Smart Channel State Information Pre-Processing for Authentication and Symmetric Key Distillation

Muralikrishnan Srinivasan, Sotiris Skaperas, Miroslav Mitev, Mahdi Shakiba Herfeh, M. Karam Shehzad, Philippe Sehier, Arsenia Chorti

Smart Channel State Information Pre-processing for Authentication and Symmetric Key Distillation

Muralikrishnan Srinivasan[1], Sotiris Skaperas[2], Miroslav Mitev[3], Mahdi Shakiba Herfeh[4], *Member, IEEE,* M. Karam Shehzad[5], *Member, IEEE,* Philippe Sehier[5], and Arsenia Chorti[4,6], *Senior Member, IEEE;*
[1]Chalmers University of Technology, Gothenburg, Sweden
[2]University of Macedonia, Greece
[3]Last Mile Semiconductor, Dresden, Germany
[4]ETIS UMR 8051 / CY Paris University, ENSEA, CNRS, 95000, Cergy, France
[5]Nokia Bell-Labs, 12 Rue Jean Bart, 91300, Massy, France
[6]Barkhausen Institute, 01187 Dresden, Germany;
mursri@chalmers.se, sotskap@uom.edu.gr {mahdi.shakiba-herfeh, arsenia.chorti}@ensea.fr
miroslav.mitev@barkhauseninstitut.org, {karam.shehzad, philippe.sehier}@nokia.com

***Abstract*—While the literature on channel state information (CSI)-based authentication and key distillation is vast, the two topics have customarily been studied separately. This paper proposes unsupervised learning techniques to disentangle deterministic from stochastic fading to decompose observed CSI vectors into "predictable" and "unpredictable" components. The former, primarily due to large-scale fading, can be used for node authentication. The latter, primarily due to small-scale fading, can be used for secret key generation (SKG). The parameterization of the decomposition is performed using the following metrics: (i) a CSI fingerprint "separability" criterion, expressed through the maximisation of the total variation distance (TVD) between the empirical CSI fingerprints; (ii) a statistical independence metric for CSI collected at different users in neighboring locations, using the d-dimensional Hilbert Schmidt independence criterion (dHSIC) test statistic; (iii) an estimation of information leakage at different users to determine the amount of necessary hashing for privacy amplification in the SKG using the FBLAEU machine learning based conditional min-entropy estimator. Employing principal component analysis (PCA), kernel PCA and autoencoders on synthetic and natural CSI datasets, this work shows that *explicit* security guarantees can be provided by using physical layer security for authentication and key agreement.**

I. Introduction

Sixth generations (6G) systems will be required to meet diverse constraints in an integrated ground-air-space global network. In particular, meeting overly aggressive latency constraints and operating in massive connectivity regimes with low energy footprint and low computational effort while providing explicit security guarantees can be challenging [1]. In addition, the extensive introduction of artificial intelligence (AI) and machine learning (ML) and the rapid advances in quantum computing are further developments that will increase the attack surface of 6G systems [2], [3]. More importantly, the massive deployment of low-end Internet of things (IoT) nodes [4], often produced following non-homogeneous production processes and with expected lifespans exceeding 10 years, poses pressing questions concerning future security architectures.

At the same time, the integration of communications and sensing, along with embedded (on-device) AI, can provide the foundations for building autonomous and adaptive security controls, orchestrated by a vertical security plane in coordination with a vertical semantic plane, dubbed as context-aware smart security [5], [6].It is in this framework that we envision the incorporation of physical layer security (PLS) schemes in 6G security protocols, introducing security controls at all layers for the first time [7]. This exciting prospect does not come, however, without challenges. Despite intense research interest in PLS for more than two decades, its incorporation in actual security products remains largely elusive. A key reason behind this is that PLS relies on the physical aspects of the transmission [8] rather than mathematical algorithms. This limitation will be overcome in 6G as channel engineering will be widely used to meet communication needs and enable PLS.

In this direction, we propose in this work smart pre-processing algorithms of observed channel state information (CSI) vectors for CSI-based authentication and secret key generation (SKG). We aim to separate the authentication signatures (CSI fingerprints) from reciprocal random components that can be used for SKG. Despite the immense bibliography on CSI-based authentication and SKG, a systematic treatment of the CSI as jointly a source of uniqueness (deterministic fading) and entropy (stochastic fading) is missing. Therefore, this paper aims at filling this gap and building pre-processing for joint SKG and authentication.

A. Unsupervised learning based pre-processing

Overall, our approach leverages unsupervised learning in the form of dimensionality reduction as a generic tool to transform CSI measurements with dependencies across different dimensions (time, frequency, space, antenna) into i)

Muralikrishnan Srinivasan was supported by the Swedish Foundation for Strategic Research (SSF, HOT-OPTICS Project). A. Chorti has been supported by the INEX Funding of Excellence (projects eNiGMA and PHEBE) and the PEPR Networks of the Future Programme of France 2030 (Projects 8 and 9)
(Corresponding author: Muralikrishnan Srinivasan.)

location dependent components and ii) random components that decorrelate over very short-distances.

The proposed pre-processing is based on the observation that large-scale fading, which is determined by path-loss and shadowing, is location-dependent and can be useful for authentication purposes [9]. On the other hand, small-scale fading is stochastic in nature and can constitute an entropy source for SKG, e.g., see [10] and [11]. To disentangle these two processes we employ unsupervised learning techniques, namely principal component analysis (PCA), kernel PCA (KPCA), and autoencoders (AEs). Our aim is to perform a power-domain decomposition (large versus small scale) of the CSI into deterministic and stochastic fading. The goal is to retrieve from the observed CSI: a predictable (deterministic) component primarily due to large-scale fading, and an unpredictable (stochastic) component primarily due to small-scale fading.

The proposed decomposition is designed to satisfy three criteria: (i) maximum separability of CSI-based authentication fingerprints, measured by the total variation distance (TVD) between empirical measures; (ii) minimum dependence between the sources of shared randomness at different nodes, measured by a normalized version of the d-dimensional Hilbert Schmidt independence criterion (dHSIC) [12] test statistic; (iii) minimization of the conditional min entropy (information leakage) at different users in neighbouring locations, measured using the FBLAEU estimator [13]. We validate our proposed pre-processing approach using both synthetic datasets generated by the Quadriga models [14] and experimentally measured outdoor CSI datasets from Nokia [15] for a massive multiple input multiple output (mMIMO) setting.

B. Contributions

The main contributions of this paper are:

1) **Unsupervised ML-based pre-processing schemes:** We propose pre-processing schemes based on PCA, KPCA, and two different AEs to disentangle the predictable components from the unpredictable components in the CSI.
2) **Maximum separability criterion of CSI fingerprints for authentication:** The paper uses the TVD to study the separability of the components used for node-authentication. This contribution provides insights into the effectiveness of the proposed approach for authentication based on the nearest neighbour classifier.
3) **Study of spatial correlation and reciprocity trade-off for SKG:** We investigate the trade-off between spatial correlations (SC) and dependencies at different locations and reciprocity between the uplink and downlink for SKG. We evaluate the degree of spatial dependence using a novel metric, referred to as normalized dHSIC.
4) **Information leakage estimation:** A conditional min-entropy estimator is used to evaluate the necessary hashing rate (privacy amplification) for SKG. Finally, the quality of the SKG keys is validated using the National Institute of Standards and Technology (NIST) test suite.
5) **Full SKG chain presented:** In SKG literature it is common to describe but not implement the privacy amplification stage. In this work, we present the full SKG chain and discuss trade-offs in terms of key rate and information reconciliation rate as well as the impact of different pre-processing schemes.

Overall, the contributions made in this paper provide a systematic treatment of the CSI as jointly a source of uniqueness and a source of entropy. The proposed ML pre-processing schemes, evaluation of spatial independence, study of randomness, and assessment of spatial uniqueness provide a fresh perspective on joint SKG and authentication using CSI data.

The rest of the paper is organized as follows. Sections II and III review the background concepts for CSI-based authentication and SKG and introduce the metrics proposed as design criteria. Section IV presents the proposed approaches for disentangling predictable CSI components from unpredictable ones using PCA, KPCA, and two different AEs. Section V presents the datasets and section VI includes numerical results. Section VII presents the discussion and concluding remarks.

II. CSI-Based Node Authentication

This section provides an overview of CSI-based node authentication and its underlying principles, including the proposed metrics for pre-processing.

A. Process and Principle

CSI-based authentication is a technique used to verify the identity of a device. The process involves the following steps:

1) The device collects CSI information, which is a set of measurements that describe the state of the wireless channel.
2) The device either processes the information locally or sends it to the authentication server, which can be located in the cloud, edge or on a local network.
3) The CSI information is used to determine the authenticity of the device by comparing it with a reference database of known and trusted CSI patterns.
4) If the CSI information is consistent with that of a known and trusted device, the authentication server grants access to the wireless network. Otherwise, access is denied.

Therefore, authentication requires a verifiable source of uniqueness, related, for example, to node positioning and fingerprinting. A promising use case concerns the identification of false base stations, leveraging public knowledge of genuine base stations locations. Early results on using azimuth and elevation angles of arrival as location features in mMIMO settings have been shown to provide resistance against location spoofing.

In all cases, the CSI fingerprints used for authentication must be statistically separable for each location. Also, it is beneficial if the fingerprints vary only slowly [16]. Various approaches to exploit different types of channel parameters for CSI-based authentication have been proposed in the literature [17]–[19]. In this work we propose pre-processing techniques to enhance the statistical separability of CSI-fingerprints, validated via a comparison of performance achieved without the pre-processing.

B. Metrics

One of the key contributions of our work is to introduce a new design criterion for pre-processing to extract highly separable fingerprints from CSI vectors. Specifically, we propose using a probability distribution distance metric as the cost function for tuning the parameters of the pre-processing step. The goal is to maximize the separability between fingerprints of nearby nodes, in order to improve the accuracy of authentication.

We utilize the TVD metric to evaluate the separability of CSI-based authentication fingerprints. TVD is a well-known distance measure between two probability distributions and measures the L1 distance between their empirical measures.

Definition. *Let μ and ν be two Borel probability measures on a metric space $\mathcal{X}$, then the TVD between μ and ν is defined as follows [20]:*

$$TVD(\mu,\nu) = \sup_{A\subset\mathcal{X}} \left|\mu(A) - \nu(A)\right|,$$

where the supremum is over Borel-measurable sets. For two probability distributions μ and ν defined on a countable configuration space $\mathcal{X}^N$, the TVD is defined as:

$$TVD(\mu,\nu) = \frac{1}{2}\sum_{x\in\mathcal{X}^N} \left|\mu(x) - \nu(x)\right|.$$

The TVD serves as a distance metric to measure the separability of the extracted fingerprints for authentication purposes. It calculates the L1 distance between two probability distributions by summing the absolute differences between their probabilities at each possible configuration in the space. The greater the TVD, the more distinguishable the two distributions and the more reliable the extracted fingerprints for authentication. Conversely, if the TVD is small, the two distributions are similar and the extracted fingerprints are less reliable.

We note in passing that it is possible to build an attack-resilient profile by utilizing various techniques to identify spoofing attacks, such as clustering analysis of the extracted fingerprints, as proposed in [21]. Additionally, the correlation between any two adjacent CSI measurements within a specific time window can be taken into account for user authentication [21]. Such aspects are outside the scope of this work and will be considered in future studies.

III. Background Concepts on SKG

In the following we review background concepts on the SKG protocol and provide a brief commentary on how it relates to cryptographic schemes for key generation and related metrics.

A. Process and Principle

The SKG protocol comprises three steps, explained below:

1) Advantage distillation: In this step, the users use pilot signals transmitted over the coherence time to excite the channel, in order to obtain highly correlated observation sequences at two remote nodes, referred to as Alice and Bob. These analog measurements are then converted into binary sequences using quantization.
2) Information reconciliation: This step is used to detect and correct discrepancies in the quantizer outputs at the legitimate parties.
3) Privacy Amplification: In this step, the legitimate users use hashing to generate a maximum entropy unpredictable secret key. The hashing rate is determined by the estimation of the conditional min entropy, accounting for observations at potential eavesdroppers.

To generate secret keys from wireless fading coefficients, three factors are exploited: (i) the channel reciprocity between two nodes, Alice and Bob, during the channel's coherence time; (ii) temporal variation due to node mobility and dynamics [22]; and (iii) spatial independence (typically measured through decorrelation) at distances of the same order of magnitude as the wavelength, i.e., in the order of a few centimetres for sub-6GHz systems. In more detail, according to Jakes' model, the channel will be uncorrelated when a third party is located half a wavelength away [23]. However, experimental results show that half-wavelength distance spatial decorrelation is valid only in very rich scattering environments [24]–[27].

Additionally, the keys need to be generated from stochastic fading components to be unpredictable; attacks on SKG from deterministic fading have been launched using ray-tracing and have been successfully demonstrated experimentally [28]. Despite this, many existing works perform SKG without systematically removing predictable components of wireless channel coefficients [29]–[31] and without explicitly accounting for them in the privacy amplification step, which is customarily omitted in published work.

In this paper, we focus on minimizing SC and dependencies at potential attackers, referred to as Eves, in the vicinity of legitimate users, explicitly as a design criterion in the pre-processing stage. Without any pre-processing of the observed CSI, an attacker can distil highly correlated observations with these at the legitimate nodes due to deterministic fading. Although these dependencies can be removed through privacy amplification, this approach requires the use of i) larger quantizers; and ii) heavier hashing, leading to less energy-efficient solutions and potentially erroneous implementations[1].

Remark: The proposed CSI pre-processing offers the advantage of linking SKG to the concept of a cryptographic pseudorandom number generator (PRG) during the advantage distillation phase. In more detail, with respect to a PRG, we refer to the following standard definition of pseudorandomness in cryptography [32]:

Definition. *The ensemble $\{G(U_n)\}_{n\in\boldsymbol{N}}$ is pseudorandom, iff for any probabilistic polynomial-time algorithm A, for any positive polynomial p, and for all sufficiently large n,*

$$|Pr(A(G(U_n);1^{l(n)})=1) - Pr(A(U_{l(n)};1^{l(n)})=1)| < \frac{1}{p(n)}.$$

[1]It should be noted that any correlations in the time, frequency, space, or antenna domains between the reconciled sequences at Alice, Bob, and potential Eves should be explicitly taken into account when evaluating the conditional min-entropy to estimate the target privacy amplification rate.

Here, the probabilistic polynomial-time algorithm A can be seen as a statistical test, and a generator G fails the test if an algorithm A exists such that the above condition does not hold. In cryptography, a semantically secure PRG has the property of "unpredictability" (resistance to next bit predictors), which means that an observer who knows i bits of the output of the PRG should be unable to predict the $(i+1)$-th bit with a probability that is greater than $\frac{1}{2}$ by more than a negligible quantity that increases only polynomially in time $\frac{1}{p(n)}$. With respect to SKG, independence between the observations of the legitimate users and the observations of Eve ensures unpredictability as independence is a stronger condition.

B. Metrics

Although various polynomial time statistical tests of unpredictability have been proposed, such as the ones in the NIST suite [33], these tests were designed to asses the randomness of a single sequence generated by a PRG. As a result, they are not equipped to evaluate dependencies and cross-correlations of sequences observed at legitimate and adversarial nodes in close proximity in the context of SKG. In order to address this issue, in this work we make use of four metrics of increasing statistical accuracy in capturing dependencies, and correspondingly of increasing computational complexity. Namely, in this work we evaluate i) the mismatch probability (MP) between observed sequences assuming a one-bit quantizer; ii) the Pearson cross-correlation coefficient; iii) a novel metric for measuring dependencies, referred to as dHSIC; and finally, iv) the conditional min-entropy of the observed sequences. The metrics are discussed in detail in the following.

1) Mismatch probability: It is important to balance the reduction of dependencies between legitimate and adversarial observed sequences during pre-processing with preserving the reciprocity between the observed sequences at Alice and Bob. To assess reciprocity, we use a one-bit quantizer about the median point along the time dimension on the CSI sequences observed in the uplink (Alice to Bob) and downlink (Bob to Alice). The mismatch probability (MP) between Alice and Bob is defined as the ratio of the number of bits in disagreement to the total number of bits. These mismatches are subsequently corrected during the information reconciliation step, but a high MP requires a lower rate reconciliation decoder to ensure zero frame-error-rate (FER) reconciliation[2], i.e., key-disagreement-free SKG, which becomes impractical for short blocklengths beyond a certain point.

2) Pearson cross-correlation coefficient: One widely used metric for measuring the degree of correlation between two variables is the Pearson cross-correlation coefficient (CC), which has been applied in the literature as an indirect measure of unpredictability in the context of PLS. However, we note that decorrelation alone may not be sufficient to prove unpredictability, in the case of non-linear dependencies and non-Gaussian distributions. Instead, the independence of observations between legitimate and adversarial entities is a stronger criterion that guarantees unpredictability and allows the design of SKG to align with the definition of a PRG at the advantage distillation step.

3) dHSIC: We propose to use the dHSIC [12] to measure dependencies between the observed CSI vectors. The dHSIC is a kernel-based statistical test of independence that applies a positive-definite kernel on N-dimensional random variables (RVs) to determine whether the dimensions are independent. The null hypothesis indicates that the vectors are mutually independent and their joint probability density function can be expressed as the product of the marginals, while the alternative hypothesis denotes that the vectors consist of at least two dependent components. An estimator of the statistical functional is defined as $dHSIC(\tilde{\boldsymbol{H}})$, which can be calculated using the following equation [12, Def 2.6]:

$$dHSIC(\tilde{\boldsymbol{H}}) = \frac{1}{M^2}\sum_{i,j=1}^{M}\prod_{l=1}^{N}\left(\mathbf{1}_{M\times M}\circ K_{ij}^{l}\right) + \frac{1}{M^{2N}}\prod_{l=1}^{N}\sum_{i,j=1}^{M}K_{ij}^{l} - \frac{2}{M^{N+1}}\sum_{i,j=1}^{M}\prod_{l=1}^{N}\left(\mathbf{1}_{M\times 1}\circ K_{ij}^{l}\right), \tag{1}$$

where the operator $\circ$ denotes the Hadamard product and $\mathbf{1}_{M\times M}$ is an $M\times M$ matrix of ones. Also, $\mathbf{K}^l = \left(\mathbf{K}_{ij}^l\right) = \left(k^l(x_i,x_j)\right) \in \mathbb{R}^{M\times M}$ is the Gram matrix of the positive semi-definite Gaussian kernel k^l, defined $\forall x_i, x_j \in \mathbb{R}$ by, $k^l = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}\right)$, with bandwidth $\sigma = \sqrt{\frac{\text{med}\left(\left\|x_i-x_j\right\|^2\right)}{2}}$ and med$(\cdot)$ is the median heuristic.

According to [12, Theorem 3.1], with respect to the hypothesis test at hand, the critical value (for a specific significance level α) can be obtained as below,

$$CV_\alpha = \left[\mathbf{D}^{dHSIC'}\right]_{\lceil (B+1)(1-\alpha)\rceil + \sum_{i=1}^{B}\mathbb{1}_{\{dHSIC'(\tilde{\boldsymbol{H}}) = dHSIC'(\tilde{\boldsymbol{H}}_i)\}}},$$

where the vector $\mathbf{D}^{dHSIC'}$ contains the B Monte-Carlo realisations of $dHSIC'(\tilde{\boldsymbol{H}})$ in an increasing order; the re-sampling function $dHSIC'\left(\tilde{\boldsymbol{H}}\right)$, $\tilde{\boldsymbol{H}} = \left(r_1(\tilde{\boldsymbol{h}}_1), \cdots, r_N(\tilde{\boldsymbol{h}}_M)\right)$ is constructed by $r_1, \cdots, r_M$ random re-samplings without replacement. The operators $\lceil . \rceil$ and $[.]_j$ denote the ceiling function and the j-th element of a vector respectively, and $\mathbb{1}_{\{.\}}$ is the indicator function.

We propose a normalised metric for measuring the level of dependence between variables, based on the $dHSIC$ test statistic and the corresponding critical value. The metric is expressed as

$$\overline{\Delta} = \frac{dHSIC(\tilde{\boldsymbol{H}})}{CV_\alpha}\mathbb{1}_{dHSIC(\tilde{\boldsymbol{H}}) > CV\alpha}, \tag{2}$$

where $\overline{\Delta}$ is close to unity when the variables exhibit low dependence and grows without bound with increasing dependence.

4) Conditional min entropy: The randomness of binary sequences is commonly evaluated using the min-entropy metric [34], [35]. The min-entropy measures the minimum number of binary strings that are required to generate the observed sequence with high probability. Specifically, the min-entropy

[2]A frame refers to a code block, and the frame error rate represents the likelihood or frequency of errors occurring within a single block.

of a sequence $\mathbf{r}$ is defined as $G_\infty(\mathbf{r}) = -\log_2 \max_{\mathbf{r}\in\mathcal{R}} p(\mathbf{r})$, where $\mathcal{R}$ is the set of all possible values of $\mathbf{r}$. In the context of key generation at the output of a PRG, the key $\mathbf{k}$ should satisfy the randomness inequality [36]:

$$|\mathbf{k}| \leq G_\infty(\mathbf{r}). \tag{3}$$

However, the min-entropy metric does not account for any leakage to an eavesdropper and is therefore only suitable when no leakage is expected, such as when keys are generated using a cryptographic PRG. In the case of SKG, to evaluate the randomness of keys in the presence of an eavesdropper, we use the conditional min-entropy [37], which measures the minimum number of binary strings that are required to generate the key $\mathbf{k}$, given the eavesdropper's observations, i.e.,

$$G_\infty(\mathbf{r}|\mathbf{r}_E) = -\log_2 \max_{\mathbf{r}\in\mathcal{R},\mathbf{r}_E\in\mathcal{R}_E} p(\mathbf{r}|\mathbf{r}_E), \tag{4}$$

where $\mathcal{R}_E$ is the set of all possible observed sequences at the eavesdropper. In fact, the difference between the min-entropy and the conditional min-entropy represents the amount of information leaked to the eavesdropper [38]:

$$\text{Leakage} = G_\infty(\mathbf{r}) - G_\infty(\mathbf{r}|\mathbf{r}_E). \tag{5}$$

The above metric provides a measure of the leakage and can be used to determine if pre-processing improves the conditional min-entropy and reduces leakage.

Finally, we also need to account for the additional leakage that occurs during the information reconciliation phase due to side information exchange. As a result, the length of the final key can be upper-bounded as follows [39]:

$$|\mathbf{k}| \leq G_\infty(\mathbf{r}|\mathbf{r}_E) - |\mathbf{s}_A|, \tag{6}$$

where $|\mathbf{s}_A|$ denotes the length of the syndrome. It is worth noting that (6) assumes a worst-case scenario where the side information is independent of the leakage during advantage distillation [39].

IV. Proposed Power Domain Pre-processing

As explained in earlier sections, we employ three different techniques: PCA, KPCA, and AEs. PCA is a linear decomposition that can effectively capture the dominant components of the CSI. KPCA and AE, on the other hand, are capable of capturing non-linear dependencies within the data. Fig. 1 provides a visual representation of our proposed pre-processing techniques and their role in implementing authentication and SKG using CSI matrices. This figure illustrates the flow of the pre-processing steps, highlighting how the techniques capture the desired components for authentication and SKG. After the pre-processing stage, the predictable components obtained can be appropriately used to train a classifier or run a hypothesis test for authentication. The unpredictable components, on the other hand, are utilized in subsequent steps of the SKG process, including quantization, information reconciliation, and privacy amplification, to generate the secret key. As discussed in previous sections, to evaluate the statistical separability of CSI-based authentication fingerprints, we use the TVD, while to evaluate the spatial dependencies after quantization we incorporate four different metrics (MP, CC, dHSIC and conditional-min entropy).

A. Pre-processing using PCA

Let $\boldsymbol{H}_u = [\boldsymbol{h}_{1u}, \cdots, \boldsymbol{h}_{Nu}]$ denote the observed CSI matrix at Bob (aggregating the CSI vectors from all Alices) and $\boldsymbol{U}$ the $M \times M$ matrix whose rows are the eigenvectors of the matrix $\text{Cov}(\mathbf{H}_u)$, sorted in decreasing order. In many scenarios, e.g., Rician and generally line of sight settings, it is plausible to assume that the first few principal components (PCs) correspond to the dominant large-scale fading terms, while the rest of the PCs correspond to small scale fading terms and noise. Using the eigenvectors $\widehat{D} \times M$ matrix $\boldsymbol{U}_{1:\widehat{D}}$ corresponding to the first $\widehat{D}$ PCs, we want to isolate the predictable part of the observed channel that will be used for CSI-based authentication, as follows,

$$\widehat{\boldsymbol{H}}_u = \boldsymbol{U}_{1:\widehat{D}}^H \boldsymbol{W}_u, \tag{7}$$

where symbol $.^H$ denotes the Hermitian transpose and the $\widehat{D} \times N$ matrix $\boldsymbol{W}_u$ is given by,

$$\boldsymbol{W}_u = \boldsymbol{U}_{1:\widehat{D}} \boldsymbol{H}_u, \tag{8}$$

and $\widehat{\boldsymbol{H}}_u = \left[\widehat{\boldsymbol{h}}_{1u}, \cdots, \widehat{\boldsymbol{h}}_{Nu}\right]$ for $u \in \{a, b\}$ is a $M \times N$ matrix. Furthermore, we want to identify a region of PCs with indices $\{\tilde{D}_1, \ldots, \tilde{D}_2\}$, corresponding to components $\tilde{\boldsymbol{H}}_u = \left[\tilde{\boldsymbol{h}}_{1u}, \cdots, \tilde{\boldsymbol{h}}_{Nu}\right]$ for $u \in \{a, b\}$ over which low dependence and correlation between Alices and Eves is achieved while keeping the MP below a threshold.

Note that the PCs beyond $\tilde{D}_2+1$ are dominated by noise and should be neglected (denoising). To efficiently disentangle the CSI matrix into predictable and unpredictable parts, the triplet $\{\widehat{D}, \tilde{D}_1, \tilde{D}_2\}$ is chosen such that the TVD is maximized for the first $\widehat{D}$ PCs while $\overline{\Delta}$, leakage and the MP are kept as low as possible for the range $\{\tilde{D}_1, \ldots, \tilde{D}_2\}$. We discuss the trade-off between minimizing $\overline{\Delta}$ (as a measure of independence) and the MP in detail in Section VI.

B. Pre-processing using KPCA

KPCA is a natural extension of PCA, that applies a kernel function to map the data into a higher-dimensional space $\mathbb{F}$ [40], allowing for capturing nonlinear characteristics of the CSI matrix. Here, the eigenvalue problem is solved in the feature space for the KPCA decomposition,

$$\widetilde{\mathbf{K}}\boldsymbol{\alpha} = N\lambda\boldsymbol{\alpha}, \tag{9}$$

where $\boldsymbol{\alpha}_i = \frac{\mathbf{V_i}}{\sqrt{\lambda_i}}$, $i = 1, \ldots, N$, and, matrix $\mathbf{V}$ and vector λ ($\lambda \geqslant 0$) denote the eigenvectors and the eigenvalues, respectively. Also, $\widetilde{\mathbf{K}}$ is the centralized Gram matrix, given by,

$$\widetilde{\mathbf{K}} = \mathbf{K} - \frac{1}{N}\mathbf{1}_{N\times N}\mathbf{K} - \frac{1}{N}\mathbf{K}\mathbf{1}_{N\times N} + \frac{1}{n^2}\mathbf{1}_{N\times N}\mathbf{K}\mathbf{1}_{N\times N} \tag{10}$$

where the Gram matrix $\mathbf{K} \in \mathbb{C}_{N,N}$ is based on the positive semi-definite complex Gaussian kernel and $\mathbf{1}_{N\times N}$ is an $N \times N$ matrix of ones.

Then, the first $\widehat{D}$ nonlinear PCs could be extracted through computing projections of the original data on the eigenvectors $\mathbf{V}_i$ in feature space $\mathbb{F}$, as above,

$$\mathbf{Y}_{1:\widehat{D}} = \boldsymbol{\alpha}_{1:\widehat{D}}^H \widetilde{\mathbf{K}}. \tag{11}$$

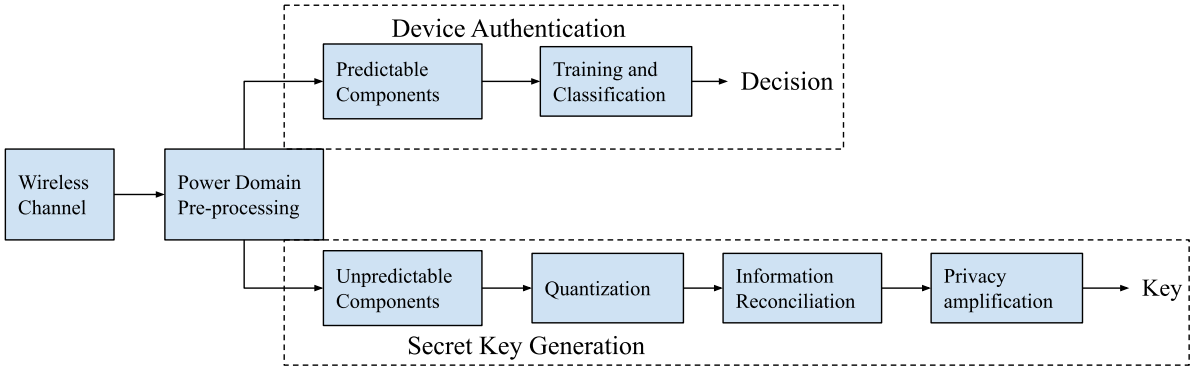


Fig. 1: Block diagram for proposed PLS-based authentication and secret key generation. The pre-processing schemes disentangle the predictable components from the unpredictable components in the CSI. The predictable components are used for authentication and the unpredictable components are used for secret key generation.

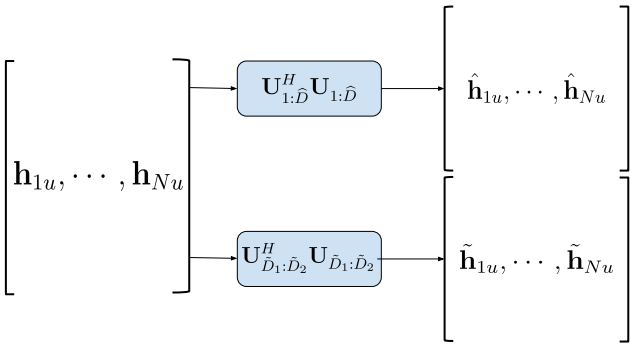


Fig. 2: A representative block diagram of PCA.

In order to estimate the predictable and the unpredictable parts as before, we apply KPCA reconstruction; unlike in PCA, this can not be done explicitly since ϕ is unknown. Hence, we approximate the reconstruction matrix based on the kernel ridge regression [41]. It follows that the predictable part of the observed channel is computed as follows

$$\widehat{\mathbf{H}}_u = \boldsymbol{\beta}\mathbf{K}_{Y_{1:\widehat{D}}}, \quad \boldsymbol{\beta} = \mathbf{H}_u(\mathbf{K}_{Y_{1:\widehat{D}}} + \gamma\mathbf{I})^{-1}, \tag{12}$$

where $\mathbf{K}_{Y_{1:\widehat{D}}} = \big(k(y_i, y_j)\big)$ is the complex Gaussian Gram matrix of the first $\widehat{D}$ PCs and γ is the hyperparameter of the ridge regression. Subsequently, we derive the unpredictable part of the observed CSI directly as the residual of removing the predictable part, i.e.,

$$\tilde{\mathbf{H}}_u = \mathbf{H}_u - \widehat{\mathbf{H}}_u, \tag{13}$$

In the case of KPCA, unlike in PCA, denoising is not performed.

C. Pre-processing using AEs

AEs are unsupervised learning architectures that utilise and learn two functions, an encoder that maps the M dimensional input matrix $\boldsymbol{h}_{nu}$ into $\widehat{D}$ dimensional encoded values $\boldsymbol{w}_{nu}$ $\forall\, n = 1, \ldots, N$ for $u \in \{a, b\}$ and a decoder that maps the encoded values back to an M dimensional output $\widehat{\boldsymbol{h}}_{nu}$, $\forall\, n = 1, \ldots, N$ and for $u \in \{a, b\}$, such that the loss-function

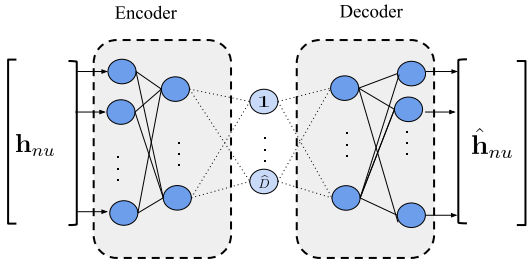


Fig. 3: A representative block diagram of AE1.

$$E_1 = \frac{1}{N}\sum_{n=1}^{N} \|\boldsymbol{h}_{nu} - \widehat{\boldsymbol{h}}_{nu}\|_2^2, \text{ for } u \in \{a, b\}, \tag{14}$$

corresponding to the mean square error (MSE) between the AE input and output, is minimal. An AE can be used to extrapolate a $\widehat{D}$-dimensional representation $\boldsymbol{w}_{nu}$, $\forall\, n = 1, \ldots, N$ that can capture the dominant components. We treat the output of the decoder $\widehat{\boldsymbol{h}}_{nu}, \forall n = 1, \ldots, N,$ for $u \in \{a, b\}$ as the dominant predictable components under the conjecture that most of the received signal strength is due to large scale fading effects. Here again, we assume that the residuals

$$\left\{\tilde{\boldsymbol{h}}_{nu}(\widehat{D})\right\}_{n=1}^{N} = \{\boldsymbol{h}_{nu} - \widehat{\boldsymbol{h}}_{nu}\}_{n=1}^{N}, \text{ for } u \in \{a, b\} \tag{15}$$

correspond to the unpredictable components of the CSI matrix.

In light of this, the value of $\widehat{D}$ is a hyperparameter that can be tuned to a more fine-grained loss function focusing on SKG; in particular, we build an alternative loss function to balance the spatial correlation with the reciprocity of the residuals in the uplink and the downlink. Since we want to

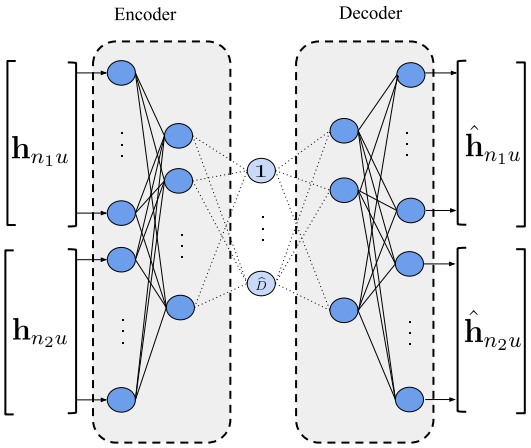


Fig. 4: A representative block diagram of AE2.

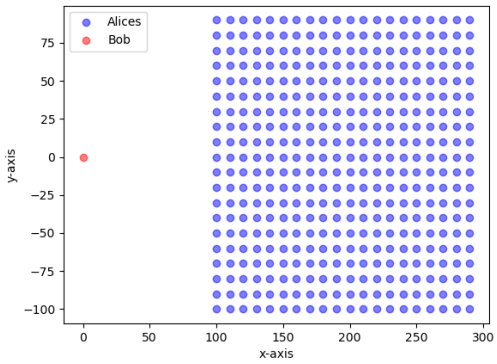


Fig. 5: Node positions in the Quadriga synthetic dataset.

lower correlation, the loss function can also explicitly specify a correlation term instead of the MSE. Consequently, the following loss function is proposed:

$$E_2 = \frac{1}{N} \sum_{\substack{n_1=1 \\ n_2 \in \mathcal{U}(n_1)}}^{N} \tilde{\boldsymbol{h}}_{n_1u}^H \tilde{\boldsymbol{h}}_{n_2u}, \qquad \text{for } u \in \{a, b\}, \tag{16}$$

as the inner product of the residual at each location and that from the neighbouring locations. Here, $\mathcal{U}(n_1)$ is the nearest neighbours of the n_1-th Alice-Bob pair.

V. Datasets

To validate the proposed methodology, we perform experiments on synthetic datasets using the Quadriga channel simulator and real experimental datasets collected from Nokia [15].

A. Quadriga synthetic dataset

We considered single-antenna legitimate nodes, referred to as Alices and a base station referred to as Bob; Alices' spatial locations are denoted by $\{\mathbf{x}_n\}_{n=1}^N$ $n = 1, \ldots, N$, where $\{x_n\}_{n=1}^N \in \mathbb{R}^L$ and L denotes the spatial dimensions considered (typically $L = 2$). We obtain the channel response at $N = 400$ equi-distant (1 m) spatial locations within a square area on the ground, between $x = 100$ and $x = 290$ and $y = -100$ and $y = 90$ and a base station located at $(x, y, z) = (0, 0, 10)$ using the "Berlin-UMa-NLOS" configuration in Quadriga channel models [14], [42], as depicted in Fig. 5. We assume that for any specific Alice, all other Alices can act as attackers (Eves).

This configuration's terrestrial Urban Macrocell parameters are extracted from measurements in Berlin, Germany. To create temporal variations in the channel, the Alices are assumed to move at a low speed of 0.5 m/s. The number of CSI snapshots per Alice is set to $M = 256$, while the carrier frequency is set to 2.68 GHz.

Let the channel function mapping the spatial locations to the $M \times 1$ CSI vectors $\{\mathbf{h}_n\}_{n=1}^N$ be denoted by $\mathcal{H} : \mathbb{R}^L \to \mathbb{R}^M$, where M is the number of snapshots in the time domain (after concatenation of the real and imaginary components into a single column vector). The CSI observations at Alice and Bob after the exchange of pilot signals can be modelled as

$$\mathbf{y}_{nu} = \mathbf{h}_n s + \mathbf{n}_{nu}, \; n = 1, \ldots, N, \; u \in \{a, b\}, \tag{17}$$

where the index a denotes an Alice, b denotes Bob; $\mathbf{n}_{na}$ and $\mathbf{n}_{nb}$ are circularly symmetric Gaussian noise variables and the pilot symbols s are drawn from the binary phase-shift keying (BPSK) constellation [43]. The zero-force CSI estimates at Alice and Bob, respectively, are denoted by $\mathbf{h}_{na} = \mathbf{y}_{na}$ and $\mathbf{h}_{nb} = \mathbf{y}_{nb}$ for $n = 1, \ldots, N$.

B. Nokia experimental dataset

A massive multiple-input multiple-output (mMIMO) channel measurement campaign was conducted on the Nokia campus in Stuttgart, Germany. The campaign area consisted of multiple roads with high buildings (15 m high approximately), acting as reflectors and blockers for the radio wave propagation. The transmit antenna array was placed on the roof-top of one of these buildings. The geometry of 64-element transmit-array was such that there were 4 rows with 16 single-polarization patch antennas, with a horizontal spacing of $\lambda/2$, and vertical spacing of λ.

The transmit antenna array transmitted 64 time-frequency orthogonal pilot signals at 2.18 GHz carrier frequency, using orthogonal frequency division multiplexing (OFDM) waveforms according to the 10 MHz LTE numerology (i.e., 600 subcarriers with 15 kHz spacing). The pilot signals have been arranged so that the sounding on 50 separate subbands (each consisting of 12 consecutive subcarriers) required 0.5 ms. Within that pilot burst period, the propagation channel was assumed to be time-invariant. The pilot bursts were sent continuously with a periodicity of 0.5 ms.

The receiver user equipment (UE) was mounted on a mobile cart and consisted of a single monopole antenna mounted at 1.5 m height, a Rohde and Schwarz TSMW receiver and a Rohde and Schwarz IQR hard disc recorder, which continuously captured the received base-band signal. Both the transmit array and the receiver were frequency synchronized via GPS. During the measurements, the receiver cart moved

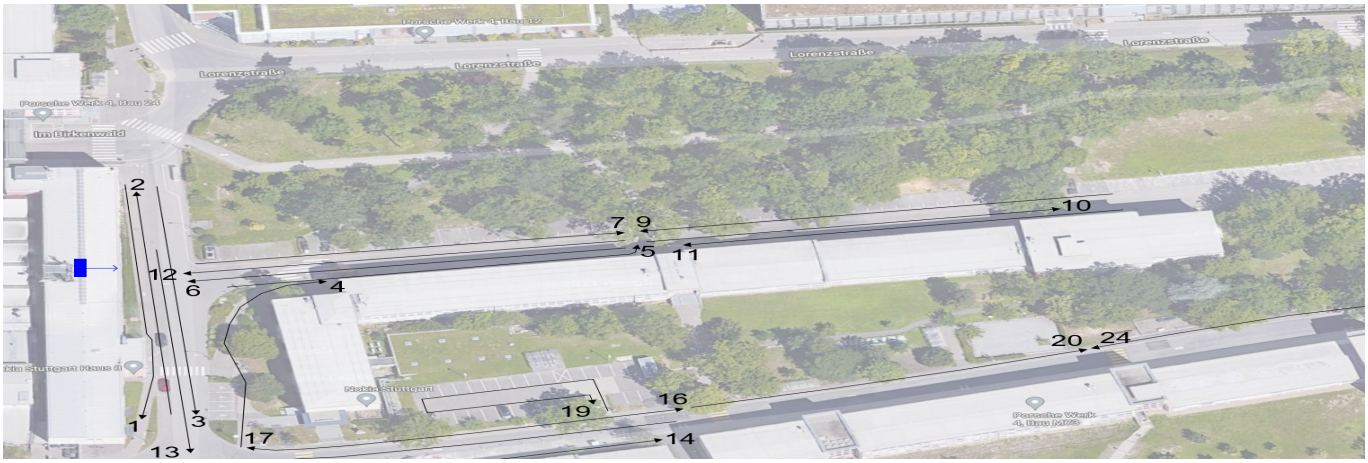


Fig. 6: Pictorial representation of the Nokia campus in Stuttgart, Germany, where the dataset was recorded. The blue bar on the left side corresponds to the location of the massive MIMO transmit antenna, placed on a rooftop, while the arrow points the antenna boresight. The black dashed lines depict the measurement tracks, along which the UE cart moved. The tracks are numbered from 1 to 24; in this work we utilized the datasets of tracks 6 and 12, which are parallel and at 1 m distance from each other [15].

along several routes at walking speed (3.6 kmph), which corresponded to a spatial channel sampling distance of less than 0.5 mm. Post-processing was used to extract, for each pilot burst and subband, the 64-dimensional CSI vector.

In this work, we used datasets on tracks 6 and 12, depicted in Fig. 6, that are parallel at a vertical distance of 1 m. In detail, we assumed that Bob is the base station and Alice walks along track-6, while Eve performs an "on the shoulder attack" and walks in parallel to Alice on track-12, i.e., legitimate and adversarial nodes are at all times 1 m away. In order to remove frequency domain (within the coherence bandwidth) and antenna domain correlations, we have downsampled the dataset. In detail, we kept the measurements from every 10th subcarrier and every 4th antenna, i.e., we used sub-sampling factors of 5 and 6, respectively.

Furthermore, as the Nokia dataset consists of only uplink data, we used alternate consecutive measurements to approximate downlink data and have further downsampled the data in the time domain, keeping every 5th channel sample. In detail, starting from sample index 1, we labeled odd index samples as uplink and even indexed samples as downlink. As a result, for each Alice and Eve, we used CSI vectors of length $M = 800$, concatenating real and imaginary parts.

It is worth noting that in practice uplink and downlink channels can be non-reciprocal due to a multitude of factors, including but not limited to: i) different dimensions of antenna arrays at Alice and Bob; ii) non linearities of power amplifiers (AM/AM and AM/PM distortions), whose characteristics are generally different in the UE and gNB; iii) antenna imperfect calibration in UE and gNB; iv) digital processing differences; v) Tx and Rx chains imbalances more generally and more importantly vi) asynchronous transmissions of the uplink and downlink in time division duplex (TDD) systems. All of these aspects need to be systematically investigated as they will impact the MP and by extension the required reconciliation rate to achieve zero FER; however, they are out of the scope this study and are left as future possible extensions of the results presented in this paper.

TABLE I: Estimated pdf parameters for the Quadriga and the NOKIA datasets.

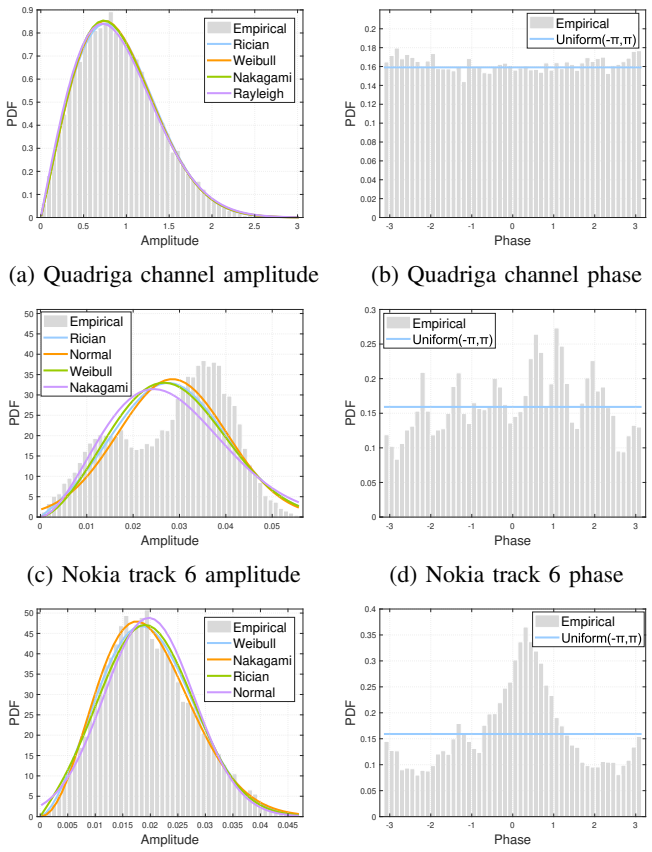
Quadriga				NOKIA track 6				NOKIA track 12			
	$\widehat{\alpha}$	$\widehat{\beta}$	p-val		$\widehat{\alpha}$	$\widehat{\beta}$	p-val		$\widehat{\alpha}$	$\widehat{\beta}$	p-val
Rician	0.56	0.59	0.11	Rician	0.02	0.01	0	Weibull	0.02	2.59	0
Weibull	1.01	2.07	0	Normal	0.03	0.01	0	Nakag.	1.52	0	0.4
Nakag.	1.05	1.01	0	Weibull	0.03	2.64	0	Rician	0.02	0.01	0
Rayleigh	0.71	-	0	Nakag.	1.34	0	0	Normal	0.02	0.01	0

Before presenting the details of the pre-processing, we discuss the statistical analysis of the two datasets.

C. Analysis of datasets

We fitted the empirical distributions of the amplitude and the phase for the Quadriga and the Nokia datasets (tracks 6 and 12) to 16 parametric probability density functions (PDFs). Table I, summarizes the estimated PDFs' parameters of the amplitude resulting in the lowest Akaike's information criterion (AIC) values, along with the p-values of the Kolmogorov-Smirnov (KS) goodness-of-fit test (the null hypothesis being that the data follow the specified distribution). Also, Fig. 7, depict the fitted PDFs of the distributions tabulated Table I on the amplitude (left column) and the fitted PDF of the uniform distribution on the phase (right column), for the Quadriga and NOKIA datasets.

According to the p-values of the KS test, the amplitude of the Quadriga-based channel follows a Rician distribution, while the phase follows a Uniform$(-\pi, \pi)$ distribution (see Fig. 7). However, the amplitude of Nokia's track 6 CSIs is bimodal (mixture distribution) and does not fit any of the chosen distributions, as shown in Fig. 7 and confirmed by the p-values of the KS test. On the other hand, for Nokia track



(a) Quadriga channel amplitude (b) Quadriga channel phase

(c) Nokia track 6 amplitude (d) Nokia track 6 phase

(e) Nokia track 12 amplitude (f) Nokia track 12 phase

Fig. 7: Fitting of the underlying channel distribution of the amplitude and the phase.

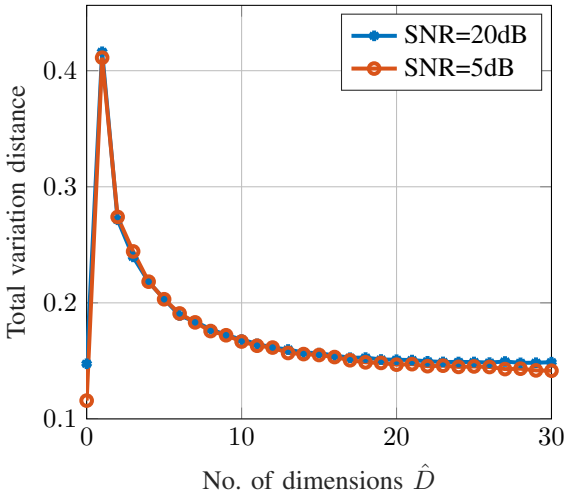
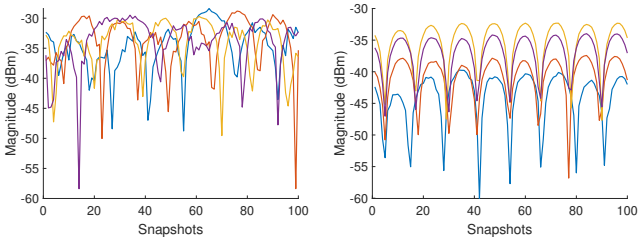


Fig. 8: TVD vs $\widehat{D}$.



(a) Original signals (b) Predictable components

Fig. 9: Separability of 6 neighbours for the original signal and the $\widehat{D} = 1$ PC for SNR= 20 dB.

12, the p-value of the Nakagami(1.52,0.0004) distribution (p-value $= 0.4$) implies that the channel is likely to follow the above distribution while the phase is uniformly distributed. In conclusion, Table I and Fig. 7 indicate that real datasets might not be well fitted to any of the usual distributions customarily used for the evaluation of SKG rates in literature, as is the case for track 6. In the future, we will consider non-parametric analysis, mixture distributions and the use of generative models for fitting real datasets.

VI. Numerical Results

In this section, we provide a comprehensive evaluation of the proposed approaches using both synthetic and real datasets. We first begin with results on maximising the separability of CSI fingerprints. We analyze the variation of TVD with respect to $\widehat{D}$. Since the results for CSI-based authentication are similar for all proposed approaches, we present them solely for the PCA for conciseness.

Subsequently, in the second half of the section, we turn our attention to unpredictable components used for SKG. We examine the reciprocity versus correlation trade-off of the PCA, KPCA, and AE by analyzing two key metrics: the average CC between locations and their nearest neighbours, and the average MP between Alice and Bob. To investigate this trade-off, we plot the variations of these metrics as the pair $\tilde{D}_1, \tilde{D}_2$ changes. Furthermore, we evaluate the randomness of the generated keys using the normalized dHSIC test statistic as well as conditional min entropy estimators. Finally, the various methods are compared in terms of overall key generation rates.

A. CSI-based node authentication

In Fig. 8, for the Quadriga dataset, the average TVD between the first $\widehat{D}$ PCs at any Alice and any of her neighbours is depicted. We observe that $\widehat{D} = 1$ results in the largest value of TVD, while the point $\widehat{D} = 0$ corresponds to the original measurements. With an increase in the SNR, there is a slight increase in the TVD; with a decrease in noise, the variance of the first PCs increases, and hence the TVD decreases.

To showcase the impact of increasing TVD, in Fig. 9, we show the variation of the amplitude of the original CSI vs. time and that of the first PC vs. time for four neighbouring Alices. We observe that when compared to the original signal in Fig. 9(a), the time series corresponding to the first PC in Fig. 9(b) are clearly distinguishable. Similar results are shown for the Nokia dataset in Fig. 10 and 11, for which, notably, the increase in the separability is more accentuated.

To provide a first validation of the proposed pre-processing, we compare the performance of CSI authentication for the original signal, the first principal component (representing large-scale fading), and the residuals (representing small-scale fading), using the NOKIA dataset. For the evaluation, we

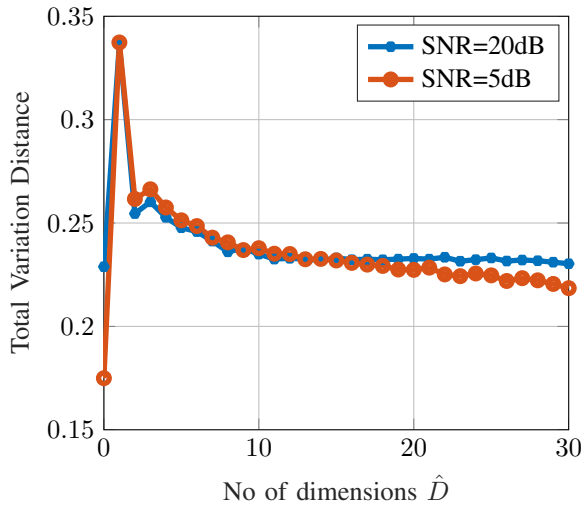
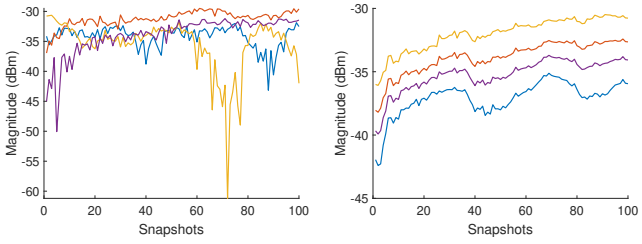


Fig. 10: TVD vs $\widehat{D}$.



(a) Original signals (b) Predictable components

Fig. 11: Original signals and first PC of 4 neighbouring Alices in the Nokia dataset for SNR= 20 dB.

employed a binary $\sqrt{M}$- nearest neighbour (NN) classifier (based on the common Euclidean distance) to distinguish between track 6 and track 12 [44]. We considered several sample sizes $M = \{250, 500, 1000, 2000, 4000\}$ and utilized 70% of each dataset for training and 30% for testing.

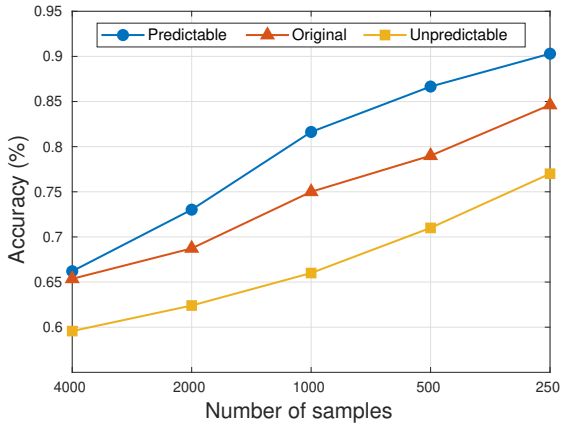
The results in Fig. 12 clearly demonstrate the superior performance of CSI-based authentication using the first PC (large-scale fading) compared to the original CSI and the small-scale fading[3]. Specifically, for sample sizes up to 1000, the large-scale fading provides an average increase in classification accuracy and f1-score of around 6% and 15% compared to the original CSI and small-scale fading. It is important to note that for larger sample sizes ($M = 2000, 4000$), the classification performance gap between the large-scale fading and the original channel is lower (up to an average of 5%). This is because the $\sqrt{M}$-NN classifier struggles to efficiently distinguish between tracks 6 and 12 due to their close proximity (1 meter). However, even at such larger sample sizes, the large-scale fading is superior to small-scale fading in terms of accuracy and f1-score.

We also evaluated the performance of CSI-based authentication using the Quadriga dataset, following the same experimental setup as before, with sample size $M = 256$ (total

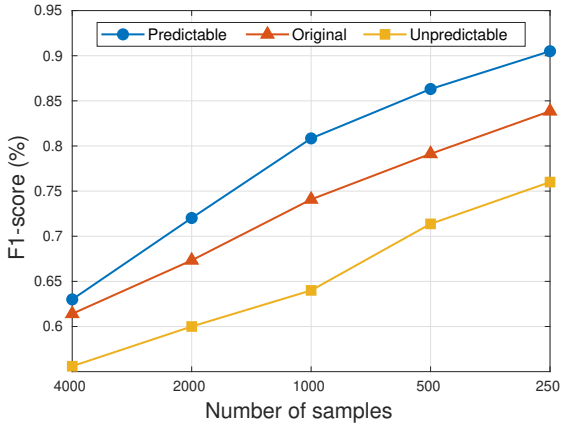
[3]The use of stochastic fading for authentication is presented here to further exemplify its unsuitability as an authentication feature.

TABLE II: Average $\sqrt{M}$-NN classification performance, in terms of classification accuracy and f1-score, for each 2 neighbouring Alices in the Quadriga dataset for SNR=20 dB for $M = 256$.

	Accuracy	f1-score
Predictable	0.793	0.7582
Original	0.5629	0.5901
Unpredictable	0.5752	0.5986



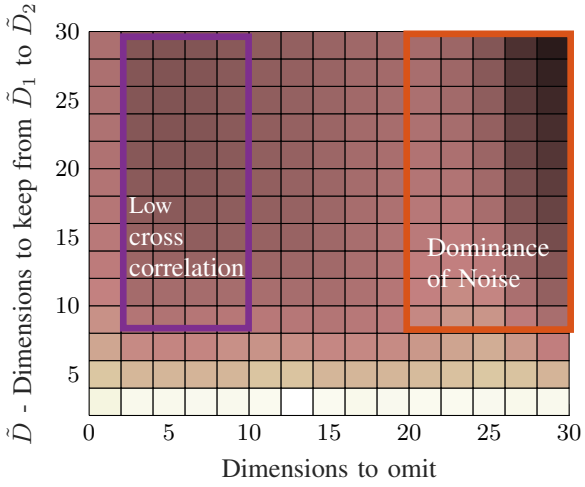
(a) Average accuracy, for several sample sizes M



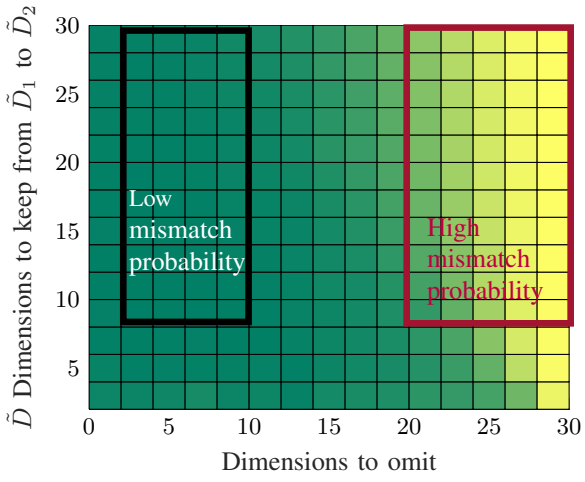
(b) Average f1-score, for several sample sizes M

Fig. 12: Average a) accuracy and b) f1-score of the binary $\sqrt{M}$-NN classifier between track 6 and track 12 of the NOKIA dataset, for the original signal, the first PC (predictable) and the residuals (unpredictable).

number of snapshots per location). The results, presented in Table II, demonstrate that accuracy is significantly improved when keeping only the first PC compared to the original CSI and the small-scale fading. On average, when keeping only the first PC leads to a 20% higher classification accuracy and a 15% higher f1-score when distinguishing between two neighbouring Alices. The table provides an overview of the average $\sqrt{M}$-NN classification performance in terms of accuracy and f1-score for each pair of neighbooring Alices in the Quadriga dataset at an SNR of 20 dB. The predictable component achieves an accuracy of 0.793 and an f1-score of 0.7582, outperforming the original channel (accuracy: 0.5629, f1-score: 0.5901) and the unpredictable component (accuracy:

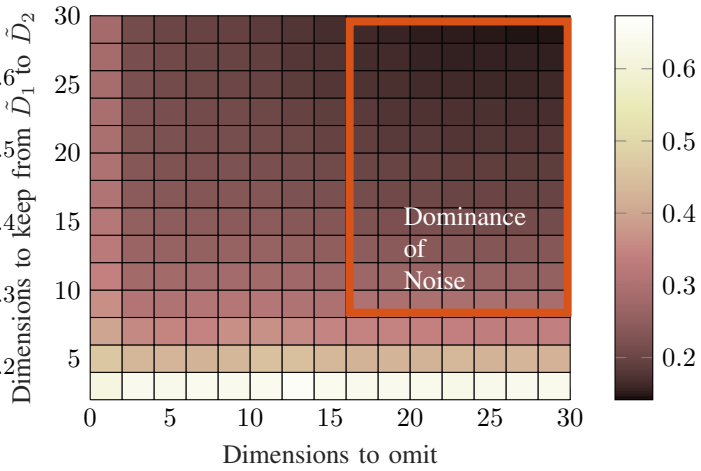


(a) Average CC (Original= 0.30)

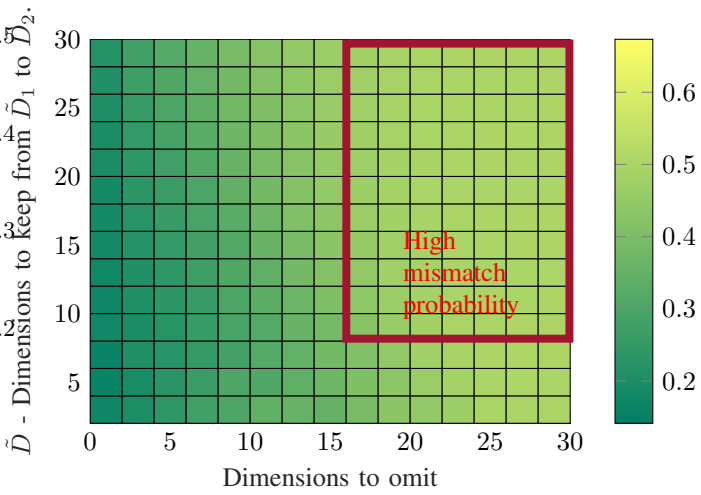


(b) Average MP

Fig. 13: Trade-off between CC and MP for SNR = 20 dB for the Quadriga dataset. Darker colours indicate lower values.



(a) Average CC (Original= 0.21)



(b) Average MP

Fig. 14: Trade-off between CC and MP for SNR = 5 dB for the Quadriga dataset.

0.5752, f1-score: 0.5986).

B. Secret key generation

1) PCA: First, we study the effect of pre-processing using PCA for SNR = 20 dB in Fig. 13, starting with the Quadriga dataset. Figs. 13(a), and (b) illustrate the variation of two metrics: i) the average CC between the locations and their nearest neighbours; and ii) the average MP between the Alices and Bob, respectively, with respect to the variation in the pair $\{\tilde{D}_1, \tilde{D}_2\}$ in steps of 2. With no pre-processing, the average CC is approximately 0.30. However, with a sufficient number of dimensions $\tilde{D}_1 - \tilde{D}_2$ retained, an increase in the number of dimensions omitted $\tilde{D}_1 - 1$ results in a decrease in the CC. Specifically, for $\tilde{D}_1 = 2$ and $\tilde{D}_2 = 20$, we observe a drop in the CC to 0.18, with no significant increase in MP.

We posit that this range of PCs captures sufficiently well small scale fading terms. This regime is indicated as "Dominance of uncorrelated components" in Fig. 13(a) while the corresponding region is referred to as "Low Mismatch Probability" in Fig. 13(b). Note that the drop in CC is more pronounced beyond $\tilde{D}_1 = 14$, beyond which noise becomes dominant, resulting in an increase of the MP. This regime is referred to as "Dominance of Noise" in Fig. 13(a). The corresponding region is marked as "High Mismatch Probability" in Fig. 13(b). In Fig. 14, the trade-off between CC and MP is shown for SNR= 5 dB. As expected, with a decrease in SNR, the effect of noise is more pronounced. Therefore, the regime of noise dominance and high MP is seen even at $\tilde{D}_1 = 10$. An important conclusion of this analysis is that for low SNRs it is possible to omit any pre-processing to avoid compromising the MP.

A trend similar to CC is observed for $\overline{\Delta}$ in Fig. 15, especially for higher values of $\tilde{D}_1$ indicating likely independence. On the other hand, $\overline{\Delta}$ for $\tilde{D}_1 < 8$ does not follow the CC drop, indicating the limitations of CC compared to $\overline{\Delta}$ to capture dependence. For example, omitting the first 6 PCs may guarantee a low CC but not a significant decrease in Δ and statistical independence. The impact of the observation vector length on $\overline{\Delta}$ will be investigated in detail in future work.

Next, we present results for the Nokia dataset starting with SNR = 20 dB in Fig. 16. With no pre-processing, the average CC is approximately 0.38. However, with a sufficient

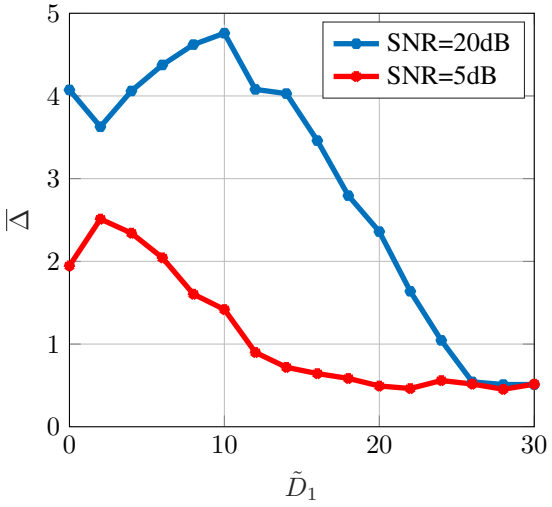


Fig. 15: Evolution of $\bar{\Delta}$ with $\tilde{D}_1$ for $\tilde{D}_2 = 30$ for the Quadriga dataset.

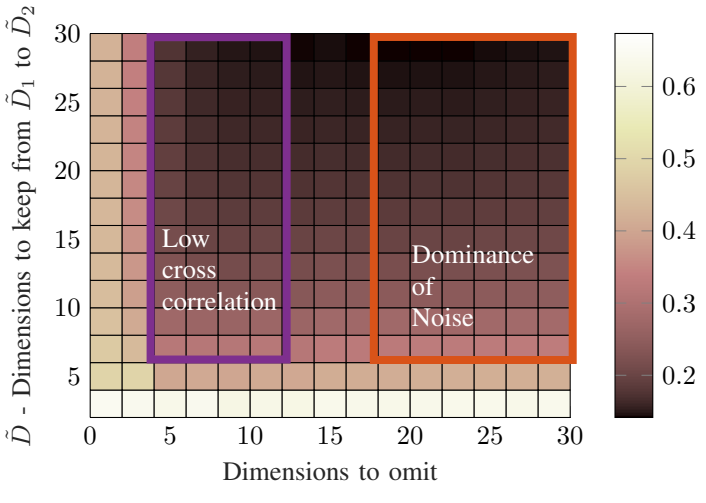
TABLE III: KPCA results for the Quadriga dataset, considering SNR=$\{5, 20\}$ dB. $\widehat{D} = 0$ denotes no pre-rpocessing.

	$\widehat{D}$	0	1	2	3	4	5	6	7	8	9	10
SNR=5dB	$\overline{\Delta}$	1.93	1.7	0.75	0.66	0.69	0.61	0.53	0.54	0.49	0.49	0.43
	MP	0.36	0.38	0.4	0.41	0.41	0.42	0.43	0.43	0.43	0.43	0.43
SNR=20dB	$\overline{\Delta}$	4.07	3.93	3.11	3.04	3.21	3.2	3.16	3.04	3.05	3.04	2.93
	MP	0.14	0.14	0.15	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16

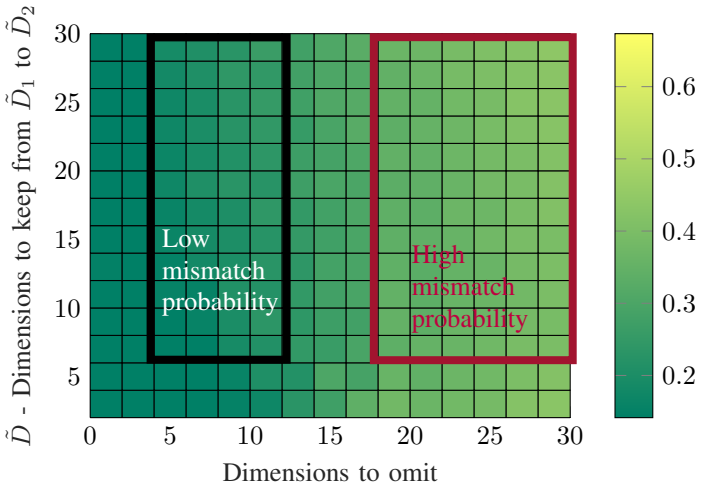
number of dimensions retained, an increase in the number of dimensions omitted decreases the CC. Specifically, for $\tilde{D}_1 = 6$ and $\tilde{D}_2 = 30$, we observe a drop in the CC to 0.15, with no significant increase in MP. As in the Quadriga dataset, we posit that this is the regime in which the predominant large scale predictable components have been removed, and the small scale fading components have been retained. Importantly, a trend similar to CC is observed in the average $\overline{\Delta}$ in Fig. 17, especially for $\tilde{D}_1 \geq 4$ after which value the dependence level collapses. Finally, similarly to Quadriga, for a low SNR=5 dB, any pre-processing would induce high MP and is therefore advised to be omitted.

2) KPCA: Next, we evaluate the performance of KPCA. Recall that, in this case, we only apply the parameter $\widehat{D}$ to derive the residuals that will be used as SKG seeds. Tables III and IV indicate that an increase in $\widehat{D}$ leads to lower values of $\overline{\Delta}$ and higher values of MP. More precisely, focusing on Table III, $\overline{\Delta}$ undergoes a significant decrease for $\widehat{D} \geqslant 2$, leading to a slight increase in MP. Moreover, comparing the outcomes of the PCA and the KPCA, KPCA seems to lead to a "faster" decrease of $\overline{\Delta}$. The results provided in Table IV concern the NOKIA dataset and similarly show a significant decrease in $\overline{\Delta}$; overall, compared to PCA, KPCA seems slightly more efficient in decreasing dependencies.

3) AE: The layers and the activation function of the AE are given in Table V and follow [45]. For brevity, the AE with



(a) Average CC (Original= 0.38)



(b) Average MP

Fig. 16: Trade-off between CC and MP for SNR $= 20$ dB for the Nokia dataset.

TABLE IV: KPCA results for the NOKIA dataset, considering SNR=$\{5, 20\}$ dB.

	$\widehat{D}$	0	1	2	3	4	5	6	7	8	9	10
SNR=5dB	Residual-$\overline{\Delta}$	10.22	4.37	2.74	1.15	0.83	0.88	0.71	0.73	0.72	0.74	0.73
	MP	0.39	0.39	0.46	0.45	0.44	0.45	0.46	0.44	0.46	0.45	0.46
SNR=20dB	Residual-$\overline{\Delta}$	32.11	25.13	17.78	10.8	9.7	10.6	9.6	9.34	9.55	9.67	9.74
	MP	0.1	0.14	0.16	0.15	0.14	0.14	0.16	0.14	0.2	0.17	0.22

MSE loss function is referred to as AE1 and that with dot-product loss function is referred to as AE2. The input to the AE2 is formed by grouping the 200×1 CSI vector (100 real and 100 imaginary) of each spatial location with 200×1 long CSI vector from each of the 8 nearest neighbours surrounding the location. In other words, the dimension at the input and the output is 400×1. This ensures that the loss function can minimize the correlation between the users while minimizing the reconstruction error between the input and the output.

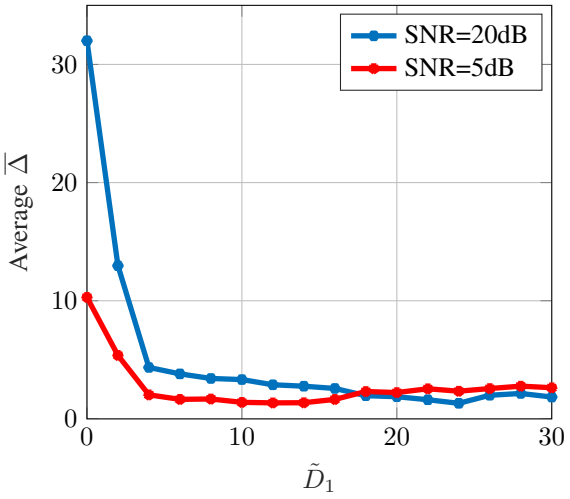


Fig. 17: Evolution of $\overline{\Delta}$ with $\tilde{D}_1$ for $\tilde{D}_2 = 30$ for the Nokia dataset.

TABLE V: The layers and activation function for AE1. For AE2 the only change is that the dimensions of the input and the output layers are 400.

Layer	Dimensions	Activation
Input	200	Linear
1	100	tanh
2	50	softplus
3	20	tanh
Intermediate	$\widehat{D}$	linear
4	20	relu
5	50	softplus
6	100	tanh
Output	200	Linear

Two types of training are possible. Either Bob and the set of Alices train separate AEs with their local datasets in localized training, or, Bob trains a global AE whose parameters are distributed to the Alices in centralized training.

From Tables VI and VII, note that, as in the case of PCA, the lower the SNR, the lower the CC and the higher the MP. Moreover, with an increase in the encoding dimensions $\widehat{D}$, the AE has more freedom to represent the predictable components. Therefore, with an increase in $\widehat{D}$, we observe a drop in the CC. For the Quadriga dataset, AE2 achieves a CC of 0.22 for $\widehat{D} = 8$ and SNR$=$ 20 dB for the residual components, without a significant increase in MP for centralized training. This CC is almost equal to what the PCA achieves for the Quadriga dataset $\tilde{D}_1 = 2$ and $\tilde{D}_2 = 20$. For the Nokia dataset, AE2 achieves a CC of 0.07 which is smaller than what PCA achieves in all cases.

Also, the $\overline{\Delta}$ of the residuals shows a significant decrease from that of the original components, especially for SNR$=$ 20 dB. Observe that in AE2, since the loss function is the dot product between residuals instead of the MSE, we can observe a significant drop in CC of the residuals for AE2 compared to AE1. However, this is accompanied by an increase in the MP, especially for localised training. Also, as expected, centralised training results in a much lower MP when compared to localised training. A very similar trend is observable in the case of the Nokia dataset also. In this case, the residual CC for $\widehat{D} = 8$ and SNR$=$ 20 dB, is much lower than what PCA attains for the same data set for $\tilde{D}_1 = 6$ and $\tilde{D}_2 = 30$.

TABLE VI: AE: Key results for Quadriga.

$\widehat{D}$	1				8			
SNR (dB)	5		20		5		20	
AE type	$AE1$	$AE2$	$AE1$	$AE2$	$AE1$	$AE2$	$AE1$	$AE2$
Original-CC	0.21	0.21	0.30	0.30	0.21	0.21	0.30	0.30
Residual-CC	0.20	0.19	0.27	0.24	0.18	0.15	0.25	0.21
Original-$\overline{\Delta}$	2.18	2.18	4.5	4.5	2.18	2.18	4.5	4.5
Residual-$\overline{\Delta}$	1.57	1.11	3.38	2.8	0.98	0.65	2.65	2.6
MP for Localized	0.37	0.39	0.19	0.27	0.39	0.45	0.17	0.36
MP for Centralized	0.35	0.36	0.14	0.15	0.34	0.40	0.15	0.15

TABLE VII: AE: Key results for Nokia dataset.

$\widehat{D}$	1				8			
SNR (dB)	5		20		5		20	
AE type	$AE1$	$AE2$	$AE1$	$AE2$	$AE1$	$AE2$	$AE1$	$AE2$
Original-CC	0.30	0.30	0.39	0.39	0.30	0.30	0.39	0.39
Residual-CC	0.13	0.11	0.25	0.17	0.04	0.05	0.10	0.07
Original-$\overline{\Delta}$	15.7	15.7	33.1	33.1	15.7	15.7	33.1	33.1
Residual-$\overline{\Delta}$	2.51	2.17	11.2	8.81	0.14	1.20	1.92	4.16
MP for Centralised	0.43	0.34	0.15	0.13	0.49	0.36	0.29	0.13
MP for Localised	0.47	0.44	0.37	0.40	0.49	0.42	0.36	0.34

TABLE VIII: Evaluation of conditional min-entropy and leakage using FBLEAU numerical estimator [13]. The table shows average values per bit.

	Min entropy	Leakage	Conditional min entropy
No-pre-processing	0.9942	0.1680	0.8262
PCA	0.9955	0.1008	0.8947
AE2	0.9990	0.0189	0.9801

C. Randomness and key-rate analysis of the pre-processing schemes

In the previous subsections, we analyzed the CC and MP trends of different schemes. However, it is also crucial to discuss the unpredictability of binary sequences at both the pre-processing and final stages using the conditional min-entropy and the leakage as metrics. In this subsection, we present the results of the proof-of-concept analysis of these metrics using only PCA and AE2 schemes, for compactness of presentation.

We estimate the conditional min-entropy and leakage at the input of the reconciliation decoder, with and without pre-processing, using the FBLEAU ML-based estimator [13] and the results are displayed in Table VIII[4]. The evaluation indicates that although the original data has high entropy, considerable leakage is observed if no pre-processing is applied. The results show that applying PCA can decrease the leakage and retain min-entropy. Additionally, we can observe that AE2 provides a further improvement by reducing the leakage to Eve. These results demonstrate that the pre-processing techniques proposed in this study can indeed reduce information leakage to neighbouring nodes, leading to larger values of conditional min-entropy and therefore lower hashing rates. These findings are aligned with the decrease seen in the normalized dHSIC test statistic and provide further proof of the decrease in dependencies of observed CSI vectors in neighbouring locations using the proposed pre-processing.

[4]We note in passing that if no reconciliation is used then the theoretically achievable secret key rate cannot be attained, resulting in sub-optimal implementations.

Next, to evaluate the final key rate expressed in Eq. (6), we perform information reconciliation using cyclic redundancy check (CRC) aided polar codes with a list size of 128 and a block length of 512 bits (considered a frame) over the quantized data [46]. In our setup, Alice encodes the channel observations y_a into u_a denoted by $u_a = y_a G_n$, where a generator matrix G_n for polar codes with a blocklength of $n = 2^m$. Alice sends a syndrome signal, s_a, through the channel, consisting of 11 CRC bits and the remaining high-entropy bits from u_a, determined by an entropy-based selection criterion. Bob's goal is to estimate Alice's channel observations, $\hat{y}_a$, using his own channel observations, y_b, and the received syndrome signal, s_a. Bob employs CRC-aided successive cancellation list decoding for this task. The frame error rate (FER) quantifies the probability of $Pr(y_a \neq \hat{y}_a)$. After Bob estimates Alice's channel observations, both Alice and Bob can independently generate secret keys from y_a.

The FER for different coding rates are depicted in Fig. 18. The coding rate in the simulations ranges from 0.3 to 0.95 with a step size of 0.05. Please note that the zero FER code rates have not been marked in the plot due to the logarithmic scale for the FER, i.e., for no pre-processing the minimum code rate to achieve zero FER is 0.35, while for PCA and AE it is 0.3 and 0.75, respectively. Note that as the MP of the quantized sequences at Bob and Alice increases, lower rate reconciliation is needed to achieve SKG without any key disagreement. Nevertheless, to determine the final key rate the amount of information leakage needs also be taken into account.

With respect to achieving zero FER when using short code-length reconciliation, in practice, the success of the information reconciliation step can be verified by transmitting parity bits. If the derived result from the information reconciliation does not match the parity check bits, Bob and Alice can request extra syndrome data from each other to finally obtain a sequence that satisfies the parity check bits. It is important to note that the mentioned parity bits are independent from the CRC bits used in the CRC-aided polar code and these extra parity checks would reduce the SKG rate further. A discussion on further parity checks is beyond the scope of the paper.

Table IX shows the bit MP of the generated sequence for each scenario and the information reconciliation coding rate that allows reliable reconciliation (zero FER in SKG), assuming that a key disagreement rate (tends to) zero. The bit MP for the original data is 0.0174 and with a coding rate of 0.35, the mismatched bits can be corrected. After applying PCA, the bit mismatch probability increases to 0.0318, which can be recovered with a coding rate of 0.3 or 0.25. On the other hand, using autoencoders (AE) significantly improves the bit mismatch probability to 0.0038, and with a coding rate of 0.75, they can be recovered. The fact that AE is non-linear provides more degrees of freedom compared to PCA, which is linear, in terms of isolating entropy-rich, reciprocal components with low information leakage to nearby nodes.

Based on the results in Tables VIII and IX, we also evaluate the overall key generation rate in bits/sec/Hz, accounting for the whole SKG chain, given by the product of the conditional min-entropy and the reconciliation code rate. The resulting

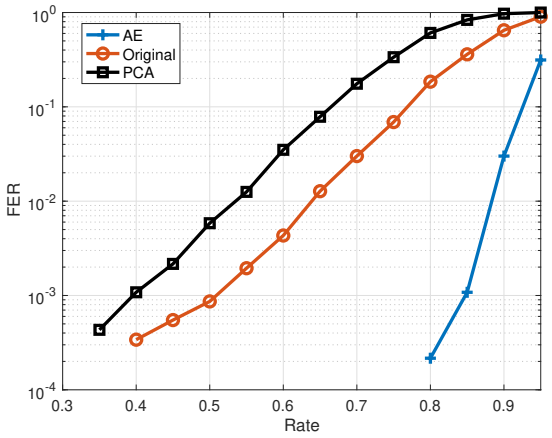


Fig. 18: The FER results for information reconciliation with different rates.

TABLE IX: The bit mismatch probability, maximum information reconciliation code rate with zero key disagreement, and the key generation rate for Nokia dataset.

Data	Bit mismatch Prob.	code rate	SKG rate (bits/sec//Hz)
Original	0.0174	0.35	0.33
PCA	0.0318	0.25 - 0.3	0.223
AE	0.0038	0.75	0.735

values are given in the last column of Table IX. The number of bits required to generate a key of size $|\mathbf{r}|$ can be evaluated as $|\mathbf{r}|/(\text{conditional min-entropy} \times \text{code rate})$ (note that this quantity represents the input size for privacy amplification, considering $|\mathbf{r}|$ to be the output size).

In this work we maintain a consistent output key size of 256 bits (that can be used for example with the advanced encryption standard (AES) with key size k=256). Table IX shows the corresponding key generation rate where the higher the key generation rate, the faster we can generate keys. Our analysis reveals that AE2 provides the highest key generation rate at 0.73 b/s/Hz, while PCA offers 0.22 b/s/Hz, which is below that achieved without pre-processing. Our previous conclusion that nonlinear pre-processing is advantageous (i.e., AE as opposed to PCA) carries over for the overall achievable key rate. Looking at the bigger picture, to obtain 256 key bits while satisfying (6), longer input sequences are required without pre-processing or when using PCA compared to AE.

After determining the maximum allowable size of the final key, we use a one-way collision-resistant compression function to compress the sequences to the desired size (i.e., 256 bits). Commonly used compression functions for this purpose are universal hash functions, while in this work we propose the use of cryptographic hashing such as SHA256, to ensure that the generated keys are maximum entropy and resistant to brute-force attacks at the input of the hash function.

Furthermore, to verify the randomness of the generated keys, we subject them to tests from the NIST randomness test suite [33]. These tests evaluate various aspects of randomness, such as uniformity, independence and unpredictability, to ensure that the generated keys are of high quality and meet the necessary security standards. In this work, we consider a subset of these tests as shown below since some tests require size larger than 10^6 bits and are not practical for the current

study [47]. The results of the tests are provided in Table X, where the success rate and average p-value of each test are reported. The results are averaged over the two pre-processing mechanisms (PCA and AE). We can see that the average success rate approaches one, hence, the generated keys pass comfortably the NIST tests.

TABLE X: Randomness evaluation using the NIST test suite from [33].

Test	Success rate	Average p-value
Frequency (monobit) test	0.9892	0.4876
Frequency within a block test	0.9922	0.4987
Runs test	0.9916	0.4817
Longest run of ones in a block test	0.9915	0.5085
Serial test	0.9804	0.4943
Cumulative sum test	0.9845	0.5092

D. Comparison of the pre-processing schemes

Summarizing the three schemes, the study indicates that for PCA omitting the first few PCs (between 2 and 14) leads to a decrease in CC without significantly increasing the MP. KPCA, on the other hand, performs better with higher values of $\widehat{D}$, which increases the MP but decreases the average dependence level. Finally, AE1 and AE2 perform better than PCA and KPCA in terms of both CC and MP. Furthermore, AE2 outperforms AE1 in terms of residual-CC, $\overline{\Delta}$, and MP for both localized and centralized training methods and achieves higher conditional min-entropy and less leakage compared to PCA. This result indicates that AE2 provides better randomness and secrecy of the generated key.

Also, the success rate and average p-value of the NIST tests are quite high, indicating that the generated keys are of high quality and meet the necessary security standards. Finally, AE2 significantly outperforms the original representation and PCA pre-processing in terms of BER and information reconciliation coding rate. Specifically, AE2 achieves a significantly lower BER with a higher information reconciliation coding rate, leading to a higher key generation rate. This suggests that AEs can provide better performance than PCA or KPCA.

In terms of computational complexity, that of PCA can be broken down into three main steps: computing the covariance matrix, computing the eigendecomposition of the covariance matrix, and projecting the data onto the eigenvectors. However, the computational complexity of PCA is dominated by the eigendecomposition step, which has a complexity of $\mathcal{O}(M^3)$, assuming a standard algorithm such as the QR algorithm or power iteration.

The computational complexity of an AE can be estimated by the number of operations required to perform a forward pass through the network. This can be estimated by counting the number of operations required to compute the matrix multiplications and activations in each layer. The total number of operations required for a forward pass through AE1 is approximately $31590 + 21\widehat{D}$. For AE2, the only difference is the dimensions of the input and output layers, which are both 400 instead of 200. Therefore, the number of operations required for a forward pass through AE2 is approximately $67570 + 21\widehat{D}$. In order to decrease the complexity and the corresponding hardware requirements, iterative pruning and quantization techniques can be investigated. For example, after initial training, one can use the Tensorflow API called Tensorflow Model Optimization that eliminates the smallest weights at the end of every training step following a polynomial decay schedule [48].

VII. Discussion and Conclusions

In this paper, we focused on dimensionality reduction techniques as pre-possessing for PLS. To demonstrate the effectiveness of our approach, we built and evaluated pre-processing approaches using PCA and AE for disentangling predictable from unpredictable components of observed CSI vectors. This allowed for the simultaneous use of CSI for authentication and key distillation. We also discussed the trade-off between correlations and dependencies at different locations, that can be extended to include time, frequency, and antenna domains, and the importance of reciprocity for the unpredictable components used in SKG.

We proposed the use of TVD as a separability measure of empirical fingerprints and used four different metrics for statistical dependence to systematize pre-processing criteria for SKG. Our evaluation showed that PCA was a straightforward approach for disentangling large from small scale fading terms in observed CSI, while KPCA and AE were shown to increase performance by capturing nonlinear structures, at the cost of explainability.

Some potential areas for future exploration in incorporating PLS in 6G security protocols could include:

- Further investigation of time, frequency, and antenna domain dependencies, potentially utilizing time domain separation techniques (preliminary results using Kalman filters have provided promising performance).
- Extensive evaluation and comparison with a range of ML-based power-domain decomposition techniques, such as independent component analysis (ICA), support vector machines (SVD), and convolutional neural networks (CNNs).
- Integration of more sophisticated AE implementations, such as those incorporating dHSIC or both dHSIC and MP into the loss function.
- Development of precise thresholds for the selection of appropriate CSI decomposition, considering both CC and MP values.
- Study of key mismatch due to various imperfections in the communication system, non-reciprocity between Tx and Rx using different size antenna arrays, phase and amplitude offsets, noise, interference, and other impairments that will impact reciprocity.

In conclusion, our proposed approach demonstrates the effectiveness of incorporating PLS technologies in 6G security protocols while providing strong security guarantees. The future explorations will provide more insight into the potential applications and performance of the proposed approach to enhance 6G security protocols.

Acknowledgement

The authors would like to thank S. Wesemann, G. Kaltbeitzel, D. Wiegner, M. Kinzler, S. Merk and S. Woerner from

Nokia for sharing the channel measurements.

REFERENCES

[1] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.

[2] V. Mavroeidis, K. Vishi, M. D. Zych, and A. Jøsang, "The impact of quantum computing on present cryptography," *Int. J. of Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 405–414, 2018.

[3] L. Chen, L. Chen, S. Jordan, Y.-K. Liu, D. Moody, R. Peralta, R. Perlner, and D. Smith-Tone, "Report on post-quantum cryptography," U.S. Dept. Commerce, Nat. Inst. Standards Technol. (NIST), Gaithersburg, MD, USA, Rep. NISTIR 8105, 2016.

[4] D. Karatzas, A. Chorti, N. M. White, and C. J. Harris, "Teaching old sensors new tricks: Archetypes of intelligence," *IEEE Sens. J.*, vol. 7, no. 5, pp. 868–881, 2007.

[5] A. Chorti, A. N. Barreto, S. Kopsel, M. Zoli, M. Chafii, P. Sehier, G. Fettweis, and H. Poor, "Context-aware security for 6G wireless, the role of physical layer security," *IEEE Commun. Standards Mag.*, vol. 6, no. 1, pp. 102–108, 2022.

[6] A. N. Barreto, S. Köpsell, A. Chorti, B. Poettering, J. Jelitto, J. Hesse, J. Boole, K. Rieck, M. Kountouris, D. Singelee, and K. Ashwinee, "Towards intelligent context-aware 6G security," *arXiv:2112.09411*.

[7] I. I. Security, "Security and Privacy, International Network Generations Roadmap (INGR) - 2021 Edition."

[8] A. Chorti and H. V. Poor, "Achievable secrecy rates in physical layer secure systems with a helping interferer," in *Proc. Int. Conf. Comput. Inf. Commun. (ICNC)*, 2012, pp. 18–22.

[9] M. Mitev, M. Shakiba-Herfeh, A. Chorti, M. Reed, and S. Baghaee, "A physical layer, zero-round-trip-time, multifactor authentication protocol," *IEEE Access*, vol. 10, pp. 74 555–74 571, 2022.

[10] M. Srinivasan, S. Skaperas, and A. Chorti, "On the use of CSI for the generation of RF fingerprints and secret keys," *in Proc. 25th Int. ITG Workshop on Smart Ant. (WSA)*, 2021.

[11] M. Srinivasan, S. Skaperas, M. S. Herfeh, and A. Chorti, "Joint localization-based node authentication and secret key generation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, South Korea, May 2022, pp. 32–37.

[12] N. Pfister, B. Buhlmann, and J. P. Scholkopf, "Kernel-based tests for joint independence," *J. Royal Statist. Soc.: Ser. B (Statist. Methodology)*, vol. 80, no. 1, pp. 5–31, 2018.

[13] G. Cherubin, K. Chatzikokolakis, and C. Palamidessi, "F-BLEAU: Fast black-box leakage estimation," in *Proc. IEEE Symp. Secur. Privacy (SP), San Francisco, CA, USA*, May 2019.

[14] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "Quadriga: A 3-d multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, 2014.

[15] M. K. Shehzad, L. Rose, S. Wesemann, and M. Assaad, "ML-based massive MIMO channel prediction: Does it work on real-world data?" *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 811–815, 2022.

[16] M. Shakiba-Herfeh, A. Chorti, and H. Vincent Poor, *Physical Layer Security: Authentication, Integrity, and Confidentiality*. Cham, Switzerland: Springer, 2021.

[17] X. Wang, P. Hao, and L. Hanzo, "Physical-layer authentication for wireless security enhancement: current challenges and future developments," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 152–158, 2016.

[18] Q. Li, H. Fan, W. Sun, J. Li, L. Chen, and Z. Liu, "Fingerprints in the air: Unique identification of wireless devices using RF RSS fingerprints," *IEEE Sensors J.*, vol. 17, no. 11, pp. 3568–3579, 2017.

[19] H. Fang, X. Wang, and L. Hanzo, "Learning-aided physical layer authentication as an intelligent process," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2260–2273, 2018.

[20] E. Meckes, *The Random Matrix Theory of the Classical Compact Groups*, ser. Cambridge Tracts in Mathematics. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[21] H. Liu, Y. Wang, J. Liu, J. Yang, Y. Chen, and H. V. Poor, "Authenticating users through fine-grained channel information," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 251–264, 2017.

[22] J. Zhang, G. Li, A. Marshall, A. Hu, and L. Hanzo, "A new frontier for iot security emerging from three decades of key generation relying on wireless channels," *IEEE Access*, vol. 8, pp. 138 406–138 446, 2020.

[23] A. Goldsmith, *Wireless communications*. Cambridge, U.K., Cambridge Univ. Press, 2005.

[24] M. Edman, A. Kiayias, Q. Tang, and B. Yener, "On the security of key extraction from measuring physical quantities," *IEEE Trans. Inf. Forensics Sec*, vol. 11, no. 8, pp. 1796–1806, 2016.

[25] C. Zenger, H. Vogt, J. Zimmer, A. Sezgin, and C. Paar, "The passive eavesdropper affects my channel: Secret-key rates under real-world conditions," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[26] R. Dautov and G. R. Tsouri, "Effects of passive negative correlation attack on sensors utilizing physical key extraction in indoor wireless body area networks," *IEEE Sens. Lett.*, vol. 3, no. 7, pp. 1–4, 2019.

[27] Z. Ji, Y. Zhang, Z. He, K. Lin, B. Li, P. L. Yeoh, and H. Yin, "Vulnerabilities of physical layer secret key generation against environment reconstruction based attacks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 693–697, 2020.

[28] P. Walther and T. Strufe, "Inference attacks on physical layer channel state information," in *Proc. 19th Int. Conf. Trust, Secur. and Privacy Comput. Commun. (TrustCom)*, Guangzhou, China, Dec. 2020, pp. 935–942.

[29] G. Li, A. Hu, J. Zhang, L. Peng, C. Sun, and D. Cao, "High-agreement uncorrelated secret key generation based on principal component analysis preprocessing," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 3022–3034, 2018.

[30] Y. Peng, P. Wang, W. Xiang, and Y. Li, "Secret key generation based on estimated channel state information for TDD-OFDM systems over fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5176–5186, 2017.

[31] W. Xi, C. Qian, J. Han, K. Zhao, S. Zhong, X.-Y. Li, and J. Zhao, "Instant and robust authentication and key agreement among mobile devices," in *Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, Vienna, Austria, Oct. 2016, pp. 616–627.

[32] A. Rock, "Pseudorandom number generators for cryptographic applications," Ph.D. thesis, Univ. Salzburg, Salzburg, Austria, 2005.

[33] L. Bassham, A. Rukhin, J. Soto, J. Nechvatal, M. Smid, S. Leigh, M. Levenson, M. Vangel, N. Heckert, and D. Banks, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," Special Publication Nat. Inst. Standards Technol. (SP NIST), Gaithersburg, MD, USA, Tech. Rep., 2010.

[34] M. Turan, E. Barker, J. Kelsey, K. McKay, M. Baish, and M. Boyle, "Recommendation for the entropy sources used for random bit generation," Nat. Inst. Standards Technol. (NIST), Gaithersburg, MD, USA, Special Publ., 800-90B, 2018.

[35] B. Espinoza and G. Smith, "Min-entropy as a resource," *Inf. Comput.*, vol. 226, pp. 57–75, 2013.

[36] D. R. L. Brown, "Formally assessing cryptographic entropy," *IACR Cryptol. ePrint Arch.*, p. 659, 2011. [Online]. Available: http://eprint.iacr.org/2011/659

[37] L. Reyzin, "Some notions of entropy for cryptography," in *in Proc. 5th Int. Conf. Inf. Theor. Secur. (ICITS)*, Berlin, Germany, 2011, pp. 138–142.

[38] G. Smith, "On the foundations of quantitative information flow," in *Found. Softw. Sci. Comput. Struct.*, L. de Alfaro, Ed., (Lecture Notes Comput. Sci.), vol. 5504, L. de Alfaro, Ed. Berlin, Germany: Springer, 2009, pp. 288–302.

[39] C. T. Zenger, J. Zimmer, M. Pietersz, J.-F. Posielek, and C. Paar, "Exploiting the physical environment for securing the internet of things," in *Proc. ACM New Secur. Paradigms Workshop (NSPW)*, Twente, Netherlands, Sep., 2015, p. 44–58.

[40] B. Scholkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[41] G. H. Bakir, J. Weston, and B. Scholkopf, "Learning to find pre-images," *Advances Neural inf. Proc. Syst.*, vol. 16, pp. 449–456, 2004.

[42] Available: http://www.quadriga-channel-model.de. [Online].

[43] A. Chorti, "Optimal signalling strategies and power allocation for wireless secret key generation systems in the presence of a jammer," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[44] L. Senigagliesi, M. Baldi, and E. Gambi, "Comparison of statistical and machine learning techniques for physical layer authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1506–1521, 2020.

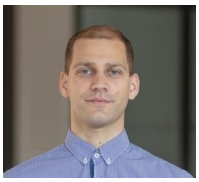
[45] C. Studer, S. Medjkouh, E. Gonultas, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[46] M. Shakiba-Herfeh and A. Chorti, "Comparison of short blocklength slepian-wolf coding for key reconciliation," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Rio de Janeiro, Brazil, 2021.

[47] J. Zhang, G. Li, A. Marshall, A. Hu, and L. Hanzo, "A new frontier for IoT security emerging from three decades of key generation relying on wireless channels," *IEEE Access*, vol. 8, pp. 138 406–138 446, 2020.
[48] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A System for Large-scale Machine Learning," in *Proc. {USENIX} Symp. Oper. Sys. Des. Imp. (OSDI)*, Savannah, GA, USA, Nov. 2016, pp. 265–283.

Muralikrishnan Srinivasan received his bachelors degree in electronics and communication engineering from College of Engineering Guindy, Anna University, India. In 2020, he received his joint masters and doctoral degree from the Indian Institute of Technology Madras after conducting his doctoral research on approximate models for generalized fading channels. In 2021, he worked as a postdoctoral researcher at ETIS, ENSEA, Cergy France, on Physical Layer Security (PLS). From 2021, he is a postdoctoral researcher at the Dept. of Electrical Engineering, Chalmers University of Technology, Sweden, working on machine learning for optical communications. He is also currently a part of the IEEE INGR pre-standardization working group for PLS. His research interests include machine learning for wireless communication systems, optical interconnects, physical layer security, massive multiple-input-multiple-output systems, aerial base stations, air-corridors, hypergeometric functions, and extreme-value theory.

Sotiris Skaperas received a B.Sc. and an M.Sc. degree in Mathematics, Aristotle University, Greece, and the Ph.D degree in 2021, from the Department of Applied Informatics, University of Macedonia, Greece. Currently, he is a post-doc researcher at Athena Research & Innovation Center and at University of Macedonia.



Miroslav Mitev obtained the Ph.D from the University of Essex, UK, with his thesis in the area of physical layer security. Next (2020), he joined the École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA) in France, where he worked in collaboration with Nokia Bell Labs on interference management. In the period 2021-2023, he was a Senior Researcher at the Barkhausen Institute, Germany. Currently, he is with Last Mile Semiconductor, Germany, working as a Wireless System Engineer. His research interests include wireless communications, physical layer security and signal processing for the IoT.

Mahdi Shakiba-Herfeh (Member, IEEE) received the B.Sc. degree from the University of Tehran, Tehran, Iran, in 2011, the M.Sc. degree from Middle East Technical University, Ankara, Turkey, in 2014, and the Ph.D. degree from Bilkent University, Ankara, in 2019. He joined ENSEA, Cergy, France, as a Postdoctoral Researcher. His research interests include various topics in information theory, wireless communications, and wireless security with a particular focus on coding techniques.

M. Karam Shehzad (Member, IEEE) is currently working as a Senior Research and Standardization Specialist at Nokia Bell-labs, Paris, France. He did his Ph.D. in electrical engineering, in 2023, at CentraleSupelec, University of Paris-Saclay, Paris, France. During his Ph.D., he was also affiliated with Nokia Bell-labs, Paris, as a Research Engineer. He received M.S. degree in electrical engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, and the B.Eng. (Hons) degree in electrical and electronic engineering from the University of Bradford, Bradford, UK. During his M.S., he also spent five months on ERASMUS+ Mobility Program at the University of Malaga, Malaga, Spain. From 2016 to 2017, he was a Research Assistant at Namal University, Mianwali, Pakistan. From November 2019 to February 2020, he served as a Research Assistant at school of electrical engineering and computer science (SEECS), NUST. His research interests include drone communication, multiple-input multiple-output communication, cognitive radio networks, and applying AI/ML to various aspects of wireless communications.

Philippe Sehier received his engineer degree from Ecole Supérieure D'électricité (Supelec) in 1984. He is a department head in Nokia Bell Labs France. His responsibilities cover fifth-generation (5G) and beyond definition, in close collaboration with standards. He has been involved in various activities in several areas of wireless telecommunications, including satellite, fixed, and mobile wireless in third- and fourth-generation technologies as a team or project manager. He has been the Nokia head of delegation of the Wimax Forum and more recently of NGMN. He has authored more than 40 publications and holds 30 patents on synchronization, channel coding, equalization, spread spectrum and transport. His current field of interest is 5G radio-access-network architectures.

Arsenia (Ersi) Chorti is a Professor at the École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA, BI Research Fellow of the Wireless Connectivity Group of the Barkhausen Institut gGmbH and a Visiting Scholar at Princeton University. Her research spans the areas of wireless communications and wireless systems security for 5G and 6G, with a particular focus on physical layer security. Current research topics include: context aware security, multi-factor authentication protocols, 5G / 6G and IoT, anomaly detection, machine learning for communications, new multiple access techniques and scheduling. She is a Senior IEEE Member, member of the IEEE INGR on Security and is Chair of the IEEE Focus Group on Physical Layer Security. She has participated in the reduction of the ITU report M.2516-0 on Future Technology Trends of Terrestrial International Mobile Telecommunications Systems Towards 2030 and beyond (sections on trustworthiness) and is a Member of the ITU Working Group on the Metaverse.