Pilot Contamination Attack Detection for 5G MmWave Grant-Free IoT Networks

Ning Wang, Weiwei Li, Amir Alipour-Fanid, Long Jiao, Monireh Dabaghchian and Kai Zeng

***Abstract*—Grant-free random access is an emerging technology for providing massive connectivity for 5G massive machine-type communications (mMTC), where non-orthogonal pilot sequences are used to simultaneously detect active users and estimate channels. However, grant-free 5G IoT networks are vulnerable to pilot contamination attacks (PCA), where the attacker can send the same pilots as legitimate IoT users to harm the active user detection and channel estimation. To defend against this attack, in this paper, we propose a physical-layer countermeasure based on the channel virtual representation (CVR). CVR can emphasize the unique characteristics of mmWave channels that are sensitive to the location of the sender. This can be utilized to counter PCA no matter if the attacker's pilots are superimposed to that of the victim or not. Based on this observation, to achieve an efficient PCA detection, a single-hidden-layer multiple measurement (SHMM) Siamese network is employed. This solution tackles the challenges of channel randomness and massive connectivity in mMTC IoT networks, and supports small sample learning. Simulation results evaluate and confirm the effectiveness of the proposed detection scheme under various scenarios. The detection accuracy can approach 99% with 128 antennas at the receiver and reach above 95% even with only 50 training samples.**

***Index Terms*—Pilot contamination attacks, Channel virtual representation, Grant-free multiple access, 5G IoT networks**

I. Introduction

It is envisioned that the fifth-generation (5G) cellular communication systems will provide massive connectivity for machine-type communications (mMTC) [1], [2]. MMTC is able to support the diverse Internet of Things (IoT) applications such as environmental sensing, smart manufacturing and smart farming. Compared with conventional human-centric communication, mMTC IoT communications have two distinctive characteristics: (i) there are a massive number of machine-type connections; (ii) only a small and random fraction of the devices are active at any given instant. To accommodate these communication characteristics, grant-free random access schemes are proposed as promising solutions to enable massive connectivity with low-latency in mMTC 5G IoT networks [3]. In grant-free schemes, each active device directly transmits its identity and data to the based station (BS) without waiting for any permission, and the non-orthogonal pilot sequences are used to identify active users and estimate the corresponding channels [4], [5].

However, despite its great potential, the grant-free random access process is very vulnerable to pilot contamination attacks (PCA), where the attacker can send the same pilot signals as legitimate users (LUs) [6]. Since pilot sequences are preassigned and fixed for all the LUs in IoT networks, PCA attackers can easily obtain the pilot by eavesdropping and masquerade as a LU by using the LU's pilot sequences. In this case, both normal active user detection and channel estimation in grant-free mMTC IoT networks will be harmed. If there are no countermeasures, the PCA attacker could gain privileges to further perform advanced attacks, e.g., man-in-the-middle attacks and denial-of-service attacks.

Unfortunately, existing countermeasures to PCA cannot be directly applied to the grant-free mMTC IoT networks. In general, there are two detection schemes against PCA: pilot design-based methods and physical layer feature-based schemes. For pilot design-based methods, legitimate users transmit a random or special pilot sequence to receivers who will raise alarms when multiple transmitters corresponding to the same pilot sequence are detected during one orthogonal resource block [7], [8]. However, in the grant-free IoT network, since the pilot is also used for user identification, it is usually fixed for each user and cannot be randomized for each transmission. Physical layer feature-based schemes can achieve PCA detection by exploiting physical layer features or mechanisms, such as the sparsity of massive MIMO mmWave channels [9], beamforming in massive MIMO [10], and uncoordinated frequency shift [11]. Nevertheless, in these detection schemes, LUs need to apply a complex multi-antenna mechanism or send multiple random frequency shifts. These mechanisms are not affordable or desirable for low-cost IoT devices which cannot support a massive number of antennas. Therefore, a PCA detection scheme that does not require any extra hardware or signal processing overhead at the IoT devices in grant-free access networks is desirable but missing in the literature.

To fill this gap, in this work, we propose a machine learning-based PCA detection scheme based on the characteristics of channel virtual representation (CVR). CVR is designed to emphasize the characteristics of mmWave communication channels, and mmWave communications have been regarded as a key wireless communication technology in 5G IoT networks. Thanks to the high directivity of mmWave communications, the multipath channel characteristics due to reflection and refraction are sensitive to the location of the sender. In CVR, the physical angle-of-arrivals (AoAs) of

Ning Wang, Weiwei Li, Amir Alipour-Fanid, Long Jiao, and Kai Zeng were with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA, US 22030 e-mail: ({nwang5, wli25, aalipour, ljiao, kzeng2}@gmu.edu).

Weiwei Li is also with school of information and electrical engineering, Hebei University of Engineering, Handan, Hebei, China 056001.

Monireh Dabaghchian is with department of computer science, Morgan State University, Baltimore, MD, US 21251, e-mail: (monireh.dabaghchian@morgan.edu).

these multi-paths can be mapped into the index of nonzero elements in the virtual AoAs defined by discrete Fourier transform (DFT). Thus, senders at different locations will result in different virtual AoAs in CVR. In grant-free IoT networks, there could be a random set of users transmitting concurrently at each time, and each LU is corresponding to a fixed pilot sequence. When there is a PCA attack, it is inevitable that more senders appear in the channel corresponding to a pilot sequence or the location of the transmitter is different. This will result in a significant change in the observed CVR corresponding to the pilot sequence. Motivated by this observation, we present a single-hidden-layer multiple measurement (SHMM) Siamese network to achieve PCA detection based on the similarity of CVR. Simulation results show that our proposed detection scheme can reach a high detection accuracy. The detection accuracy can approach 99% with 128 antennas at the receiver and reach above 95% even when there are only 50 training samples.

The contributions of this work are threefold:

- To the best of our knowledge, this is the first work to investigate the PCA problem in grant-free 5G IoT networks. Based on the observation that the virtual AoAs are sensitive to the location of the sender, the similarity of CVR is exploited to detect PCA.
- For the characteristics of the CVR in grant-free 5G IoT networks, we present a tailored SHMM-Siamese network to achieve PCA detection.
- We provide theoretical analysis and simulation results under various scenarios to verify the feasibility and effectiveness of the proposed detection scheme.

The proposed SHMM-Siamese network-based detection scheme has the following salient features:

- *No change or any extra computation/communication requirements on IoT devices.* The proposed detection scheme and the corresponding signal processing are conducted at BS/AP. It does not need any modification or computation/communication requirements on IoT devices, which makes the proposed scheme suitable for low-cost IoT applications.
- *Applicable to the scenario of 5G massive IoT devices in mMTC IoT networks.* The unique characteristics of mmWave communications are employed in the proposed detection model, and the presented Siamese network can address the problem of model training under massive IoT devices. Therefore, the proposed scheme is applicable to 5G mMTC IoT networks.
- *Efficient detection model training.* The input training samples for the proposed detection model are sample pairs which consist of two CVR samples, and the corresponding training output is only either the similar or dissimilar label. Hence, it does not need to label every device in the networks when training the detection model.

In the remainder of this article, we first introduce the related work in Section II; then the system model and motivation are given in Section III. Theoretical analysis and challenges are provided in Section IV. Section V presents the SHMM-Siamese network-based detection scheme. Section VI provides the simulation results, and finally Section VII concludes this paper.

Notations: Superscript $(\cdot)^T$ denotes transpose, superscript $(\cdot)^*$ indicates conjugate transpose, $\mathbf{I}$ denotes identity matrix with appropriate dimensions, and $E[\cdot]$ represents expectation operation, and $\|\cdot\|_2$ denotes the ℓ_2 norm.

II. Related work

PCA was first proposed in [12]. Later, an overview of PCA for massive MIMO communication was given in [10], [13]. In general, existing PCA detection strategies can be categorized into pilot design-based methods and physical layer feature-based schemes.

Pilot design-based methods try to counter PCA by using random or special pilot sequences. Random phase-shift keying (PSK) pilot scheme was first proposed in [7], and an improved variation which considers the case of three or more observations was provided in [14]. The other random pilot scheme is based on random frequency shifts [11]. Unfortunately, these PCA detection schemes based on random pilots can hardly be applied to grant-free 5G IoT networks, where fixed pilots must be preassigned to the devices to identify different IoT devices. The special design of the pilot sequence also can mitigate pilot contamination and detect PCA. For example, the authors in [15] proposed superimposing a random sequence on the pilot sequence and use source enumeration methods to detect PCA. A two-way training scheme via a whitening-rotation based semi-blind approach was proposed to design the pilot sequence to thwart PCA [16]. However, these pilot design methods are designed for the orthogonal multiple access technology and need to modify current communication protocol, which make these schemes unsuitable for off-the-shelf IoT devices and non-orthogonal pilot sequences used by the users in grant-free access networks.

For physical layer feature-based schemes, a special beamformer has been designed to detect the activity of PCA based on massive MIMO [10]. Moreover, a PCA detection scheme based on CVR in NOMA communications with mmWave and massive MIMO technologies was proposed [9]. These physical layer feature-based schemes assume that the LUs are equipped with massive antennas and can achieve some features of massive MIMO, such as beamforming and channel sparsity. However, the LUs in the grant-free IoT networks might be low-cost IoT devices with a single antenna, which would not have the capability to carry out the function of massive MIMO. In addition, there are two significant differences between this work and [9]: (i) [9] focuses on attack mode one, while this work can cover both of the two attack modes (These attack modes will be introduced in Section III-A2.); (ii) The detection strategy and physical features are different. A tailored Siamese network and the similarity of CVRs are exploited to achieve PCA detection in this work, whereas a simple machine learning algorithm and the sparsity of the massive MIMO channel are used in [9].

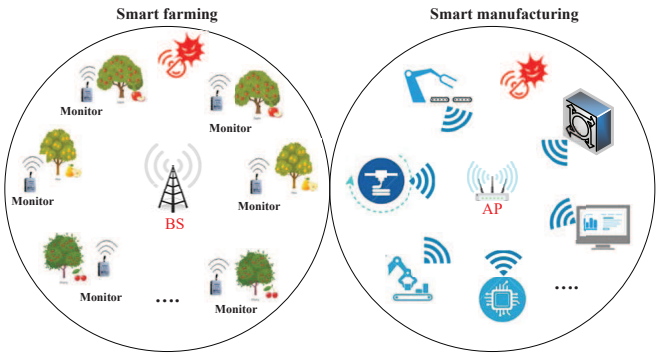


Fig. 1: MMTC IoT application scenarios: Smart farming and smart manufacturing.

III. System model and motivation

This section will introduce the system model and motivation of this study.

A. System model

1) Communication system: Consider a typical mMTC IoT network, e.g., smart farming and smart manufacturing illustrated in Fig. 1, where the BS/AP can serve a large number of low-cost IoT devices. The locations of all LUs are fixed in the network, such as the monitors in smart farming and the actuators in smart manufacturing. The BS/AP can have multiple antennas and sufficient computing resources, and the IoT devices are resource-constrained and limited computation power. Moreover, all transceivers in this network can support mmWave communications.

To enable massive connectivity, the grant-free random access scheme is exploited to connect BS/AP and LUs. Every user is preassigned with a unique pilot sequence as the metadata, which is used to estimate the channel and serves as the identification (ID) during all of the time slots [3]. During the communication phase, the active users transmit metadata and content data to the BS/AP. The metadata involves preamble for active user detection and channel estimation, and the content data can be directly sent after the metadata free from waiting for the grant from BS/AP. At each time slot, based on the metadata, the BS/AP will first detect the active users and estimate the corresponding channels, then decode the content data with the estimated channels [17]. To tackle the challenge that orthogonal pilot sequences are failed to be assigned to all of the users in the mMTC IoT network, the approximate message passing (AMP) algorithm based on compressed sensing techniques are used to achieve the active user detection and channel estimation [3], [17].

Formally, we assume there are K users in the system, which are represented by the set $\mathbf{K} = \{1, ..., k, ..., K\}$ and the BS/AP is equipped with N_r antennas. In this case, the channel from user k to the BS/AP is $\mathbf{H}_k \in \mathbb{C}^{N_r \times 1}$. In each coherent time slot, the user activity indicator function can be represented as

$$\tau_k = \begin{cases} 1 & User \text{ k is active;} \\ 0 & Otherwise. \end{cases} \tag{1}$$

Furthermore, each user k is assigned with one pilot sequence $\mathbf{p}_T \in \mathbb{C}^{C \times 1}$ with $\|\mathbf{p}_T\|^2 = 1$, where C indicates the length of the pilot sequence during a window of discrete-time length T. It is worth noting that these pilot sequences are non-orthogonal to each other. Thus, the signal at the BS/AP can be represented by

$$\mathbf{Q}_R = \mathbf{X}_P \mathbf{P}_T + \mathbf{E}_P, \tag{2}$$

where $\mathbf{Q}_R = [\mathbf{q}_{R,1}, ..., \mathbf{q}_{R,C}] \in \mathbb{C}^{N_r \times C}$ is the matrix of received signals, $\mathbf{P}_T = [\mathbf{p}_{T,1}, ..., \mathbf{p}_{T,K}]^T \in \mathbb{C}^{K \times C}$ is the collection of pilot sequences of all users, $\mathbf{X}_P = [\mathbf{x}_{P,1}, ..., \mathbf{x}_{P,K}] \in \mathbb{C}^{N_r \times K}$ with $\mathbf{x}_{P,k} = \tau_k \mathbf{H}_k$ denoting the effective channel of user k, and $\mathbf{E}_P$ is the independent additive white Gaussian noise.

Based on the received signal $\mathbf{Q}_R$, the goal for the BS/AP is to detect the active users and estimate the corresponding channel. In grant-free networks, the pilot sequences $\mathbf{P}_T$ can be seen as the measurement matrix (or called sensing matrix), the entries in the measurement matrix suggests the active users, and recovering $\mathbf{X}_P$ based on the noisy observation $\mathbf{Q}_R$ is exploited to estimate the channel. From the compressed sensing perspective, the sparse signal reconstruction problem under the case of multiple antennas is a multiple-measurement vector (MMV) problem. There is a group of measurement vectors for a group of signal vectors that are assumed to be jointly sparse and share a common support. To address this problem, the AMP algorithm is an efficient iterative thresholding approach [5], [17]. This algorithm provides an estimate $\hat{\mathbf{x}}_P$ based on $\mathbf{q}_R$ that minimizes the mean-squared error,

$$MSE = E[\|\hat{\mathbf{x}}_P(\mathbf{q}_R) - \mathbf{x}_P\|_2^2]. \tag{3}$$

By continuously updating the residuals, the AMP algorithm can gradually approaches the optimal solution, where each iteration can be represented as

$$\mathbf{x}_{P,k}^{\delta+1} = \eta_{k,\delta}((\mathbf{R}_P^{\delta})^* \mathbf{p}_{T,k} + \mathbf{x}_{P,k}^{\delta}), \tag{4}$$

and the residual can be updated by

$$\begin{aligned} \mathbf{R}_P^{\delta+1} &= \mathbf{Q}_R - \mathbf{X}_P^{\delta+1} \mathbf{P}_T \\ &+ \frac{K}{C} \mathbf{R}_P^{\delta} \sum_{k=1}^{K} \frac{\eta'_{k,\delta}((\mathbf{R}_P^{\delta})^* \mathbf{p}_{T,k} + \mathbf{x}_{P,k}^{\delta})}{K}, \end{aligned} \tag{5}$$

where $\delta = \{0, 1, ...\}$ indicates the index of the iteration, $\mathbf{R}_P \in \mathbb{C}^{N_r \times C}$ denotes the corresponding residual, $\eta_{k,\delta}$ is the so-called denoiser [5], and $\eta'_{k,\delta}$ denotes the first-order derivative of $\eta_{k,\delta}$.

2) Attack model: We assume there are one or multiple attackers who can masquerade as legitimate users by modifying its own pilot sequence into that of the legitimate user. It may even compromise the authentication key after sniffing the communication between the BS/AP and LUs for enough time. The location of the PCA attacker in the network is arbitrary but cannot be the same as LUs. The PCA attacker can take either attack mode below [18]:

- *Attack mode one*: The PCA attacker can transmit the deceiving pilot signal to the receiver regardless of the LU's activity.

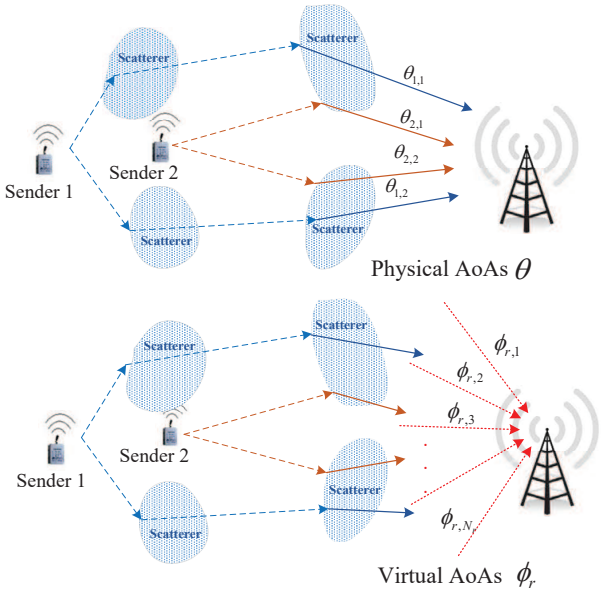


Fig. 2: Physical AoAs and virtual AoAs.

- *Attack mode two*: The PCA attacker can listen to the channel of the victim and start to transmit a deceiving pilot to the receiver when the victim is not transmitting.

As such, PCA attackers expect that the receiver would accept the spoofing signal from the PCA attacker as an LU, thus they can gain the privilege to further perform advanced attacks.

B. Motivation of our method

Our PCA detection scheme is motivated by the following observations:

1) *MmWave channels are sensitive to the location of the sender.* Due to the high directivity of mmWave, the signals from different senders are likely to travel different multipaths when arriving at the receiver. As illustrated in Fig. 2, different senders with different locations will lead to different physical AoAs at BS/AP, since the facing scatterers are different.
2) *Virtual AoAs in CVR can reflect the unique features of physical AoAs.* Although the virtual receive directions in CVR is fixed, the nonzero elements of the virtual AoAs can reveal the unique features of physical AoA. In other words, different physical AoAs will lead to different elements of virtual AoAs, as illustrated in Fig. 2.
3) *Virtual AoAs are easy to acquire compared with physical AoAs.* In contrast to physical AoAs, virtual AoAs are not difficult to acquire. Section IV will provide a theoretical analysis.

Based on these observations, we propose to exploit the features of CVR to counter PCA. In the 5G grant-free IoT network, the state and the statistical property of the CVR corresponding to each fixed pilot sequence can be easily recorded, since the BS/AP generally has sufficient storage and computing resources. Furthermore, due to the high carrier frequency used in mmWave such as 28GHz and 60GHz, if the PCA attacker wants to gain a similar CVR to the LU, the distance between them must be less than 5mm and 2.5mm, respectively [19]. In this case, it is very difficult for an attacker to be so close to a LU without being detected.

Thus, if the pilots are only coming from the genuine sender, the current received CVR and the recorded one should be similar since they have the same location. Conversely, the similarity would be degraded under the PCA case, in which there are more transmitters in the channel (attack mode one) or the location of the transmitter is different (attack mode two). Section IV-C will introduce the characteristics of CVR that can be used to distinguish these different situations. As a result, we can use the similarity between the recorded CVR and the current one corresponding to the same device ID to detect whether the current communication is corrupted by PCA.

IV. CVR Characteristics and Challenges

This section will provide a theoretical analysis of CVR, introduce the estimation of virtual AoAs, show the characteristics of CVR, and point out the challenges to achieve a desirable PCA detection.

A. CVR

CVR was first discussed in [20] and has been applied in mmWave communications. The model of channel virtual representation is based on a mmWave geometry channel model, where a sparse multi-path structure of mmWave communication can be represented by a discrete physical model [20]. To solve the issue that the physical angular directions are difficult to acquire, channel virtual representation is proposed to characterize the mmWave channel by fixed virtual receive and transmit directions. CVR can be seen as a mapping of the real channel in a special space based on the discrete Fourier transform (DFT). Furthermore, a fast estimation method to obtain CVR was provided in [21].

Based on the geometry channel model with L scatters formed by ray tracing, a sparse multipath structure of mmWave can be represented by a discrete physical model [20]. Based on Section III-A, we consider that BS/AP with N_r antennas communicates with K users each with a single antenna. Thus, the channel between the BS/AP and the k^{th} user can be represented as the sum of L paths in the form

$$\mathbf{H}_k = \sqrt{\frac{N_r}{\rho}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_r(\theta_l), \tag{6}$$

where ρ is the average path loss, α_l represents the fading coefficient of the $l-\text{th}$ propagation path, $\theta_l \in [-\pi/2, \pi/2]$ denotes the physical AoA, and the vector $\mathbf{a}_r(\theta_l) \in \mathbb{C}^{N_r \times 1}$ indicates the response array for receiving antenna arrays.

CVR can be applied to represent the mmWave channel by fixed virtual receive directions at BS/AP [20], [22]. If there is an antenna array consisting of an N_v dimensional uniform linear array, the virtual representation corresponds to system representation with respect to uniformly spaced spatial angles

$\vartheta_i = i/N_v, i = 0, ..., N_v - 1$. The corresponding steering vectors can be defined by $\phi_i = \arcsin(\lambda\vartheta_i/d)$, where d denotes the antenna spacing and λ indicates the wave-length of operation [22]. Thus, a unitary Discrete Fourier Transform (DFT) matrix $N_v \times N_v$ can be obtained,

$$\mathbf{U} = \frac{1}{\sqrt{N_v}}[\mathbf{a}(\phi_0), ..., \mathbf{a}(\phi_{N_v-1})]^T, \tag{7}$$

where $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I}$, and $\mathbf{U}^*$ is the conjugate transpose of $\mathbf{U}$.

Based on this unitary DFT matrix, we have the channel virtual representation,

$$\mathbf{H} = \mathbf{U}_r\mathbf{H}_V = \sum_{\omega=1}^{N_r} H_V(\omega)\mathbf{a}_r(\phi_{r,\omega}), \tag{8}$$

where $\mathbf{U}_r \in \mathbb{C}^{N_r \times N_r}$ is the unitary DFT matrix, which can reflect the fixed virtual receive angles that uniformly sample the unit angle space. $\mathbf{H}_V \in \mathbb{C}^{N_r \times 1}$ is the CVR matrix, in which the entry $H_V(\omega)$ captures the gains of the corresponding paths, and $\{\phi_{r,\omega}\}$ denotes virtual AoAs. It is worth noting that the gain of paths is random due to the channel fading effect [20].

Based on Eq. (6) and Eq. (8), and according to [20], the relationship between channel virtual representation and the physical channel model can be represented by

$$H_V(\omega) = \sum_{l=1}^{L} \alpha_l f(\theta_l - \frac{\omega}{N_r}), \tag{9}$$

where the function $f(.)$ represents a smoothing kernel that get peaky around the origin with increasing N_r [20].

From Eq. (9), we can see that the virtual representation $H_V(\omega)$ are samples of a smoothed version of scatters at virtual angles [20]. Thus, the index of these path gains on the virtual angle vectors (i.e., the location of the nonzero elements in virtual AoAs), can reflect the characteristics of the physical angles of all main scatters.

B. Obtaining virtual AoA

If we have a channel estimator to observe the received training signal, based on Eq. (8), Eq. (2) can be transformed into the angular domain via spatial DFT,

$$\mathbf{Q}_{\mathrm{VR}}^{(k)} = \mathbf{U}_r^*\mathbf{Q}_R^{(k)} = \tau_k\mathbf{H}_V^{(k)}\mathbf{p}_{T,k} + \mathbf{U}_r^*\mathbf{e}_P, \tag{10}$$

where $\mathbf{Q}_{\mathrm{VR}}^{(k)} \in \mathbb{C}^{N_r \times C}$ indicates the received signals corresponding to k^{th} user on angular domain based on spatial DFT, $\mathbf{e}_P \in \mathbb{C}^{N_r \times C}$ indicates the additive Gaussian noise.

Ideally, the virtual AoAs at the receiver side are indexed by the nonzero elements of $\mathbf{H}_V^{(k)}$. The locations of the nonzero elements of virtual AoAs in $\mathbf{H}_V^{(k)}$ can be identified as long as the corresponding elements in $\mathbf{Q}_R$ are nonzero. In practice, the additive noise $\mathbf{U}_r^*\mathbf{e}_P$ will influence the elements in $\mathbf{Q}_{\mathrm{VR}}^{(k)}$. In this case, a simple energy detection can be an optimal detector to estimate $\mathbf{Q}_{\mathrm{VR}}^{(k)}$. Thus, by collecting the received energy on each element in $\mathbf{Q}_{\mathrm{VR}}^{(k)}$, we can use a predefined threshold to decide whether the element in $\mathbf{Q}_{\mathrm{VR}}^{(k)}$ is noise or not, and this threshold can be decided by the noise variance. As a result, we can easily obtain the virtual AoAs based on pilot signal $\mathbf{p}_{T,k}$ and the corresponding received signals $\mathbf{Q}_{\mathrm{VR}}^{(k)}$, without complicated signal processing and estimation techniques.

Furthermore, if there is a pilot contamination attacker who sends the same pilots as k^{th} user, the received signal on angular domain based on spatial DFT can be represented as

$$\widetilde{\mathbf{Q}}_{VR}^{(k)} = \mathbf{U}_r^*\widetilde{\mathbf{Q}}_R = \widetilde{\mathbf{H}}_V^{(k)}\mathbf{P}_T^{(k)} + \mathbf{U}_r^H\mathbf{N}, \tag{11}$$

where $\widetilde{\mathbf{H}}_V^{(k)}$ denotes the CVR of the channel between PCA attacker and the receiver, and $\widetilde{\mathbf{Q}}_R^{(k)}$ indicates the corresponding received signals.

From Eq. (11), we can see that even with the same pilot sequences, the received signal $\widetilde{\mathbf{Q}}_{VR}^{(k)}$ is different from $\mathbf{Q}_{VR}^{(k)}$ since the channel is different. Thus, the corresponding $\widetilde{\mathbf{H}}_V^{(k)}$ is distinct from $\mathbf{H}_V^{(k)}$ when the location of the sender is different.

C. CVR Characteristics

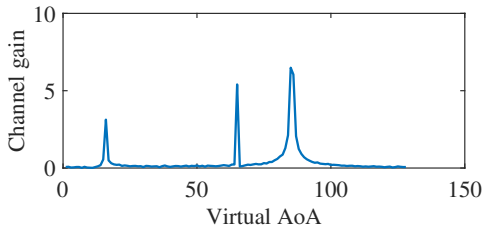
Fig. 3 illustrates some examples of CVR in mMTC IoT networks under various scenarios, where a 28GHz mmWave channel with 1×128 antennas is considered and the number of main multi-paths is 3. From these examples, we notice three interesting properties of CVR:

- *CVR is sparse, and the positions of path gains on the virtual AoAs are distinct when the physical AoA is different.* Fig. 3(a) shows a CVR with one transmitter and the physical AoAs are $\theta_{1,2,3}^{(1)}$ under the non-line-of-sight (NLOS) scenario. Fig. 3(b) shows a similar CVR, in which it has the same transmitter and the same location as Fig. 3(a). We can see that they have similar virtual AoAs but different channel gains. Meanwhile, another CVR with a different transmitter with different physical AoAs ($\theta_{1,2,3}^{(2)}$) is given in Fig. 3(c). Compared with Fig. 3(a) and Fig. 3(b), the virtual AoAs and the corresponding channel gains in Fig. 3(c) are totally different.
- *The CVR with two transmitters has more paths than the channel with only one transmitter.* The CVR with two transmitters under the NLOS scenario is illustrated in Fig. 3(d). We can see that there are more paths in Fig. 3(d) than those in Fig. 3(a) or Fig. 3(b). This property suggests that a single transmitter and multiple transmitters have significant differences on CVR, even if the used pilot sequences are same.
- *CVR can reflect the characteristics of the channels under NLOS and LOS, respectively.* Fig. 3(e) presents a CVR with one transmitter under the LOS scenario. We notice that there is a single strong path gain in CVR, which is significantly distinct from Fig. 3(a) or Fig. 3(c) under NLOS scenarios.

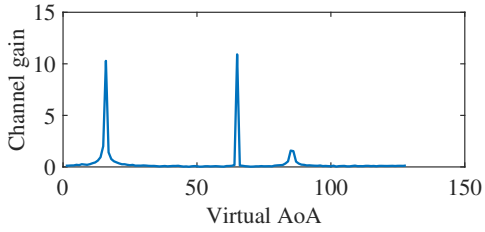
These properties imply that CVR can reflect the unique characteristics of the channel under various scenarios, and that can be used for PCA detection.

D. Challenges

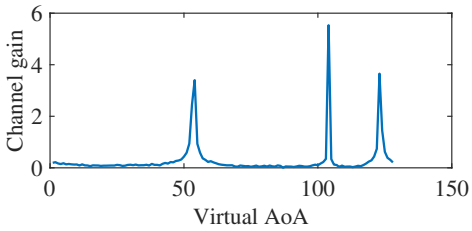
In fact, to achieve a desirable PCA detection based on CVR in grant-free 5G IoT networks is non-trivial. We must tackle the following three issues:



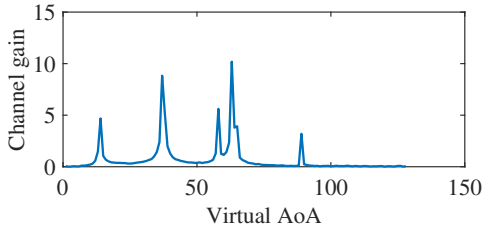
(a) One NLOS transmitter with physical AoAs ($\theta^{(1)}_{1,2,3}$).



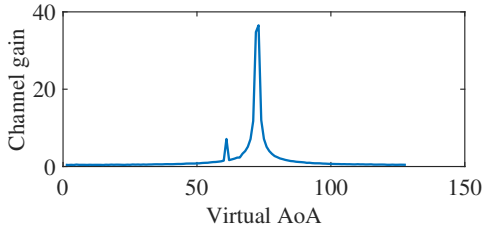
(b) One NLOS transmitter with physical AoAs ($\theta^{(1)}_{1,2,3}$).



(c) One NLOS transmitter with physical AoAs ($\theta^{(2)}_{1,2,3}$).



(d) Two transmitters under NLOS.



(e) One transmitter under LOS.

Fig. 3: CVR under various scenarios. (The number of antennas at BS/AP is 128 and the number of the main multi-paths is 3.)

1) *Channel gains in CVR are random.* Although the physical AoAs hinge on the location of the sender, the corresponding channel gain for each path is random due to the channel fading.
2) *It is difficult to define a threshold to differentiate the PCA case from normal case.* Since it is hard to obtain the statistical distribution of CVR, the traditional detection schemes, such as Neyman-Pearson (NP) testing, cannot be exploited directly. Furthermore, owing to the randomness of channel gains in CVR and the influence of noise, directly using a distance metric to separate the two cases makes it hard to achieve a desirable PCA detection performance.
3) *It is hard to label every IoT device in mMTC grant-free IoT networks.* Classification machine learning-based methods can provide a nonlinear feature learning to separate the different CVRs from different transmitters with different locations. However, training a classification machine learning model to recognize all of the LUs will be a huge task in the mMTC grant-free IoT network since there are huge numbers of IoT devices. Furthermore, it is difficult to obtain the training samples of PCA attackers.

As a result, we need an effective method to achieve a desirable PCA detection based on CVR in mMTC grant-free IoT networks.

V. SHMM-Siamese Network

The first Siamese network was reported in [23], where the Siamese network is used to distinguish the signature. In recent works, Siamese networks were explored to achieve face verification [24], image recognition and object tracking [25]–[27]. For example, the Siamese network consisting of convolutional neural networks (CNN) was introduced to transform the problem of image object tracking into a matching problem of path blocks [25]. Different from these Siamese networks, in this work, we propose a novel efficient tailored Siamese network construction consisting of a single hidden layer and a measurement layer that can achieve multiple distance measurements, i.e., SHMM-Siamese network. This structure of the neuron network has the advantages of low training and storage overhead. It enables an effective PCA detection with a small size of samples in grant-free networks.

The goal of the SHMM-Siamese network is to optimize its network parameters to enhance the distinguishability between the normal case and the PCA case. Resorting to the unique characteristics of Siamese neural networks and GANs, this detection scheme has the following three salient features: (i) The dependency on labels is significantly reduced. In the training stage, the correlation state between two samples are labeled. That is, we only need to tell the training model whether the two samples are from the same device or not, rather than labeling each sample corresponding to the specific device; (ii) The scalability of this model is remarkable, where untrained data from new devices also can be recognized in the detection stage; (iii) It can be supported by small sizes of training samples. Since the goal of this learning model is the correlation state between samples rather than specific devices, it does not require a large number of training samples to train the model parameters. Furthermore, GANs can improve the diversity of the training dataset even under a small size of training samples.

In the following section, the problem of PCA detection is transformed into a similarity detection problem, then the network architecture, loss function and GANs based detection model training are introduced in turns.

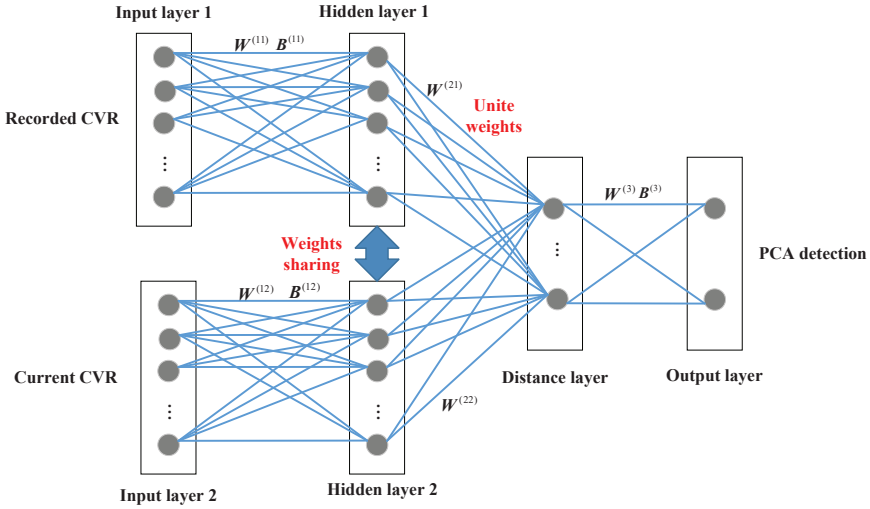


Fig. 4: SHMM-Siamese network architecture.

A. PCA detection to similarity detection

To tackle the challenges in Section IV-D, in this section, the PCA detection in mmWave grant-free IoT is transformed into a similarity detection based on CVR. Based on the analysis in Section III-B, we can detect the presence of the PCA attacker by judging whether the received CVR and the recorded CVR are similar.

Let $\mathbf{X}^{(1)} = [\mathbf{x}_n^{(1)}]$ and $\mathbf{X}^{(2)} = [\mathbf{x}_n^{(2)}]$ be a set of training sample pairs, $\mathbf{x}^{(i)} \in \mathbb{R}^{N_r \times M_{TR}}$, where $i \in \{1, 2\}$, $n = 1, ..., M_{TR}$ is the number of the training samples and N_r indicates the dimension of CVR, i.e, the number of the antennas at BS/AP. Here, the raw complex number in CVR is converted to real number. The corresponding state only has the target label, i.e., $\mathbf{Y} = [\mathbf{y}_n]^{2 \times M_{TR}}$. Thus, the SHMM-Siamese network aims to use the discriminant function $f : \chi \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ to train a corresponding machine learning model. For a set of testing data $\mathbf{X}_{test,m}^{(1)}$ and $\mathbf{X}_{test,m}^{(2)}$, the classifier can give the corresponding predictive values, i.e., $\widehat{\mathbf{Y}} = [\widehat{\mathbf{y}}_1, .., \widehat{\mathbf{y}}_{M_{TE}}]$, where $\widehat{\mathbf{y}} \in \{[0\ 1]^T, [1\ 0]^T\}$ and $m = [1, ..., M_{TE}]$ denotes the number of testing CVR pairs.

Let $\widehat{\mathbf{y}}_m = [0\ 1]^T$ indicate the normal case and $\widehat{\mathbf{y}}_m = [1\ 0]^T$ denote the spoofing attack case. The problem of PCA detection can be represented as

$$f(\mathbf{x}_{test,m}^{(1)}, \mathbf{x}_{test,m}^{(2)}) = \begin{cases} \widehat{\mathbf{y}}_m = [1\ 0]^T & \text{Normal} \\ \widehat{\mathbf{y}}_m = [0\ 1]^T & \text{PCA Attack.} \end{cases} \tag{12}$$

Next, we will introduce the architecture of the SHMM-Siamese network to achieve this machine learning-based PCA detection.

B. Network architecture

The SHMM-Siamese network consists of four layers, including an input layer, a twin hidden layer, a distance layer and an output layer, as shown in Fig. 4.

- *Input layer.* Different from traditional classification machine learning models that require one label corresponding to one type of samples, the input layer will receive data pairs as the input samples, where the correlation between this pair is the goal of learning. The number of input layer nodes equals to the dimension of CVR, i.e., $\mathbf{x}_{input}^{(1)} = [x_{input,1}^{(1)} \ ... \ x_{input,N_r}^{(1)}]$ and $\mathbf{x}_{input}^{(2)} = [x_{input,1}^{(2)} \ ... \ x_{input,N_r}^{(2)}]$.
- *Twin hidden layer.* In order to design a proper hidden layer which can be fed by the two different input layers, we construct a twin hidden layers structure as shown in Fig. 4, where two sets of neural nodes have the same weights and bias, i.e., $\mathbf{W}^{11} = \mathbf{W}^{12}$ and $\mathbf{B}^{11} = \mathbf{B}^{12}$. Furthermore, the output of this layer can be represented by

$$\begin{aligned} \mathbf{z}^{(11)} &= g(\mathbf{W}^{(11)} \cdot \mathbf{x}_{input}^{(1)} + \mathbf{B}^{(11)}), \\ \mathbf{z}^{(12)} &= g(\mathbf{W}^{(12)} \cdot \mathbf{x}_{input}^{(1)} + \mathbf{B}^{(12)}), \end{aligned} \tag{13}$$

  where $g(\cdot)$ is the activation function such as sigmoid function.
- *Distance layer.* In this layer, the outputs of hidden layer are measured by multiple measurement metrics. Here, we use Euclidean distance and cosine distance as the two measurement metrics. Moreover, unite weights are used, i.e., $\mathbf{W}^{(21)} = \mathbf{W}^{(22)} = [1, ..., 1]$. The Euclidean distance is employed as the first metric,

$$D_1(\mathbf{z}^{(11)}, \mathbf{z}^{(12)}) = \left\| \mathbf{z}^{(11)} - \mathbf{z}^{(12)} \right\|^2, \tag{14}$$

  where $\|\cdot\|^2$ is the Frobenius norm.
  The consine distance is as the second metric,

$$D_2(\mathbf{z}^{(11)}, \mathbf{z}^{(12)}) = \frac{\mathbf{z}^{(11)} \odot \mathbf{z}^{(12)}}{\left\|\mathbf{z}^{(11)}\right\| \cdot \left\|\mathbf{z}^{(12)}\right\|}, \tag{15}$$

  where $\odot$ means inner product, and $\|.\|$ denotes the norm of vectors.
- *Output layer.* The output layer involves a two class classification problem, where two classes are corresponding to similar pairs and dissimilar pairs in the input layer,

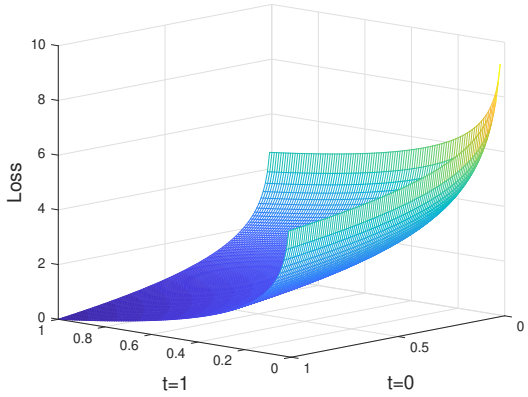


Fig. 5: Graph of the loss function.

respectively. The output of this layer can be given by

$$\mathbf{z}^{(2)} = g(\mathbf{W}_{i,n}^{(2)} \cdot \mathbf{D} + \mathbf{B}^{(2)}), \tag{16}$$

where $\mathbf{D} = [D_1 \ D_2]$.

C. Loss function

Here, by minimizing the loss function, the parameters in the model are continuously optimized to gain a better capacity for distinguishing the dissimilar signals from similar signals. The employed loss function can be represented as

$$Loss = (t\,(-\mathbf{y}_1 \ln \mathbf{a}_1) + (1-t)\,(-\mathbf{y}_2 \ln \mathbf{a}_2)) + \lambda \cdot \|\mathbf{W}\|_2, \tag{17}$$

where $t = 0$ denotes that two input channels are from a same transmitter with the same location, while $t = 1$ indicates the input channels with different transmitters, and $\mathbf{y}_1 = [1\ 0]^T$ and $\mathbf{y}_2 = [0\ 1]^T$. Moreover, $\mathbf{a}_i = e^{\mathbf{z}^{(i)}} / \sum e^{\mathbf{z}^{(i)}}$ and $i \in \{1, 2\}$. $\lambda \cdot \|\mathbf{W}\|_2$ indicates weight decay and λ denotes attenuation coefficient (empirical value $\lambda = 0.001$).

Furthermore, the graph of the loss function is provided in Fig. 5. It can optimize the parameters in the network that reduce the distance between the similar samples and increase the distance between the dissimilar samples, at the same time.

D. Network training based on GANs

Loosely speaking, the proposed SHMM-Siamese network is similar to a two-class classification machine learning model. In general, the training of the classification machine learning model needs to gain all types of training samples. However, for PCA detection in grant-free IoT networks, it is hard to predict the activity and the location of attackers. In other words, how to obtain negative training samples is a challenge for establishing a desirable machine learning model.

To tackle this issue, GANs are introduced to generate the negative training sample. GANs can learn a generative model as a two-player zero-sum game between a generator and a discriminator [28]. The generator tries to generate the target information from the candidate resource pool for the given simulation task, and the discriminator tries to distinguish the input data from the fake ones. When the discriminator in GAN fails to distinguish between the real data and the fake ones, the generator could generate high-quality signals with the same distribution as the input signals. Utilizing this feature of GANs, we can feed GANs the device samples as the real data and get a set of fake ones. These GAN samples have the same distribution as the IoT devices. Thus, we can use these GAN samples as the negative training samples to train the neural networks, i.e., the proposed SHMM-Siamese network. Furthermore, according to the principles of GANs, the process of generating a stable GAN model is random since the fake data in the GAN model is originally coming from a noise signal [28]. Thus, due to the randomness of GANs, the diversity of the training dataset can be significantly improved. The training process based on GANs is illustrated in Fig. 6. The legitimate IoT users can provide legitimate CVR as the similar sample pairs, and these CVR can be used to generate a number of GAN samples as the dissimilar sample pairs. Then these sample pairs can be used to train the proposed SHMM-Siamese network to achieve PCA detection.

TABLE I: Simulation parameters.

Notations	Parameters
$N_r = \{32, 64, 128, 256\}$	The number of antennas at BS/AP
$L = 3$	Channel scatters
0.01	The learning rate
$\lambda = 0.001$	The attenuation coefficient
$50, 100, 200, 300, 400$	The size of training sample pairs
1000	The size of testing sample pairs
$R_{AP} = \{1, 2, 4\}$	Attack-user power ratios
$28GHz$	Carrier frequency
$1 - 1000$	Iteration
$20, 50, 100, 300$	The number of hidden layer neurons
$SNR = \{-5, 0, 5, 10, 15\}dB$	Signal-to-noise-ratio

VI. Simulation Results

In this section, Monte Carlo experiment simulation results will be provided, involving the evaluation of the proposed SHMM-Siamese network and the validation of the proposed PCA detection scheme under various scenarios. We use MATLAB and Python to evaluate the collected measurement data on a general computer, which operates on a 64-bit system with a 16G memory and an i7-7700 CPU.

Moreover, to simplify the analysis, we define the PCA detection accuracy as P_A for various testings:

$$P_A = 1 - (P_{MD} + P_{FA}), \tag{18}$$

where $P_{MD} = 1 - P_D$, P_{MD} and P_D denote the miss detection rate and the detection rate respectively, and P_{FA} is the false alarm rate. Therefore, a higher P_A indicates a better detection performance. Furthermore, Table I shows the relevant simulation parameters and notations.

A. Evaluation of SHMM-Siamese network

For the SHMM-Siamese network model, we evaluate the impact of the loss function and the neurons in the hidden layer on the detection accuracy.

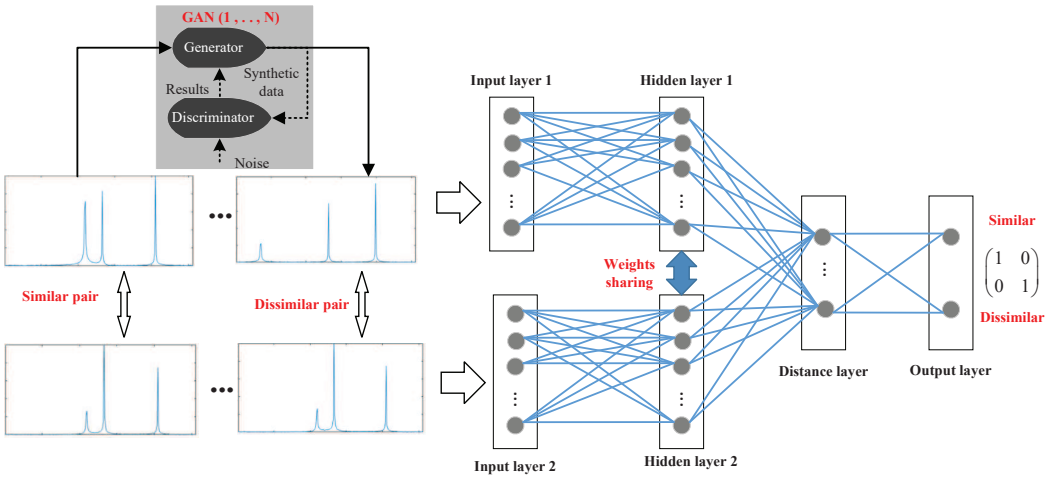


Fig. 6: SHMM-Siamese network training.

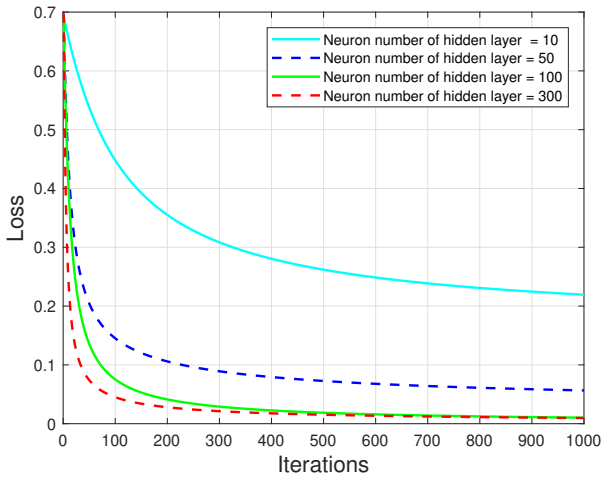


Fig. 7: Iteration number VS. Loss under different neuron numbers of the hidden layer. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

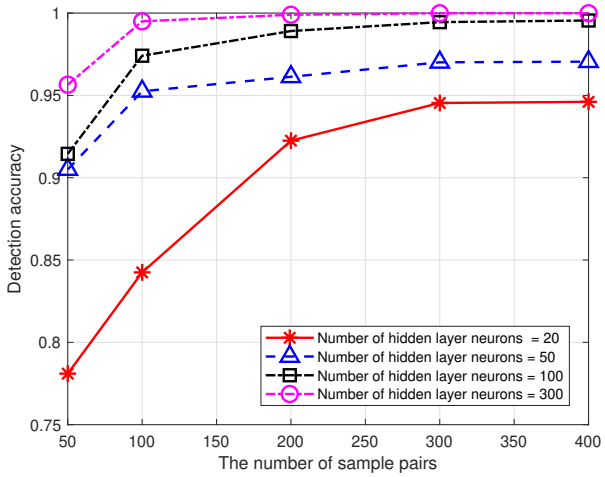


Fig. 8: Detection accuracy VS. Sample number under different neuron numbers of the hidden layer. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the loss is 10^{-2}; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

1) Effectiveness of loss function: The effect of the hidden layer on the loss function is shown in Fig. 7. We can see that the value of the loss function decreases steadily as the number of iterations increases. This implies that the proposed machine-learning scheme does not overfit. Under the same testing samples and parameters, the network model with more neurons in the hidden layer is easier to quickly converge to a lower loss value. For example, when the iteration number is 1,000, the loss value of the network model with 10 neurons in the hidden layer gradually reaches 0.2, while the loss value of the model with 300 neurons in the hidden layer can approach 0.01. Furthermore, we notice that the curve of the network model with 100 neurons in the hidden layer is close to that of the network model with 300 neurons in the hidden layer.

2) Detection performance under different training sample sizes: In Fig. 8, the detection performance of the proposed Siamese network with different training sample pairs is provided. As the training sample pairs increase, the detection accuracy of the Siamese network model is improved gradually. We can see that the model with more hidden layer nodes can achieve a higher detection performance with fewer training samples. For instance, when the number of hidden layer neurons is 300, the corresponding detection performance can reach 99% even if the size of the training sample pairs is 200. Moreover, when the number of sample pairs is more than 300, the detection performance becomes stable. This implies that the proposed Siamese network does not need a large number of training samples to train the machine learning model.

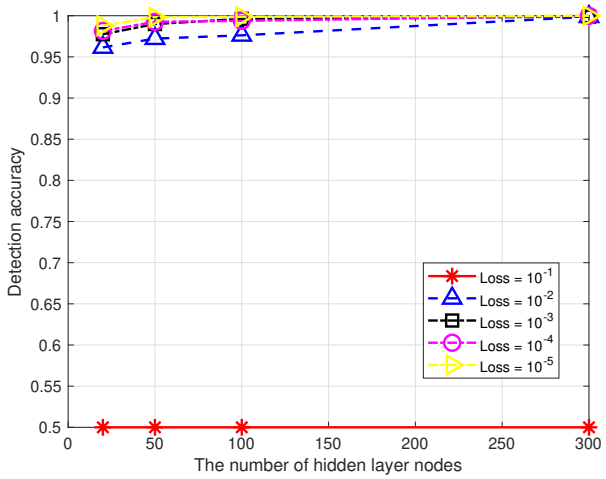


Fig. 9: Detection accuracy VS. Neuron number of the hidden layer under different loss levels. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

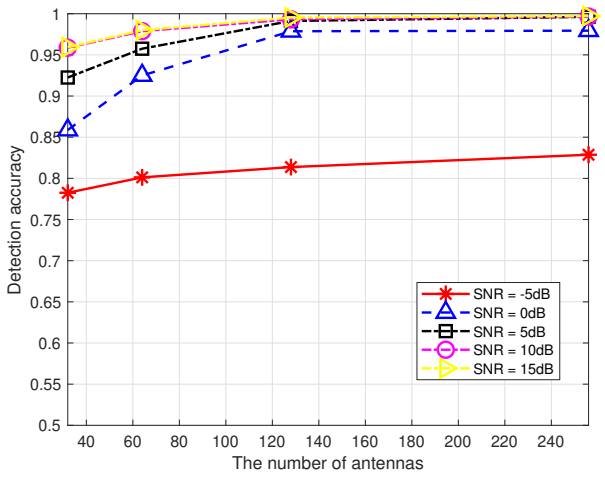


Fig. 10: Detection accuracy under attack mode one with different antenna numbers and different SNRs. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

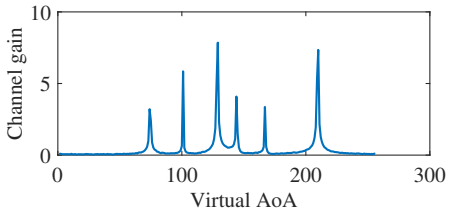
3) Detection performance with different hidden layer sizes: Fig. 9 shows the effect of the loss value on detection performance. When the loss level is set at 10^{-1}, the detection accuracy is around 50%. That means that the detection mechanism cannot achieve the detection task. Meanwhile, when the loss level is higher than 10^{-2}, the best detection performance that the detection model can achieve will exceed 95%. For example, as the number of neurons in the hidden layer is 50, the detection accuracy can be higher than 96% if there are sufficient training sample pairs. This indicates that the proposed detection model does not need a very low value of the loss function to achieve a better detection performance.

B. Effectiveness in different PCA detection scenarios

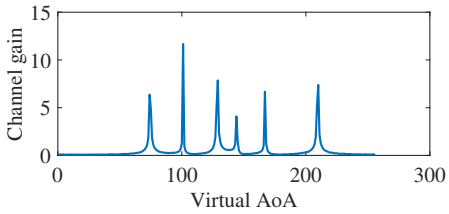
Here, the effectiveness of the proposed detection scheme will be verified under different PCA scenarios, including different attack modes illustrated in Section III-A2, and LOS and NLOS scenarios.

1) Detection performance under the attack model one: The detection performance under the attack mode one is shown in Fig. 10. Different signal-to-noise-ratio (SNRs) and different antenna numbers at AP could impact detection performance. From Fig. 10, we can see that under the same SNR condition, using more antennas can improve detection performance. Take the curve with SNR=0 as an example; when the number of antennas is 32, the detection accuracy is just around 85%, while it can reach 97% as the number of antennas is increased to 128. We also notice that when the SNR condition is higher than 0dB and the number of antennas is more than 64, the detection accuracy can exceed 90%. This implies that the proposed model has a good anti-noise property.

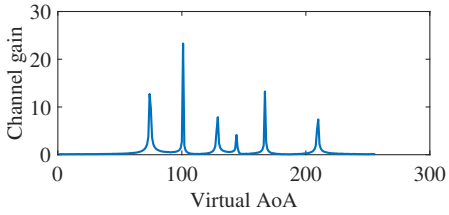
Furthermore, the proposed scheme with different transmit powers of the attacker is evaluated. For consistency, we



(a) Attacker-user power ratio $R_{AP} = 1$.



(b) Attacker-user power ratio $R_{AP} = 2$.



(c) Attacker-user power ratio $R_{AP} = 4$.

Fig. 11: CVR under different attacker-user power ratios scenarios. (The number of antennas at AP is 256 and the number of the main multi-paths is $L = 3$.)

consider an attacker-user power ratio as a metric to indicate the power of the attacker, given by

$$R_{AP} = \frac{P_{attacker}}{P_{user}} \tag{19}$$

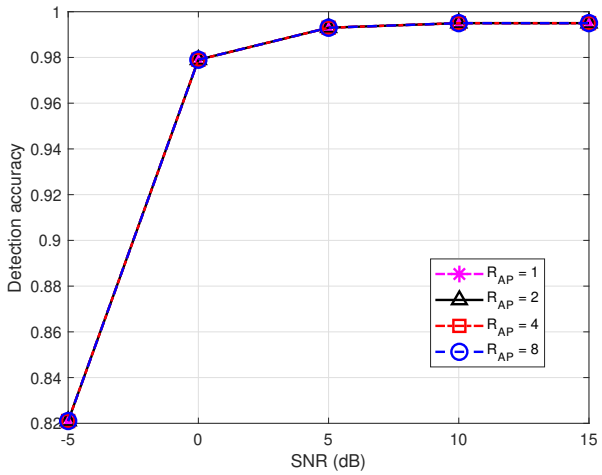


Fig. 12: Detection accuracy under different attacker-user power ratios. In this example, the carrier frequency is 28GHz; the number of antennas at BS/AP is 128; the size of training sample pairs is 400; the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

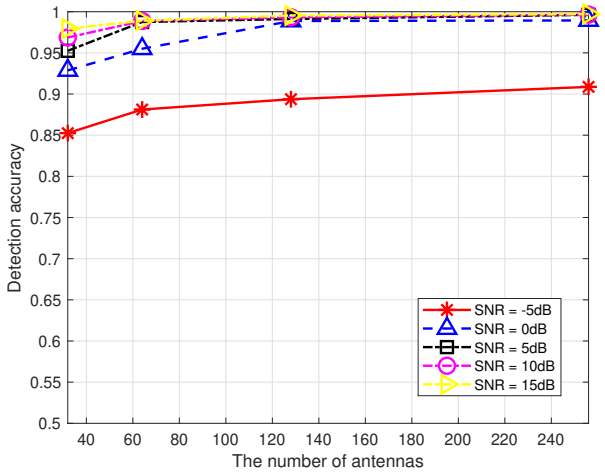


Fig. 13: Detection accuracy under attack mode two with different antenna numbers and different SNRs. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

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

The CVR with different attacker-user power ratios is illustrated in Fig. 11. Fig. 11(a) shows the scenario when the transmitting power of the attacker is the same as the legitimate user's power. Fig. 11(b) and Fig. 11(c) show the channel gains VS. AoAs for $R_{AP} = 2$ and $R_{AP} = 4$, respectively. We notice that although the number of paths with a high gain in 11(c) is more than that of Fig. 11(a), the position of the CVR is not affected by the attacker's power level. Fig. 12 shows the detection performance under different attacker-user power ratios. We can see that there is no significant difference between different attacker-user power ratios. It implies that for the proposed scheme, the attacker's power does not affect the detection performance under the attack model one.

2) Detection performance under the attack model two: Fig. 13 shows the detection performance with different SNRs under the attack mode two. Compared with the attack mode one, the difference between the PCA case and the normal case is more significant in the attack mode two since there are more multi-paths in the CVR under this attack case, as shown in Fig. 3. For this reason, from Fig. 13, we can see that all curves have a better detection performance. Take 128 antennas as an example, the detection accuracy can reach 98% when the SNR=0dB, and it will be close to 90% even when the SNR is -5dB.

3) Detection performance under attack model one mixed attack model two: Both attack mode one and attack mode two are simultaneously considered in Fig. 14, where half of the training and testing samples are taken from attack mode one and the other half are from attack mode two. We notice that in Fig. 14, the detection performance these curves can reach falls in between Fig. 10 and Fig. 13. For example, the detection performances in Fig. 10 and Fig. 13 are around

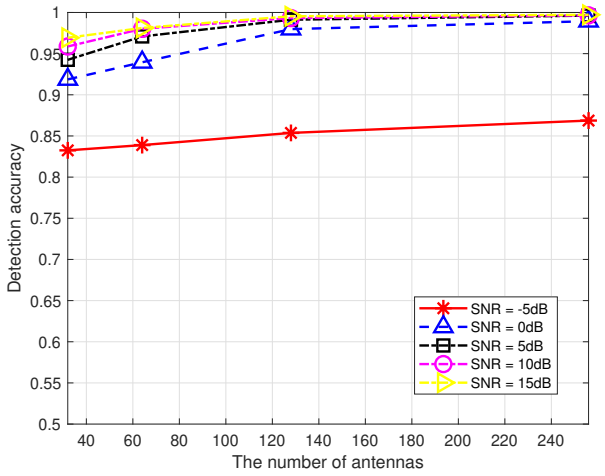


Fig. 14: Detection accuracy under attack mode one and mode two. In this example, the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

80% and 90%, respectively, when SNR=-5dB. By contrast, the detection accuracy lies in 85% in Fig. 14, under the same environment. This implies that the proposed PCA detection scheme can handle different attack modes even the mixed attack mode.

4) Detection performance under LOS and NLOS scenario: Fig. 15 illustrates the detection performance of the proposed scheme under the LOS scenario, where there is a strong LOS path in the channel. We can see that the proposed detection can gain a high detection accuracy under the LOS scenario. For instance, when the SNRs are higher than 0dB, all performance curves can exceed 0.95. Meanwhile, Fig. 16

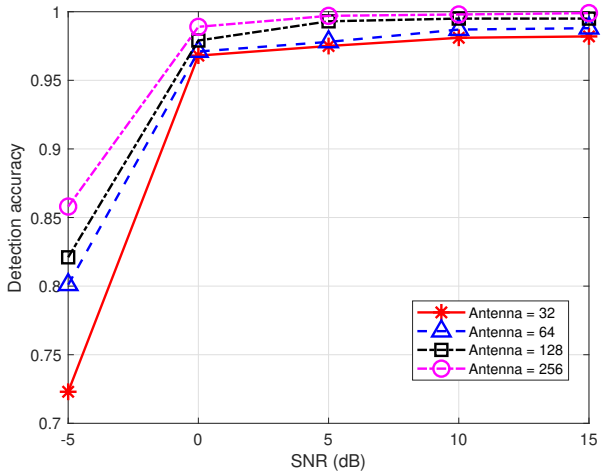


Fig. 15: Detection accuracy under LOS scenario with different antenna numbers and different SNRs. In this example, the attack mode one is considered; the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

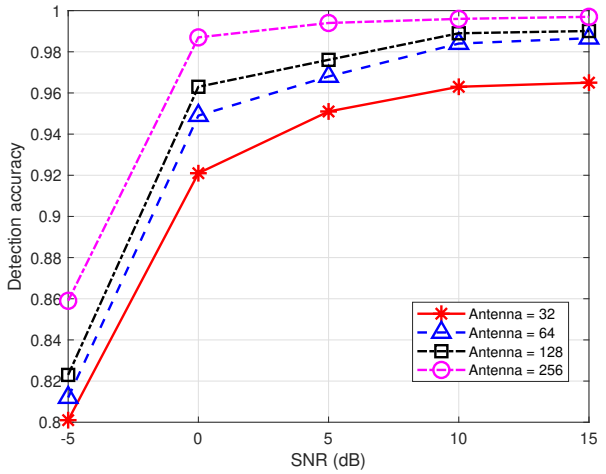


Fig. 16: Detection accuracy under NLOS scenario with different antenna numbers and different SNRs. In this example, the attack mode one is considered; the carrier frequency is 28GHz; the number of antennas at AP is 128; the size of training sample pairs is 400; NLOS scenario is considered and the number of path is $L = 3$; the learning rate of the learning model is 0.01 and the attenuation coefficient is $\lambda = 0.001$.

shows the detection performance of the proposed scheme under the NLOS scenario with the same conditions. We notice that most curves in Fig. 15 have a better detection performance than those in Fig. 16. Take the curve with 64 antennas as an example; with the SNR=0dB, the detection accuracy under the NLOS scenario is 96%, while it can reach 97% under the LOS scenario. It suggests that the distinguishability between the attack case and the normal case under the LOS scenario is more significant than that under the NLOS scenario.

VII. Conclusion

In this paper, we investigated the issue of the pilot contamination attack in mMTC grant-free 5G IoT networks. To counter this physical layer threat, CVR was introduced to detect the pilots of the IoT devices in grant-free 5G IoT networks, where the PCA detection was transformed into a classification problem based on the similarity of CVR. Furthermore, we proposed a novel SHMM-Siamese network to achieve this detection, where a new loss function was provided and GANs were introduced to enhance the model training. Simulation results showed that the proposed detection scheme can achieve desirable detection performance under various scenarios. The detection accuracy can approach 99% with 128 antennas at the BS/AP and it can reach higher than 95% even when the SNR is 0dB. For future works, there are two interesting and valuable research directions: (i) how to achieve the detection of pilot contamination attacks under mobile scenarios using CVR; (ii) how to tackle the co-located attacks in the grant-free 5G IoT networks.

References

[1] N. A. Mohammed, A. M. Mansoor, and R. B. Ahmad, "Mission-critical machine-type communication: An overview and perspectives towards 5g," *IEEE Access*, vol. 7, pp. 127 198–127 216, 2019.

[2] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120–128, 2017.

[3] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. De Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, 2018.

[4] Y. Du, B. Dong, Z. Chen, X. Wang, Z. Liu, P. Gao, and S. Li, "Efficient multi-user detection for uplink grant-free noma: Prior-information aided adaptive compressive sensing perspective," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2812–2828, 2017.

[5] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.

[6] N. Wang, L. Jiao, A. Alipour-Fanid, M. Dabaghchian, and K. Zeng, "Pilot contamination attack detection for noma in 5g mm-wave massive mimo networks," *IEEE Transactions on Information Forensics and Security*, vol. 15, no. 1, pp. 1363–1378, 2019.

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

[8] X. Wang, M. Liu, D. Wang, and C. Zhong, "Pilot contamination attack detection using random symbols for massive mimo systems," in *IEEE 85th Vehicular Technology Conference (VTC Spring)*. IEEE, 2017, pp. 1–7.

[9] N. Wang, L. Jiao, and K. Zeng, "Pilot contamination attack detection for noma in mm-wave and massive mimo 5g communication," in *2018 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.

[10] 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.

[11] W. Zhang, H. Lin, and R. Zhang, "Detection of pilot contamination attack based on uncoordinated frequency shifts," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2658–2670, 2018.

[12] X. Zhou, B. Maham, and A. Hjorungnes, "Pilot contamination for active eavesdropping," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 903–907, 2012.

[13] N. Wang, P. Wang, A. Alipour-Fanid, L. Jiao, and K. Zeng, "Physical-layer security of 5g wireless networks for iot: Challenges and opportunities," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8169–8181, 2019.

[14] D. Kapetanovic, A. Al-Nahari, A. Stojanovic, and F. Rusek, "Detection of active eavesdroppers in massive mimo," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2014, pp. 585–589.

[15] J. K. Tugnait, "Self-contamination for detection of pilot contamination attack in multiple antenna systems," *IEEE Wireless Communications Letters*, vol. 4, no. 5, pp. 525–528, 2015.

[16] J. Yang, S. Xie, X. Zhou, R. Yu, and Y. Zhang, "A semiblind two-way training method for discriminatory channel estimation in mimo systems," *IEEE Transactions on Communications*, vol. 62, no. 7, pp. 2400–2410, 2014.

[17] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[18] M. H. Yılmaz and H. Arslan, "A survey: Spoofing attacks in physical layer security," in *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*. IEEE, 2015, pp. 812–817.

[19] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2017.

[20] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.

[21] 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.

[22] 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.

[23] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," in *Advances in neural information processing systems*, 1994, pp. 737–744.

[24] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, pp. 539–546.

[25] R. Tao, E. Gavves, and A. W. Smeulders, "Siamese instance search for tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2016, pp. 1420–1429.

[26] Q. Wang, Z. Teng, J. Xing, J. Gao, W. Hu, and S. Maybank, "Learning attentions: residual attentional siamese network for high performance online visual tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2018, pp. 4854–4863.

[27] A. He, C. Luo, X. Tian, and W. Zeng, "A twofold siamese network for real-time object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018, pp. 4834–4843.

[28] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C.-X. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 22–27, 2019.