downlink transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6010–6023, 2016.
[15] D. Kapetanovic, A. Al-Nahari, A. Stojanovic, and F. Rusek, "Detection of active eavesdroppers in massive mimo," *2014 Ieee 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (Pimrc)*, pp. 585–589, 2014. [Online]. Available: ¡Go to ISI¿://WOS:000392729300113
[16] W. Zhang, H. Lin, and R. Zhang, "Detection of pilot contamination attack based on uncoordinated frequency shifts," *IEEE Transactions on Communications*, 2018.
[17] T. Kim and D. J. Love, "Virtual aoa and aod estimation for sparse millimeter wave mimo channels," *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 146–150, 2015. [Online]. Available: ¡Go to ISI¿:WOS:000380547100030
[18] Y. Wang, Z. Tian, S. Feng, and P. Zhang, "A fast channel estimation approach for millimeter-wave massive mimo systems," in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*. IEEE, 2016, pp. 1413–1417.
[19] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
[20] W. Zhang, H. Lin, and R. Zhang, "Detection of pilot contamination attack based on uncoordinated frequency shifts," *IEEE Transactions on Communications*, 2018.
[21] W. Zhang, F. Gao, S. Jin, and H. Lin, "Frequency synchronization for uplink massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 235–249, 2017.
[22] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2014.
[23] M. K. Samimi and T. S. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels," pp. 3483–3489, 2014.
[24] L. Zhao, A. Alipour-Fanid, M. Slawski, and K. Zeng, "Prediction-time efficient classification using feature computational dependencies," 2018.
[25] F. Bach, R. Jenatton, J. Mairal, G. Obozinski *et al.*, "Optimization with sparsity-inducing penalties," *Foundations and Trends® in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.
[26] C. Robert, "Machine learning, a probabilistic perspective," 2014.
[27] A. Alipour-Fanid, M. Dabaghchian, N. Wang, P. Wang, L. Zhao, and K. Zeng, "Machine learning-based delay-aware uav detection over encrypted wi-fi traffic," 2019.
[28] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: http://www-stat.stanford.edu/ tibs/ElemStatLearn/
[29] K. P. Murphy, *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press, 2013.
[30] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.
[31] A. Sayeed and J. Brady, "Beamspace mimo for high-dimensional multiuser communication at millimeter-wave frequencies," in *Global Communications Conference (GLOBECOM), 2013 IEEE*. IEEE, Conference Proceedings, pp. 3679–3684.