Deep Learning for Detection and Identification of Asynchronous Pilot Spoofing Attacks in Massive MIMO Networks

Fuad Choudhury, Aissa Ikhlef, *Senior Member, IEEE*, Walid Saad, *Fellow, IEEE* and Merouane Debbah, *Fellow, IEEE*

***Abstract*—Massive multiple-input multiple-output (MIMO) networks are highly vulnerable to an active eavesdropping attack called pilot spoofing attack. The pilot spoofing attack causes information leakage to the active eavesdropper (ED) and also weakens the strength of the signal received by the attacked legitimate user equipment (UE) during the downlink transmission. In this paper, a deep neural network, called identification network (IDNet), is proposed to detect asynchronous pilot spoofing attacks and identify the attacked UE. We show that an asynchronous pilot spoofing attack leads to increasing the signal subspace dimension by one unlike the synchronous one. This property is then exploited to improve the attack detection/identification accuracy. In the proposed IDNet, the input features are the eigenvalues of the sample covariance matrix of the received signal at the base station (BS) as well as the ratio between the power of the received signal at the BS projected onto the pilot signals and its expected value. Numerical results show the effectiveness of IDNet in identifying the attacked UE and reveal that the larger the timing and/or frequency mismatches of the ED, the higher the identification accuracy confirming that asynchronous pilot spoofing attacks can be identified more accurately than synchronous pilot spoofing attacks.**

***Index Terms*—Deep neural networks, massive multiple-input multiple-output, physical layer security, asynchronous pilot spoofing attack, attack detection and identification.**

I. Introduction

Massive multiple-input multiple-output (MIMO) techniques have demonstrated their effective ability to help meet the exponentially increasing demand in wireless networks [1]–[3]. Indeed, massive MIMO is an important enabler of the fifth generation (5G) of wireless networks and is expected to play an important role in the upcoming sixth generation (6G) of wireless networks [4], [5]. In particular, massive MIMO systems can achieve unprecedented gains in terms of spectral and energy efficiency [1]. However, despite this promising outlook of massive MIMO, wireless networks that exploit this technology are highly vulnerable to active eavesdropping attacks known as pilot spoofing attacks [6], [7]. In a massive MIMO network, given the large number of antennas at the base station (BS), it is more practical to adopt the time division duplex (TDD) transmission mode while performing channel estimation during uplink. In a pilot spoofing attack, during the uplink channel estimation phase, an active eavesdropper (ED) transmits the same pilot signal as that of the legitimate user equipment (UE) under attack. As a result, the BS estimates the uplink composite channel of the UE under attack and the ED. Hence, in the downlink phase, a portion of the signal intended for the UE under attack will be leaked towards the ED, without the knowledge of the BS. Naturally, a larger ED pilot signal power will lead to a larger leakage and a weaker received signal at the UE under attack. Hence, there is a need to address several challenges related to pilot spoofing attacks, that include detection and design of efficient and robust countermeasures.

Several works in the literature investigated the above challenges by proposing detection schemes and countermeasures. The authors in [8] presented a detection approach, called the energy ratio approach, that exploited the asymmetry of the received signal power at the BS and UE and derived a detection threshold without knowledge of the UEs and ED channel state information (CSI). However, as the detection happens at the UE, the BS needs to send the calculated average power received in the uplink to the UE so that the UE can determine the presence or absence of attack. Additionally, the UE needs to notify the BS of an attack, if any, so that the BS can take appropriate countermeasures. This exchange between the BS and UE will lead to a communication overhead including frame format change, which is undesirable. Similarly, the authors in [9] proposed three detection schemes for massive MIMO systems leveraging the received signal power at the UE and BS to derive an energy ratio that only requires knowledge of the noise variance and large-scale fading coefficients. Joint detection and localization of the eavesdropper in MIMO systems using spatial spectrums was proposed in [10]. The authors in [11] proposed a detection method that is based on the likelihood ratio test principle and uses the channel estimate. [12] proposed a detection method that is also based on likelihood ratio test principle but employs pilot manipulation in which UEs randomly partition their pilot sequences into two parts and multiply the second part by a

The work of Aissa Ikhlef has been supported by the CHEDDAR: Communications Hub for Empowering Distributed ClouD Computing Applications and Research funded by the UK EPSRC under grant numbers EP/Y037421/1 and EP/X040518/1. The work of Walid Saad was supported by the Office of Naval Research (ONR) under MURI grant N00014-19-1-2621.

F. Choudhury was with the Department of Engineering, Durham University, Durham, DH1 3LE, UK. He is now with the Markets Post Trade Group, Barclays Investment Bank, Knutsford, WA16 9EU, UK (e-mail: fuad3501@outlook.com).

A. Ikhlef is with the Department of Engineering, Durham University, Durham, DH1 3LE, UK (e-mail: aissa.ikhlef@durham.ac.uk).

W. Saad is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA, USA (email: walids@vt.edu).

M. Debbah is with the 6G KU Research Center, Khalifa University of Science and Technology, P O Box 127788, Abu Dhabi, UAE (e-mail: merouane.debbah@ku.ac.ae).

diagonal matrix. However, the works in [9] and [10] only considered a single user system and [9]–[12] assumed that the ED is perfectly synchronized with the UE.

The body of literature in [13]–[16] developed detection methods using source enumeration approaches, which are used in array signal processing to detect the number of incident signals. In these approaches, UEs are treated as sources and if the number of detected sources is equal to the number of UEs in the system then there is no ED and if the number of sources is greater than the number of UEs then an ED is present. The works in [13] and [14] applied source enumeration to detect the ED by superimposing a random signal onto the pilot signal of the UE and estimating the signal subspace dimension using the minimum description length (MDL) estimator. By doing so, in the presence of the ED, the signal subspace dimension increases by one and, hence, source enumeration can help determine the presence or absence of the ED. However, this comes at the expense of reducing the training power budget of the pilot signal thereby decreasing CSI estimation accuracy [13]. The authors in [15] also used an eigenvalue-based estimator using Akaike's information criterion (AIC), which outperformed MDL for shorter pilot signals. The authors in [16] presented a jamming attack detection scheme in a massive MIMO network where random matrix theory was used to analyze the largest eigenvalues of the covariance matrix and outperformed MDL. Other source enumeration methods could also be adapted for ED detection [17]–[19].

All of the prior art [8]–[16] and references therein, assumed perfect timing and frequency synchronization between the attacked UE and ED during the uplink pilot signal transmission phase, which is difficult to achieve in practice. Also, the existing works [8]–[10], [13]–[16] used conventional signal processing techniques to detect whether an attack is present but they are unable to identify the attacked UE. Being able to identify the attacked UE is of paramount importance as in this case only the service to the identified attacked UE could be interrupted and not to all the UEs in the network.

In contrast to these prior works, the main contribution of this paper is a novel framework that uses deep neural networks (DNNs) to simultaneously detect asynchronous pilot spoofing attacks and identify the attacked UE in massive MIMO systems. Although DDNs were used in the detection of the number of sources in array signal processing [20]–[22], to our knowledge, they have not been exploited for the problem of the detection and identification of pilot spoofing attacks in massive MIMO systems. In contrast to prior works such as [23] that used support vector machines (SVMs) to detect detect active eavesdropping attacks in single antenna multiuser communications, our approach has five distinguishing characteristics: 1) We consider a massive MIMO setup while [23] considered a single antenna multiuser setup, 2) unlike [23]; we consider the more practical case in which the ED is not in perfect synchronization with the BS, 3) we use DNNs instead of SVMs, 4) we use different features such as the eigenvalues of the covariance matrix of the received signal as input to the DNNs, 5) in addition to detection of attack, we propose to identify the attacked UE. In summary, our key contributions include:

- We consider a new practical scenario in which the ED, unlike the legitimate UEs, is not in perfect synchronization with the BS during the uplink pilot signal transmission.
- We show that if the ED is present and not in perfect timing and/or frequency synchronization with the BS, the dimension of the signal subspace increases by one, which we exploit to detect the presence of the ED and identify the attacked UE.
- We propose a DNN-based approach, called identification network (IDNet), deployed at the BS to simultaneously detect pilot spoofing attacks and identify the attacked UE. This approach enables the BS to selectively terminate the downlink transmission to the affected UE, minimizing downtime for unaffected UEs. To our best knowledge, no existing work has considered identifying attacked UEs in massive MIMO networks suffering from pilot spoofing attacks.
- We also propose to use multiple frames to improve the identification accuracy.
- Comprehensive simulation results show that the proposed IDNet is effective in identifying the attacked UE and that asynchronous pilot spoofing attacks are easier to be detected and identified compared to synchronous pilot spoofing attacks.

The rest of the paper is organized as follows. In Section II, we introduce the proposed model with the asynchronous pilot spoofing attack. In Section III, the proposed DNN architecture and feature vectors are presented. Section IV provides simulation results and discusses the performance of the proposed architecture. In Section V, we draw key conclusions.

Notation: $\mathbb{E}[\cdot]$ and $\|\cdot\|_F$ denote the expectation and the Frobenius norm of a matrix, respectively. The operators $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and complex conjugate (Hermitian) transpose, respectively. $\boldsymbol{I}_M$, $\mathbb{R}^{m\times n}$, $\mathbb{C}^{m\times n}$ denote the $M\times M$ identity matrix, the real and complex spaces of dimension $m\times n$, respectively. $\mathrm{diag}(x_1, x_2, \ldots, x_L)$ denotes an $L\times L$ diagonal matrix with diagonal elements given by $x_1, x_2, \ldots, x_L$. $\boldsymbol{x}\sim\mathcal{CN}(0,\boldsymbol{\Sigma})$ means that $\boldsymbol{x}$ is a circularly symmetric complex Gaussian (CSCG) random vector with zero mean and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model

As illustrated in Fig. 1, we consider a single-cell massive MIMO network with a BS equipped with a large antenna array comprising M antennas and serving a set $\mathcal{K}$ of K single-antenna (legitimate) UEs in the presence of a single-antenna active ED. The system employs a TDD transmission mode and exploits the reciprocity of the uplink and downlink channels. During the uplink channel estimation phase, all UEs simultaneously transmit mutually orthogonal pilot signals to the BS. Let $\boldsymbol{\phi}_i\in\mathbb{C}^{L\times 1}$ be the pilot signal of UE $i\in\mathcal{K}$, where L is the pilot signal length. Since the pilot signals are mutually orthogonal, we have $\boldsymbol{\phi}_i^H\boldsymbol{\phi}_j = L$ if $i=j$ and $\boldsymbol{\phi}_i^H\boldsymbol{\phi}_j = 0$ if $i\neq j$. The ED is assumed to have full knowledge of the pilot signal used by the UE under attack (targeted UE). Thus, during the uplink channel estimation phase, the ED transmits

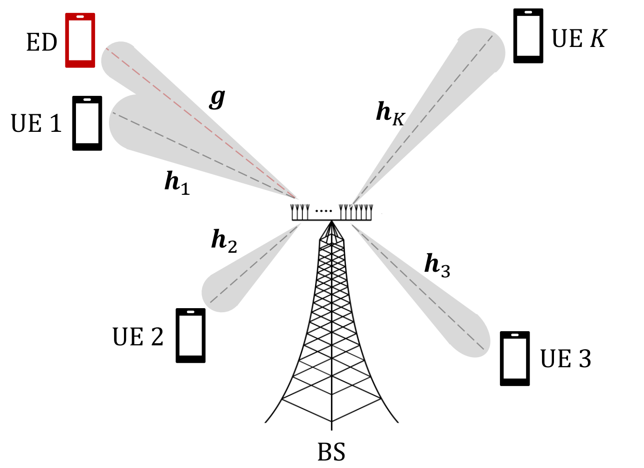

Fig. 1. An illustration of the proposed multiuser single-cell massive MIMO network with active eavesdroppers performing pilot spoofing attacks.

the same pilot signal as that of the attacked UE. As a result, the BS estimates the uplink composite channel of the attacked UE and the ED. Consequently, in the downlink transmission, a portion of the signal intended for the attacked UE is leaked to the ED, without the knowledge of the BS. We assume that if there is a pilot spoofing attack, only one UE is attacked[1].

We assume that the channels in the system exhibit block fading, i.e., they remain constant over a time block and change independently from one block to another. We define $\boldsymbol{h}_i \in \mathbb{C}^{M\times 1}$ as the channel between UE i and the BS and $\boldsymbol{g} \in \mathbb{C}^{M\times 1}$ as the channel between the ED and the BS. We assume that all the channels are Rayleigh flat-fading channels, i.e.,

$$\boldsymbol{h}_i = \sqrt{\beta_i}\ \tilde{\boldsymbol{h}}_i,\ i \in \mathcal{K},\ \text{and}\ \boldsymbol{g} = \sqrt{\beta_{\text{ED}}}\ \tilde{\boldsymbol{g}}, \tag{1}$$

where β_i and β_{ED} represent, respectively, the large-scale fading coefficients of the channels between UE i and the BS and between the ED and the BS. These coefficients are assumed to remain constant throughout the communication and only β_i is known at the BS. $\tilde{\boldsymbol{h}}_i \in \mathbb{C}^{M\times 1},\ i \in \mathcal{K}$, and $\tilde{\boldsymbol{g}} \in \mathbb{C}^{M\times 1}$ are vectors with elements modeled as $\mathcal{CN}(0,1)$ and are assumed to be unknown at the BS.

Prior works [6]–[16], [24], [25] on pilot spoofing attacks in massive MIMO networks assume the ED to always be perfectly synchronized with the BS. However, it could be difficult for the ED to achieve perfect timing and frequency synchronization in practice. Particularly, unlike UEs that use timing advance to achieve timing synchronization with the BS, the ED is an illegitimate user that cannot participate in this synchronization phase and must only rely on overhearing the exchanges between the BS and targeted UE to estimate when to start transmitting. For this reason, the ED timing synchronization might not be perfect. Hence, hereinafter, we consider that the ED will have a timing mismatch of $\tau = eT_s$, $e \in [0, 1)$, where T_s is the symbol period. Also, because the ED is an illegitimate user, it is practical to assume that there is a frequency mismatch between the ED and the BS local oscillators since it is difficult to synchronize them. To accurately model the timing synchronization mismatch of the ED, we adopt the discrete-time model of the imperfect timing synchronization of pilot contamination that was developed in [26]. In particular, for a timing mismatch of τ and a normalized frequency offset of Δf between the ED and BS and assuming that the ED attacks the kth UE[2], the received signal at the BS during the uplink channel estimation phase, $\boldsymbol{Y} \in \mathbb{C}^{M\times L}$, is given by

$$\boldsymbol{Y} = \sum_{i=1}^{K} \sqrt{P}\ \boldsymbol{h}_i \boldsymbol{\phi}_i^T + \alpha\sqrt{P_{\text{ED}}}\ \boldsymbol{g}\boldsymbol{\phi}_k^T \boldsymbol{F}^T \boldsymbol{U} + \boldsymbol{N}, \tag{2}$$

where P and P_{ED} are the transmit powers of the UEs and the ED during the uplink channel estimation phase, respectively. $\alpha = 1$ indicates the presence of the ED and $\alpha = 0$ its absence. The imperfect timing synchronization of the ED is represented by $\boldsymbol{F} \in \mathbb{C}^{L\times L}$, which is a circulant matrix and its ith row is the cyclic shift of the row vector $\boldsymbol{f}_\tau = [f_{\tau,0}, \ldots, f_{\tau,-L_1}, f_{\tau,L_2}, \ldots, f_{\tau,1}]$ to the right by i positions, with $L = L_1 + L_2 + 1$ [26]. $f_{\tau,i} = F(iT_s - \tau)$, where $F(iT_s - \tau)$ is the baseband-equivalent impulse response of the convolution of the transmit pulse shaping filter and the corresponding matched receive filter with a timing mismatch of $\tau = eT_s$. We note that in the case of perfect timing synchronization, i.e., $\tau = 0$, we have $F(0) = 1$ and $F(iT_s) = 0$, $\forall i \neq 0$. Therefore, it is clear that for perfect timing synchronization, i.e., $\tau = 0$, $\boldsymbol{F} = \boldsymbol{I}_L$ which results in $\boldsymbol{\phi}_k^T \boldsymbol{F}_j^T = \boldsymbol{\phi}_k^T$. The imperfect frequency synchronization of the ED with the BS is represented by $\boldsymbol{U} \in \mathbb{C}^{L\times L} = \text{diag}(e^{-j2\pi\Delta f}, \ldots, e^{-j2\pi L\Delta f})$. $\boldsymbol{N} \in \mathbb{C}^{M\times L}$ is the additive noise matrix with entries modeled as circularly symmetric complex Gaussian random variables with zero mean and variance σ_n^2, i.e., $\mathcal{CN}(0, \sigma_n^2)$. The following proposition shows that the timing and/or frequency asynchronicity of the ED with the system increases the subspace dimension of the received signal at the BS by one.

Proposition 1. *The subspace dimension of the received signal at the BS is K and $K+1$ in the absence and presence of the asynchronous pilot spoofing attack (i.e., the ED has a timing and/or frequency mismatch with the BS), respectively. This is unlike the synchronous attack where the signal subspace dimension is always K regardless of the absence or presence of the ED.*

Proof. Let $\boldsymbol{H} = [\sqrt{P}\boldsymbol{h}_1, \ldots, \sqrt{P}\boldsymbol{h}_K, \alpha\sqrt{P_{\text{ED}}}\boldsymbol{g}] \in \mathbb{C}^{M\times(K+1)}$, $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K,\ \overline{\boldsymbol{\phi}}_k]^T \in \mathbb{C}^{(K+1)\times L}$, where $\overline{\boldsymbol{\phi}}_k = \boldsymbol{U}\boldsymbol{F}\boldsymbol{\phi}_k$. Therefore, the received signal at the BS in (2) can be recast as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{\Phi} + \boldsymbol{N}. \tag{3}$$

The covariance matrix of the received signal $\boldsymbol{Y}$ can then be obtained as

$$\boldsymbol{R} = \frac{1}{L}\mathbb{E}[\boldsymbol{Y}\boldsymbol{Y}^H] = \frac{1}{L}\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{\Phi}^H\boldsymbol{H}^H + \sigma_n^2\boldsymbol{I}_N = \boldsymbol{R}_s + \sigma_n^2\boldsymbol{I}_N, \tag{4}$$

[1]The case of multiple EDs attacking a single or multiple UEs is beyond the scope of this work and is left for future research.

[2]This means that the ED uses the same pilot signal as that of the attacked user, UE k, i.e., $\boldsymbol{\phi}_{\text{ED}} = \boldsymbol{\phi}_k$.

where $\boldsymbol{R}_s = \frac{1}{L}\boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{\Phi}^H\boldsymbol{H}^H$. Since the pilot signals ϕ_i are mutually orthogonal and for $e \neq 0$ and/or $\Delta f \neq 0$, $\overline{\phi}_k \neq \phi_k$, then $\mathrm{rank}(\boldsymbol{\Phi}\boldsymbol{\Phi}^H) = K+1$ given that $L \geq K+1$. Also, as all the channels in the system are independent and $N \gg K+1$, hence $\mathrm{rank}(\boldsymbol{H}) = K+1$ with probability 1. Consequently, it is clear that $\mathrm{rank}(\boldsymbol{R}_s) = K+1$. So, clearly the imperfect timing/frequency synchronization of the ED with the system led to a subspace dimension of $K+1$ unlike the case where the ED is perfectly synchronized where the subspace dimension is K. □

Proposition 1 means that the covariance matrix of the received signal could be exploited to determine the presence or absence of the ED when it is not perfectly synchronized with the BS. In particular, the well known MDL method can be used to estimate the dimension of the signal subspace d using the eigenvalues of the sample covariance matrix. If $d = K$ then there is no ED and if $d \geq K+1$ then an ED is present.

In the next section, in contrast to prior works, we propose a deep learning (DL) approach to detect and identify pilot spoofing attacks in massive MIMO networks when the ED is not perfectly synchronized with the system.

III. Attack Detection and Identification

In this section, we propose a DNN, called IDNet, to detect the presence of pilot spoofing attacks and identify the attacked UEs. We then describe the key features included in the feature vector.

We propose to formulate the detection of an attack and identification of the attacked UE as a multi-class classification problem, where the output of the DNN indicates which UE are under attack (or if no attack is present). Identifying the attacked UEs could be paramount as, for example, it enables the BS to selectively terminate/pause transmission to the attacked UEs, minimizing unnecessary network downtime and disruption for the unaffected UEs. We adopt the multilayer perceptron (MLP) neural network architecture as it can learn complex non-linear relationships [27], [28] and can be generally applied to any classification problem if the feature is not too complex.[3].

The proposed IDNet consists of three segments: an input layer, a number of hidden layers, and an output layer. The input layer simply has the same dimension as that of the feature vector. The number of hidden layers and units per each hidden layer are determined from preliminary testing to observe what minimizes the training and validation losses to prevent over-fitting. We use four hidden layers with NM units per layer, where N is the number of frames. The rectified linear unit (ReLU) is used as the activation function for each hidden layer due to its versatility across many classification tasks [29]. Each hidden layer is preceded by a dropout and batch normalization (BatchNorm) layer to minimize over-fitting and provide stable optimization, respectively. An early-stopping callback is used to cease training when there is no or little improvement in the validation loss after a certain number of training epochs. Finally, the output layer consists of $K+1$ units, i.e., $K+1$ classes, with the first K elements indicating the presence/absence of an attack for each UE and the final unit indicating that no attack is present. A softmax activation function is used to normalize the output from the weighted sum values into a vector of probabilities that sum up to one [29], with the element with the highest probability being chosen as the predicted class. It must be noted that, intuitively, IDNet can also be used for the detection of an attack as in the case of an attack, if any of the first K classes are predicted it still correctly identifies there is an ED present even if the predicted attacked UE is incorrect. The architecture of the proposed IDNet is given in Fig. 2 and Table I. The IDNet is created using the sequential application programming interface (API) to create a Keras model with TensorFlow 2.12.

Layer	# Units	Input Shape	Activation
Dense	NM	$NM+K$	ReLU
Dropout	-	NM	-
BatchNorm	-	NM	-
Dense	NM	NM	ReLU
Dropout	-	NM	-
BatchNorm	-	NM	-
Dense	NM	NM	ReLU
Dropout	-	NM	-
BatchNorm	-	NM	-
Dense	NM	NM	ReLU
Dropout	-	NM	-
BatchNorm	-	NM	-
Dense	$K+1$	$K+1$	Softmax

TABLE I
IDNet structure.

The proposed IDNet is trained using datasets containing the vectors of features and their associated labels. Feature selection is a crucial step in designing any DL algorithm as it significantly impacts its performance. For our detection and identification problem, the features are derived from the uplink channel estimation phase, namely, the received signal in (3). To reduce the effect of fading, we propose to use $N \geq 1$ frames for the calculation of the features. Let $\boldsymbol{Y}^{(j)}$ be the received pilot signal at the BS for the jth frame. Hereinafter, we will describe the features that will be used to train and test the IDNet.

1) Average Power: One of the main effects of the presence of the ED is the increase in the power of the received signal. Thus, it is natural to use the average power (AP) of the received signal, i.e., $p = \frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{Y}^{(j)}\|_F^2$, as a feature.

2) Projected Average Power: Although the average power feature, p, could be useful for the detection of the ED, it does not help identify the attacked UE. For this reason, we propose to isolate the power coming from each UE and cancel out the contributions from the remaining UEs. To this end, we project the received signal at the BS onto all the K pilot signals and use the average power of the resulting signal, called projected average power (PAP), as a feature. Let $\boldsymbol{y}_i^{(j)} = \boldsymbol{Y}^{(j)}\boldsymbol{\phi}_i^{\star}$, $i \in \mathcal{K}$, be the projection of the received pilot signal corresponding to

[3]Note that we also considered similar NN architectures to IDNet, such as CNNs, and more complex NN architectures such as ResNet-50, AlexNet, and VGG but they did not provide any significant improvement in identification accuracy during preliminary testing. This shows that our problem does not require complex NN architectures and a simpler architecture, such as the proposed IDNet, is sufficient to achieve good performance.

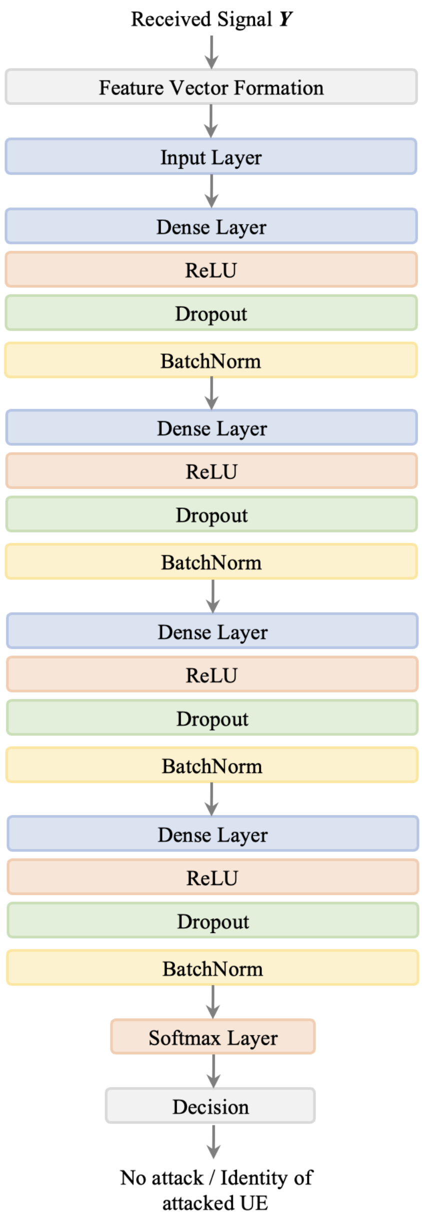

Fig. 2. Architecture of the proposed IDNet.

the jth frame onto the ith pilot signal. The average power of the resulting signal over N frames is: $p_i = \frac{1}{N}\sum_{j=1}^{N} \|\boldsymbol{y}_i^{(j)}\|^2$, $i \in \mathcal{K}$.

3) Power Ratio: Inspired by the detection methods in [8] and [9], we propose to use the ratio of the instantaneous received power and the average power at the BS in the absence of the ED as a feature as it has the potential to reveal the presence of the ED. In particular, in the absence of the ED, the power ratio is expected to be close to one, but in the case of an attack, this ratio is expected to be larger than one due to the added power from the ED. This defines a threshold above which an attack is present for the DNN to learn. The average power of the received pilot signal in the absence of an ED, i.e., $\alpha = 0$ in (3), can easily be obtained as

$$\overline{\mathcal{P}} \triangleq \mathbb{E}\left[\|\mathbf{Y}\|_F^2\right] = PML\sum_{i=1}^{K}\beta_i + ML\sigma_n^2. \tag{5}$$

The power ratio (PR) is then given by

$$\gamma = \frac{\sum_{j=1}^{N}\|\mathbf{Y}^{(j)}\|_F^2}{N\overline{\mathcal{P}}}. \tag{6}$$

4) Projected Power Ratio: In order to be able to identify the attacked UE, we propose to use the power ratio for each UE. The ith power ratio, called the projected power ratio (PPR), is defined as

$$\gamma_i = \frac{\frac{1}{N}\sum_{j=1}^{N}\|\boldsymbol{y}_i^{(j)}\|^2}{\mathbb{E}\left[\|\boldsymbol{y}_i^{(j)}\|^2\right]} = \frac{\sum_{j=1}^{N}\|\boldsymbol{y}_i^{(j)}\|^2}{N(PML^2\beta_i + ML\sigma_n^2)}. \tag{7}$$

From (6) and (7), the BS must have knowledge of the large-scale fading coefficients for each UE prior to pilot training, which is reasonable to assume in a massive MIMO system[4].

5) Eigenvalues: As stated in Proposition 1, the asynchronous attack increases the rank of the covariance matrix of the signal part of the received signal from K to $K+1$. Since the rank and the eigenvalues of a covariance matrix are closely related, it is hence natural to use the eigenvalues of the covariance matrix of the received signal as a feature. The estimate of the covariance matrix of the received signal in the jth frame, $\boldsymbol{Y}^{(j)}$, is defined as

$$\widehat{\boldsymbol{R}}^{(j)} = \frac{1}{L}\mathbf{Y}^{(j)}\mathbf{Y}^{(j)H} = \boldsymbol{U}^{(j)}\boldsymbol{\Lambda}^{(j)}\boldsymbol{U}^{(j)H}, \tag{8}$$

where $\boldsymbol{\Lambda}^{(j)}$ is the diagonal matrix containing the eigenvalues of $\widehat{\boldsymbol{R}}^{(j)}$ and $\boldsymbol{U}^{(j)}$ is the matrix containing the corresponding eigenvectors. Let the real ordered eigenvalues of $\widehat{\boldsymbol{R}}^{(j)}$ be denoted as $\lambda_1^{(j)} \geq \lambda_2^{(j)} \geq \ldots \geq \lambda_M^{(j)}$. Hence, the eigenvalues (EVs) $\lambda_i^{(j)}$, $i = 1, \ldots, M$, $j = 1, \ldots, N$ can be used as features in IDNet.

The potentially attacked UE is assumed to be unknown; therefore, the eigenvalues alone do not provide enough information about the UE being attacked. However, when an ED attacks the targeted UE, the PAPs and PPRs corresponding to the attacked UE will be significantly higher compared to those of the other non-attacked UEs, due to the added power from the ED. This enables a clear identification of the UE being attacked by the ED.

The complexity of IDNet in terms of the number of multiplications is $L^2(M+L) + LMK + A(M+K) + A^2(B-1) + A(K+1)$, where A is the number of neurons per hidden layer and B is the number of hidden layers. Note that the main complexity of IDNet is due to the calculation of the eigenvalues of the received signal covariance matrix that are used in the FV. Although IDNet has higher complexity compared to conventional approaches, it provides much superior performance than them, as shown in Fig. 7 in the next section.

The impact of each of the above features (or a combination of them) on the detection and identification accuracy of IDNet will be discussed in the next section.

[4]The large-scale fading coefficients are assumed to stay constant for long coherence blocks, therefore the BS can accurately estimate them [1].

IV. Simulation Results and Analysis

In this section, we evaluate the performance of the proposed IDNet. We will first describe how the dataset that is used to train and test IDNet is generated. Then, we discuss how IDNet is trained. Finally, we present comprehensive results about the performance of IDNet for different system parameters.

A. Dataset Generation

In order for IDNet to detect deviations from expected behavior, it must be trained on a sufficiently large dataset to learn the expected received signal characteristics for when an active ED is present or not. This enables the trained DNN to classify new unseen data as either normal or abnormal. As such, given the lack of datasets built using real measurements at this stage, we generate a synthetic dataset for different system parameters such as $M, K, L, D, \Delta f$, and N to observe how they affect the detection/identification accuracies.

To build the dataset, we randomly generate the locations of K UEs and the ED within a square cell area, with the BS located at the center. The attacked UE is chosen as the closest UE to the ED. Using the locations of the UEs and ED, the large-scale fading coefficients, β_i and β_{ED}, are calculated according to the 3GPP channel model in [30] to mimic a non-line-of-sight (NLoS) urban-micro street canyon environment. For each randomly generated locations of the UEs and ED, we generate a very large number (thousands) of realisations of the UEs and ED channels to simulate the small-scale fading characteristics of the channels, $\tilde{\boldsymbol{h}}_i$ and $\tilde{\boldsymbol{g}}$, where each realisation follows a Rayleigh distribution to reflect the Rayleigh fading channel model adopted in (1). This allows us to capture a wide range of fading scenarios. This helps in training the IDNet model to be robust against various fading conditions and generalize to different fading scenarios during deployment. This is then repeated for other randomly generated UEs and ED locations, with the aim of averaging the channel realizations over the entire coverage area. During each training phase, a random timing mismatch of $\tau = eT_s$, $e \in [0, 1)$ is generated. The received samples used to build the dataset are calculated based on (3).

Channel realizations are also generated without an ED present, with a 50:50 split to ensure no class imbalance in the dataset would cause poor performance in the minority class and over-fitting during training. Each channel realization for each location is associated with a one-hot encoded vector to indicate the unique UE label (ID) being attacked or no attack at all. Finally, the generated samples are randomly shuffled within the dataset to obtain a uniform spread of data for model training, validation, and testing.

B. Model Training

Prior to model training, the input features are scaled to have zero mean and unit variance. Preliminary testing is conducted to determine suitable hyperparameters using grid search for model training and are shown in Table II. The train-test-validation partitions of the datasets are split into a 75:15:10 ratio. We use the adaptive moment estimation (Adam) optimizer, which is a popular stochastic gradient descent-based optimizer that has an adaptive learning rate. The loss function used is the categorical cross-entropy (CE) function. The number of training epochs is determined using an early-stopping callback to detect when there is no appreciable improvement in identification accuracy.

Hyperparameter	IDNet
Loss Function	Categorical CE
Optimiser	Adam
epochs	30
Batch Size	200
minimum delta	0.001
Patience	5

TABLE II
Training Hyperparameters for IDNet

C. Results

Unless otherwise stated, the considered system parameters are set as follows: $M = 100$, $L = 16$, $K = 4$, $N = 1$, and $P = 20$ dBm. The communication bandwidth is $W = 20$ MHz and the noise power is $\sigma_n^2 = W\kappa_b T_0 N_F$. $N_F = 5$ dB is the noise figure, $T_0 = 290$ Kelvin is the thermal noise temperature, and $\kappa_b = 1.381 \times 10^{-23}$ Joules/Kelvin is the Boltzmann constant. Different ED powers are investigated to see the impact on the detection/identification accuracy, as the ED can smartly adapt its transmission power to remain undetected [16]. The ED's transmit power, P_{ED}, was swept between a range of -10 dBm and 30 dBm, and eight possible timing mismatch values are used, i.e., $\tau = eT_s \in \{0, \frac{T_s}{8}, \frac{T_s}{4}, \frac{3T_s}{8} ..., \frac{7T_s}{8}\}$. For notational convenience, the timing mismatch can also be written as a factor of the symbol delay called the delay factor, $D = \frac{8\tau}{T_s}$ i.e., $D \in \{0, 1, ..., 7\}$. Finally, the UEs and ED are uniformly randomly distributed within a square area of 200×200 m^2. We adopt the detection/identification accuracy metric to evaluate the performance of IDNet[5].

We first investigate the impact of the size of the dataset on the identification accuracy to identify where the performance saturates while also having a practical dataset size. Note that collecting real data to build a dataset consumes a lot of resources (time, financial and human) and determining a suitable dataset size will help optimize this paramount task. The size of the dataset is influenced by the number of random locations, n_{loc}, and the number of channel realizations, n_s, generated per ED power. In Fig. 3, we examine the synchronous case and show the identification accuracy versus the ED power for IDNet. Different combinations of n_{loc} and n_s are used to observe how many locations and samples are sufficient to average over the 200×200 m^2 square cell. For $n_s = 1000$, when increasing n_{loc} from 50 to 100, we observe a maximum increase in identification accuracy of 9% at $P_{\text{ED}} = 20$ dBm. For $n_{\text{loc}} = 50$, when increasing n_s from

[5]Note that other performance metrics could be used such as the detection/identification probability. However, detection/identification accuracy is preferred as, unlike detection/identification probability which only captures true positives (TPs), it encapsulates both the TPs and true negatives (TNs). This is essential since incorrectly classifying cases with no attack present as attack present would increase unnecessary downtime/disruption in the network.

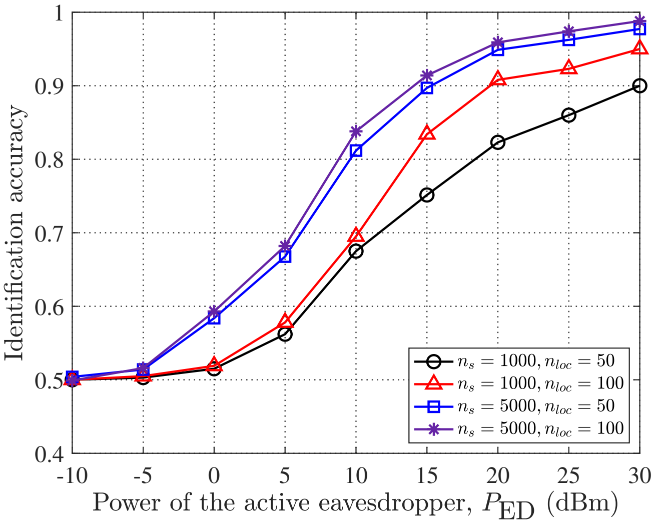

Fig. 3. IDNet identification accuracy versus ED transmit power for different dataset sizes, for synchronous ED pilot transmission (i.e., $D = 0$ and $\Delta f = 0$). $M = 100$, $L = 16$, $K = 4$, and $N = 1$.

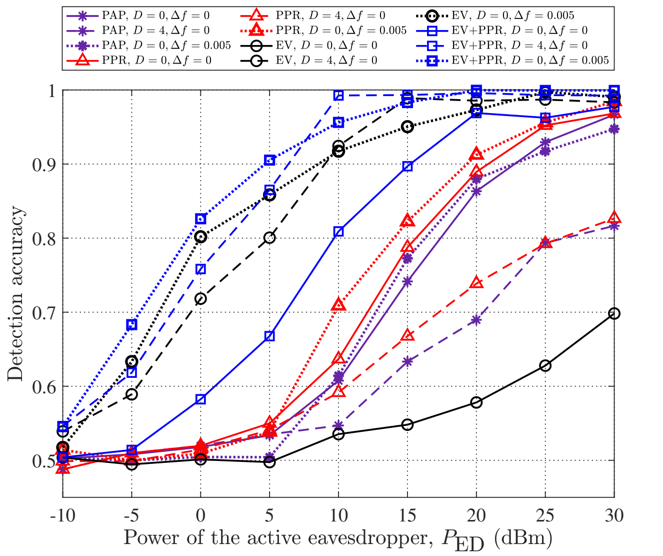

Fig. 4. IDNet detection accuracy versus ED transmit power for different feature vectors. $M = 100, L = 16, K = 4, N = 1$.

1000 and 5000 a maximum increase in identification accuracy of 15% is observed at $P_{\text{ED}} = 15$ dBm. As noted by [31], DL algorithms can exhibit improved performance as the size of the dataset increases, as this enables better generalization of the data. Using more samples has a greater impact on accuracy compared to the number of locations as sampling 5000 times over 100 locations only provided at most an increase of 2% in identification accuracy. Nonetheless, this improvement comes at the cost of a lengthier training time. To balance the size of the dataset and constraints of the model training time available, subsequent datasets that were generated used a Monte-Carlo simulation of 5000 realizations over 50 random locations to sufficiently average over the coverage area.

Fig. 4 shows the detection accuracy of IDNet versus the ED

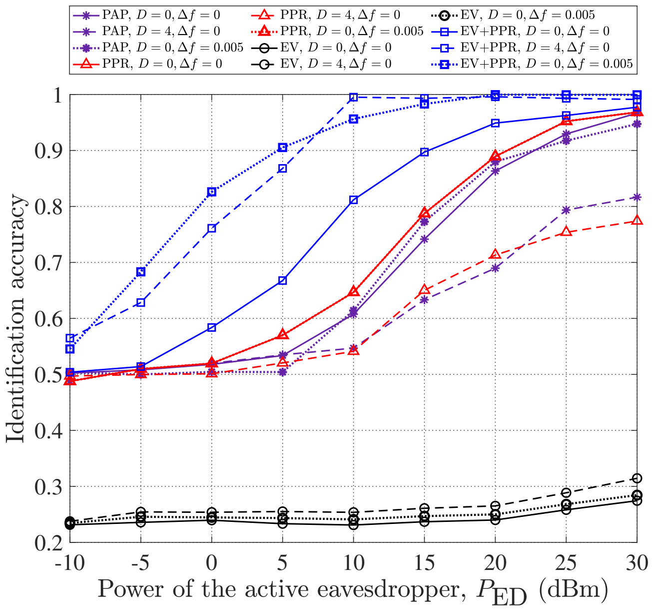

Fig. 5. IDNet identification accuracy versus ED transmit power for different feature vectors. $M = 100, L = 16, K = 4, N = 1$.

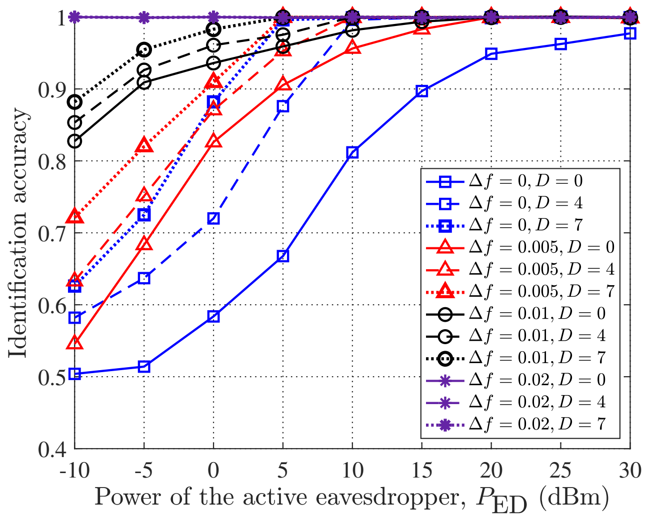

Fig. 6. IDNet identification accuracy versus ED transmit power for different combinations of the delay factor, D, and the frequency offset, Δf. $M = 100, L = 16, K = 4$, and $N = 1$.

power for four different cases of feature vectors representing varying degrees of complexity. These four cases are: PAPs, PPRs, EVs, and EVs and PPRs[6]. For each one of these feature vectors we consider both synchronous (i.e., $D = 0$ and $\Delta f = 0$) and asynchronous (i.e., only a timing mismatch of $D = 4$ or only a frequency mismatch of $\Delta f = 0.005$) pilot spoofing attacks. For the synchronous case ($D = 0$ and $\Delta f = 0$), we can see that the feature vector consisting of

[6]The AP and PR features provided the lowest performance and were not included to keep the figure clear. The power only based features are still represented by PAPs and PPRs, which performed much better than AP and PR.

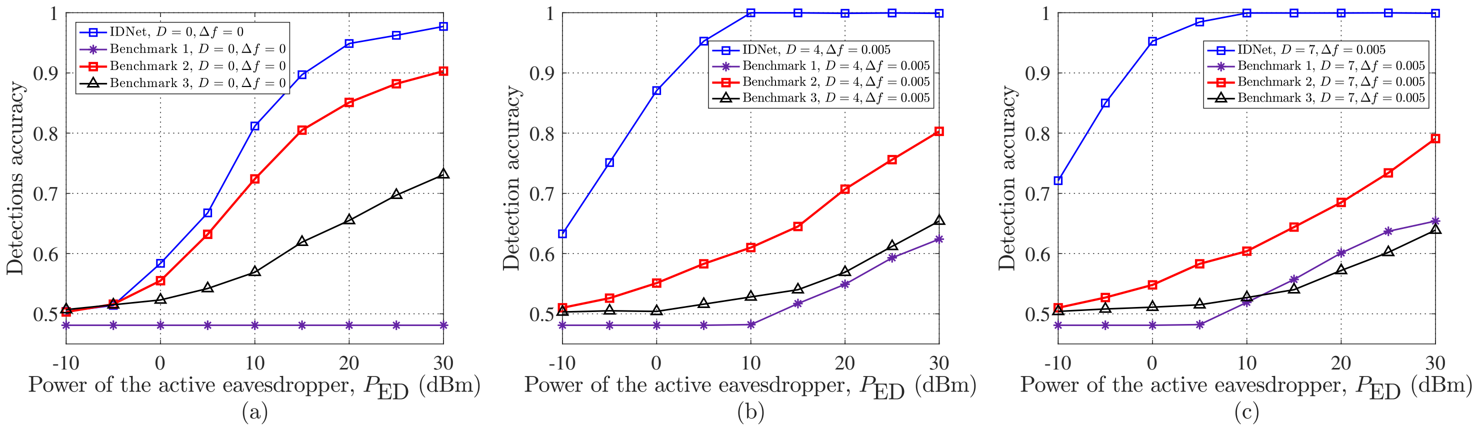

Fig. 7. Comparison of IDNet and other methods in terms of detection accuracy versus ED transmit power. (a) shows the synchronous case, i.e., $D = 0$ and $\Delta f = 0$, (b) shows the asynchronous case where $D = 4$ and $\Delta f = 0.005$, and (c) shows the asynchronous case where $D = 7$ and $\Delta f = 0.005$. $M = 100$, $L = 16$, $K = 4$, and $N = 1$.

EVs and PPRs provides the best detection accuracy across all ED powers. We can also observe that the feature vector consisting only of PPRs performs better than that consisting only of EVs. This is due to the fact that, when the ED pilot signal is synchronous with that of the targeted UE, the signal subspace dimension is the same as that of no attack case. The real benefit of the EVs is shown in the asynchronous cases, i.e., when $D = 4$ or $\Delta f = 0.005$, where using EVs significantly outperforms the energy-only based feature vectors, i.e., PAPs and PPRs. The detection accuracy increases when using the EVs as the timing/frequency mismatch in the ED pilot transmission causes the signal to become less correlated with the attacked UE pilot signal when sampled at the BS. The low correlation between these two pilot signals effectively creates an extra source signal (i.e., increases the signal subspace by one) that the IDNet can distinguish. It should be noted that in the case of only timing mismatch (i.e., $D = 4$ and $\Delta f = 0$), unlike in the case of frequency mismatch (i.e., $\Delta f = 0.005$ and $D = 0$), using only PAPs or PPRs decreases the detection accuracy. This is because as the timing mismatch increases, unlike frequency mismatch, the ED pilot signal becomes less correlated with the attacked UE, so less energy is recovered from the projection of the received signal onto the attacked UE pilot signal. EVs provide insights into the overall signal structure and dimensionality of the signal subspace, while PPRs offer more granular information about signal power distribution. Leveraging these synergistic features, allows to achieve the best detection accuracy for both synchronous and asynchronous cases.

Fig. 5 shows the identification accuracy of IDNet versus the ED power for the same feature vectors above for both the synchronous and asynchronous cases. Similar to the results in Fig. 4, we observe that the feature vector consisting of the EVs and PPRs provides the best identification accuracy. We can also clearly see that using only the EVs does not provide any identification capability. This can be explained by the fact that the EVs by themselves do not provide any indication of which UE is attacked, whereas projecting the received signal onto each training sequence can help isolate the power of the attacked UE and ED. From Figs. 4 and 5 it is apparent that the feature vector containing both the EVs and PPRs outperforms all the other feature vectors for both synchronous and asynchronous attacks. Therefore, the feature vector containing both the EVs and PPRs will be used in the rest of this section.

Fig. 6 shows the identification accuracy of IDNet versus the ED power for different combinations of the delay factor, D, and the frequency offset, Δf. We can clearly see that the identification accuracy improves with increasing timing and/or frequency mismatches. Specifically, Fig. 6 shows that a 100% identification accuracy is reached for $\Delta f = 0.02$ for any timing mismatch for an ED power of -10 dBm. So, higher timing and/or frequency mismatches enhances the identification accuracy.

Fig. 7 compares IDNet with three conventional methods, used as benchmarks, in terms of detection accuracy versus the ED power. The three benchmarks are: 1) Benchmark 1: the well known MDL method [13][7], 2) Benchmark 2: the method in [12], and 3) Benchmark 3: the method in [11]. For comparison fairness, when implementing the MDL-based detector, no power budget was allocated to a random signal superimposed onto the pilot signal transmitted from the attacked UE. We can see that the proposed IDNet significantly outperforms all the benchmarks for both the synchronous (in Fig. 7(a)) and asynchronous (in Figs. 7(b) and (c)) cases. It is observed that the gap between IDNet and Benchmarks 2 and 3 increases with increasing time delay factor D. In particular, for the synchronous case of ($D = 0$, $\Delta f = 0$) in Fig. 7(a), IDNet outperforms all the benchmarks across the entire power range and that the detection accuracy gap increases with P_{ED}. It is worth noticing that MDL has no capability of accurately detecting an ED in the synchronous case. This is expected as MDL cannot distinguish between the attacked UE and ED unless a random signal is superimposed onto the pilot signal transmitted from the attacked UE as

[7]It is important to note that MDL cannot identify the attacked UE and can only detect whether there is an ED.

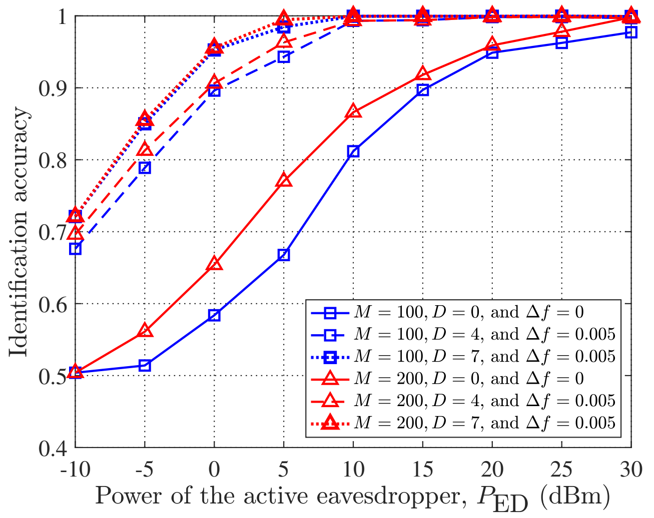

Fig. 8. IDNet identification accuracy versus ED transmit power for different combinations of the number of BS antennas, M, the delay factor, D, and the frequency offset, Δf. $L = 16, K = 4$, and $N = 1$.

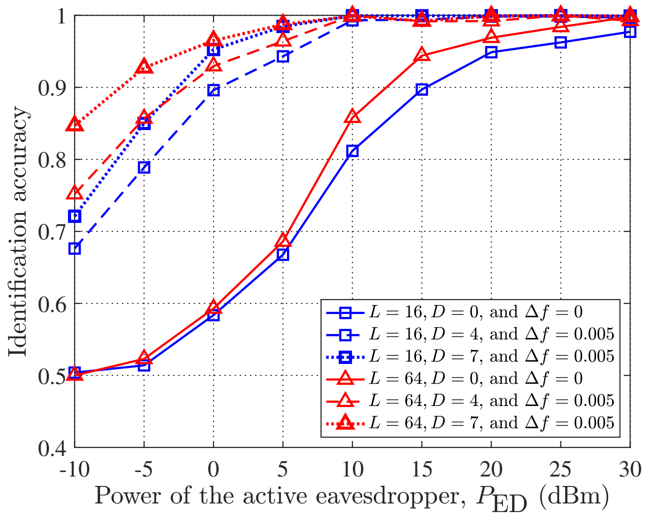

Fig. 9. IDNet identification accuracy versus ED transmit power for different combinations of the pilot signal length, L, the delay factor, D, and the frequency offset, Δf. $M = 100, K = 4$, and $N = 1$.

in [13]. The average detection accuracy of 50% for MDL implies that it predicts no ED present for most trials, implying a high false negative rate. Unlike Benchmarks 2 and 3, the performance of MDL[8] improves in the asynchronous cases of $(D = 4, \Delta f = 0.005)$ and $(D = 7, \Delta f = 0.005)$ but IDNet still significantly outperforms it. Also, from Figs. 7(b) and (c), we observe that the detection accuracy of IDNet improves in the asynchronous cases but, on the contrary, the detection accuracies of Benchmarks 2 and 3 degrade. Therefore, this clearly shows the advantage of IDNet over the benchmarks especially for asynchronous attacks. Since it is difficult for the ED to achieve perfect time/frequency synchronization, IDNet is therefore a more compelling choice in the more realistic scenario of asynchronous PSA attacks. It should be noted that the outstanding performance of IDNet comes at an increased complexity cost compared to the benchmarks.

Next, in Fig. 8 we investigate the impact of the number of antennas, M, at the BS on the identification accuracy of IDNet for different combinations of the delay factor, D, and the frequency offset, Δf. It is notable that the number of antennas at the BS has a clear impact on the identification accuracy. For example, in the synchronous case $(D = 0, \Delta f = 0)$, increasing M from 100 to 200 leads to 10.3% increase in the identification accuracy at $P_{\text{ED}} = 5$ dBm. This is because larger antenna arrays improve the spatial diversity and hence reduce the impact of fading. This benefit of using more antennas at the BS is less apparent for $P_{\text{ED}} < -5$ dBm and $P_{\text{ED}} \geq 15$ dBm. For the asynchronous cases of $D \in \{4, 7\}$ and $\Delta f = 0.005$, we also observe a slight increase in the identification accuracy by increasing the number of antennas at the BS even at $P_{\text{ED}} = -10$ dBm. However, increasing the number of antennas for $(D = 7, \Delta f = 0.005)$ resulted in negligible identification accuracy improvement.

[8]Note that according to Proposition 1 the signal subspace dimension increases by one in the presence of the asynchronous ED and hence MDL can be used to estimate the signal subspace dimension and consequently the detection of the ED.

Fig. 9 shows the impact of the pilot signal length, L, on the identification accuracy of IDNet for different combinations of the delay factor, D, and the frequency offset Δf. We can see that using a longer pilot signal does enhance the accuracy in all cases. However, the largest gains can be seen in the asynchronous cases at low ED powers, e.g., at -10 dBm. This is because longer pilot signals allow the BS to average the noise but more importantly they allow to obtain a better estimate of the sample covariance matrix and consequently more accurate eigenvalues that are used as input features. Note that increasing both M and L in tandem can help increase the identification accuracy, whilst also providing other benefits such as reducing the effect of pilot contamination and improving the accuracy of the CSI.

Fig. 10 investigates the impact of the number of frames on the identification accuracy. We consider three cases, namely $N = 1$, 3, and 5. We can clearly see that the identification accuracy increases with the increase of the number of frames. It is worth noticing that for $D = 4$ and $D = 7$, there is a noticeable improvement in identification accuracy even for low ED powers confirming that it is easier to detect and identify asynchronous attacks even if the ED power is very low. This can be justified as follows:

- The PPRs are averaged over the set of available frames, providing a more accurate value that is more robust against the effect of noise and channel fluctuations.
- By using multiple frames, the number of eigenvalues available to learn from increases, thereby suppressing the impact of noise if one or more frames become more corrupted with noise, compared to other frames.

We can also observe that for $D = 4$ and 7, increasing N from 3 to 5 leads to a smaller gain compared to increasing N from 1 to 3. So, it is expected that increasing N beyond 5 will result in increased complexity with diminishing accuracy improvement.

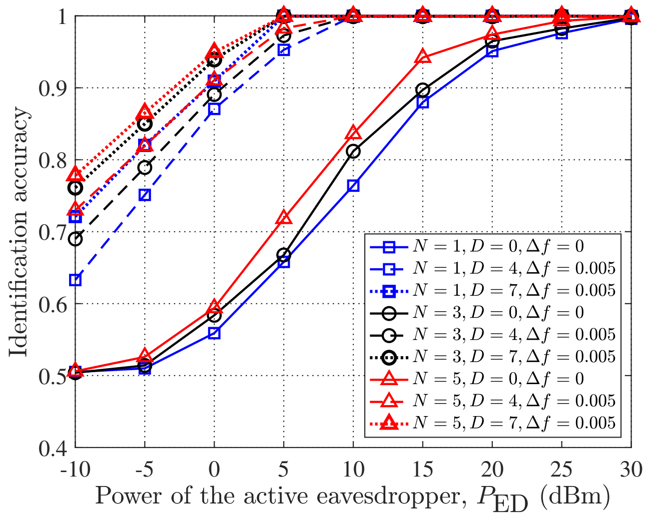

Fig. 10. IDNet identification accuracy versus ED transmit power for different combinations of the number of frames, N, and the delay factor, D. $M = 100$, $L = 16$, and $K = 4$.

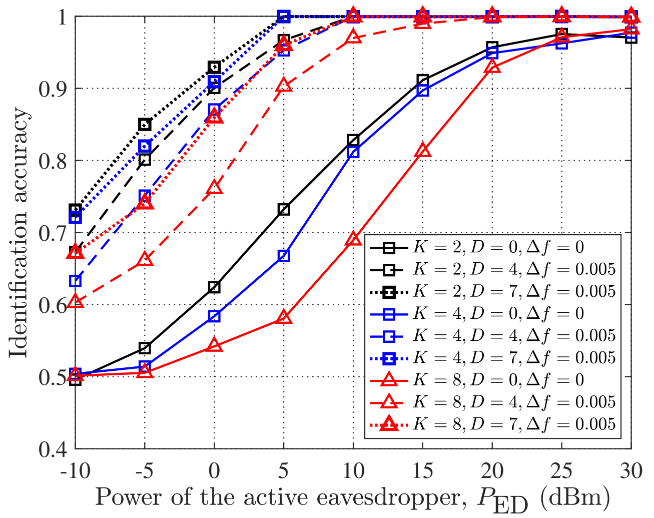

Fig. 11. IDNet identification accuracy versus ED transmit power for different combinations of the number of UEs, K, and the delay factor, D. $M = 100$, $L = 16$, and $N = 1$.

In Fig. 11 we evaluate the impact of the number of UEs on the identification accuracy. We can see that as the number of UEs increases from $K = 2$ to $K = 8$ the accuracy decreases for both the synchronous and asynchronous cases. This decrease could be explained by the fact that as K increases the dimension of the feature space increases leading to increased complexity and variability of the data making it more difficult for the model to generalize from the training data. From Figs. 6, 8, and 10, to improve the identification accuracy for a certain value of K we can increase the number of BS antennas and/or the length of the pilot signal and/or the number of frames.

Fig. 12 shows the confusion matrices which evaluate the performance of the categorical class predictions made by IDNet for $K = 4$ and $D \in \{0, 7\}$. The labels UE_i,

Fig. 12. IDNet confusion matrices for $D = 0$ (top) and $D = 7$ (bottom).$M = 100$, $K = 4$, $L = 16$, $P_{\text{ED}} = 20$ dBm, $N = 3$, and $\Delta f = 0$.

$i = 1, ..., K$, denote the case of UE i being attacked. By observing the leading diagonal, the percentage of correctly identified attacked UEs is higher when the delay factor is higher with an average increase of 7.1%. In both cases, the true negative precision is high with an average accuracy of 98.4% resulting in a low false alarm rate of 1.6%. In the case of an attack, we can clearly see that most miss-classifications are classified as no attack which minimises the risk of terminating downlink communication to unaffected UEs, limiting unnecessary disruptions to the network.

V. Conclusion

In this paper, we have proposed a DNN framework, called IDNet, to detect pilot spoofing attacks and identify the attacked UE in massive MIMO networks. We have developed a new model in which the active ED is asynchronous with the system and showed that this asynchronous attack results in increasing the received signal subspace dimension by one, which was exploited for attack identification. We have used the eigenvalues of the sample covariance matrix the received pilot signal as well as the ratio between the power of the received signal projected onto the pilot signals and its expected value as input features to IDNet. It was shown that IDNet is effective in identifying the attacked UE even for small ED's powers. We have also showed that IDNet outperforms conventional approaches for all ED's powers and different timing/frequency mismatches. Simulation results also showed that the larger the timing and/or frequency mismatches of the ED, the higher the identification accuracy meaning that asynchronous pilot spoofing attacks are easier to be identified compared to synchronous pilot spoofing attacks. We have also showed that increasing the number of antennas at the BS, the pilot signal length, or the number of frames lead to enhanced identification accuracy. Given the effectiveness of IDNet in detecting pilot spoofing attacks and identifying the attacked UE, the BS can use it to selectively terminate or pause transmission to only the attacked UE, minimizing unnecessary network downtime and disruption for the unaffected UEs. These findings, therefore, make a compelling case for DL-aided pilot spoofing attack detection in massive MIMO networks. In the future, it is worthwhile extending our IDNet to a more general system where multiple EDs can attack different UEs.

References

[1] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, pp. 154–655, 2017. [Online]. Available: https://nowpublishers.com/article/Details/SIG-093

[2] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Commun. Standards Mag.*, vol. 1, no. 4, pp. 24–30, Dec. 2017.

[3] H. Jin, K. Liu, M. Zhang, L. Zhang, G. Lee, E. N. Farag, D. Zhu, E. Onggosanusi, M. Shafi, and H. Tataria, "Massive MIMO evolution toward 3GPP release 18," *IEEE J. on Sel. Areas Commun.*, vol. 41, no. 6, pp. 1635–1654, Jun. 2023.

[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.

[5] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[6] X. Zhou, B. Maham, and A. Hjorungnes, "Pilot contamination for active eavesdropping," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 903–907, Mar. 2012.

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

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

[9] 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)*, 2014, pp. 585–589.

[10] L. Ning, B. Li, C. Zhao, Y. Tao, and X. Wang, "Detection and localization of the eavesdropper in MIMO systems," *IEEE Access*, vol. 8, pp. 94 984–94 993, 2020.

[11] S. Xu, W. Xu, H. Gan, and B. Li, "Detection of pilot spoofing attack in massive MIMO systems based on channel estimation," *Signal Process.*, vol. 169, p. 107411, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168419304633

[12] W. Xu, C. Yuan, S. Xu, H. Q. Ngo, and W. Xiang, "On pilot spoofing attack in massive MIMO systems: detection and countermeasure," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1396–1409, 2021.

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

[14] ——, "Pilot spoofing attack detection and countermeasure," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2093–2106, May 2018.

[15] K. Yuan, L. Guo, C. Dong, and T. Kang, "Detection of active eavesdropper using source enumeration method in massive MIMO," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–5.

[16] J. Vinogradova, E. Björnson, and E. G. Larsson, "Detection and mitigation of jamming attacks in massive MIMO systems using random matrix theory," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016, pp. 1–5.

[17] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.

[18] P. Jiang, K. Yan, H. Zhang, G. Yao, and L. Li, "A Hankel-based singular vector source enumeration for low signal-to-noise ratio," in *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, 2015, pp. 1–4.

[19] J. Wang, F. Ji, F. Chen, and Z. Hu, "Correlated source number estimation with Gerschgorin radii of partitioned matrices products," *Wireless Personal Commun.*, vol. 107, pp. 1077–1091, Jul. 2019.

[20] W. Hu, R. Liu, X. Lin, Y. Li, X. Zhou, and X. He, "A deep learning method to estimate independent source number," in *2017 4th International Conference on Systems and Informatics (ICSAI)*, 2017, pp. 1055–1059.

[21] J. Rogers, J. E. Ball, and A. C. Gurbuz, "Estimating the number of sources via deep learning," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–5.

[22] Y. Yang, F. Gao, C. Qian, and G. Liao, "Model-aided deep neural network for source number detection," *IEEE Signal Process. Lett.*, vol. 27, pp. 91–95, 2020.

[23] T. M. Hoang, T. Q. Duong, H. D. Tuan, S. Lambotharan, and L. Hanzo, "Physical layer security: Detection of active eavesdropping attacks by support vector machines," *IEEE Access*, vol. 9, pp. 31 595–31 607, 2021.

[24] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3880–3900, Jul. 2016.

[25] M. Alageli, A. Ikhlef, and J. Chambers, "SWIPT massive MIMO systems with active eavesdropping," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 233–247, 2019.

[26] A. Pitarokoilis, E. Bjömson, and E. G. Larsson, "On the effect of imperfect timing synchronization on pilot contamination," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[27] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Comput. Sci.*, vol. 2, Aug. 2021.

[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[29] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *CoRR*, vol. abs/1710.05941, 2017. [Online]. Available: http://arxiv.org/abs/1710.05941

[30] 3rd Generation Partnership Project (3GPP), "5G study on channel model for frequencies from 0.5 to 100 GHz," pp. 28–29, 11 2020. [Online]. Available: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=55982

[31] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. Abou Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Appl. Sci.*, vol. 11, p. 796, 2021.