Comparison of Statistical and Machine Learning Techniques for Physical Layer Authentication

Comparison of Statistical and Machine Learning Techniques for Physical Layer Authentication

Linda Senigagliesi, Marco Baldi and Ennio Gambi

***Abstract*—In this paper we consider authentication at the physical layer, in which the authenticator aims at distinguishing a legitimate supplicant from an attacker on the basis of the characteristics of a set of parallel wireless channels, which are affected by time-varying fading. Moreover, the attacker's channel has a spatial correlation with the supplicant's one. In this setting, we assess and compare the performance achieved by different approaches under different channel conditions. We first consider the use of two different statistical decision methods, and we prove that using a large number of references (in the form of channel estimates) affected by different levels of time-varying fading is not beneficial from a security point of view. We then consider classification methods based on machine learning. In order to face the worst case scenario of an authenticator provided with no forged messages during training, we consider one-class classifiers. When instead the training set includes some forged messages, we resort to more conventional binary classifiers, considering the cases in which such messages are either labelled or not. For the latter case, we exploit clustering algorithms to label the training set. The performance of both nearest neighbor (NN) and support vector machine (SVM) classification techniques is evaluated. Through numerical examples, we show that under the same probability of false alarm, one-class classification (OCC) algorithms achieve the lowest probability of missed detection when a small spatial correlation exists between the main channel and the adversary one, while statistical methods are advantageous when the spatial correlation between the two channels is large.**

***Index Terms*—Clustering, machine learning, physical layer authentication, wireless communications.**

I. Introduction

Classical authentication protocols rely on cryptographic primitives to allow the supplicant to prove his identity to the authenticator. Although such approaches are well consolidated and implemented in real world applications, they may exhibit some limitations when adopted in new scenarios, like that of the Internet of Things (IoT). In fact, cryptographic approaches are usually characterized by a relatively large complexity, since security relies on the difficulty for a computationally-constrained attacker to break some mathematical trapdoor.

In IoT applications, huge numbers of resource-constrained cyber-physical devices are deployed, and they need to be authenticated in order to avoid impersonation and falsification attacks. In such a context, the use of classic authentication protocols based on cryptographic primitives may result cumbersome, due to computing power and memory constraints of embedded devices. Opposed to approaches based on cryptographic primitives, physical layer security (PLS) approaches exhibit some features that make them particularly suitable for IoT applications, that is:

- they do not require any assumption on the computing power of attackers;
- they only rely on the communication channel's characteristics and do not need pre-shared credentials;
- they are characterized by low complexity.

For this reason, a recent trend in the literature is focused on PLS solutions for IoT applications [1]–[4]. Low-complexity security frameworks for key generation based on PLS have also been proposed [5], together with solutions that consider adversaries having infinite computational capabilities [6] over realistic wireless channels [7].

Authentication is one of the first and most important tasks in secure communications. It aims at recognizing messages coming from a legitimate supplicant while detecting those which may be forged by a malicious attacker. At the physical layer, authentication is performed by distinguishing the source of a message based on the unique characteristics of the communication channel [8].

In this paper, we consider and design some physical layer authentication (PLA) protocols based on several decision criteria, both based on statistical hypothesis testing and machine learning approaches. We consider wireless parallel channels affected by time-varying fading to assess the performance of the PLA protocols we examine. The legitimate receiver does not know the exact realizations of the channel between the transmitter and itself, but he may know some of its statistical properties, such as the noise variance, or even not have any channel state information (CSI). In the latter case, authentication must hence be performed blindly with respect to the channel characteristics. Concerning decision methods, we start from the statistical techniques used in [9] for a flat fading wireless channel model, and consider several others decision criteria. We also start from statistical criteria based on a hypothesis testing approach [10], and their corresponding optimal attacks. Then we consider decision criteria based on machine learning, whose application to PLA has started to attract interest in the literature in the last years under different scenarios [11], [12]. For this purpose, we consider several classification methods based on machine learning, and compare the performance they achieve with that obtainable through classical statistical methods. In particular, we use

The material in this paper was presented in part at the IEEE Global Communications Conference (Globecom 2019), Waikoloa, HI, USA, Dec. 2019, and at the IEEE International Workshop on Information Forensics and Security (WIFS 2019), Delft, The Netherlands, Dec. 2019.

L. Senigagliesi, M. Baldi and E. Gambi are with Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60131 Ancona, Italy (e-mail: l.senigagliesi@staff.univpm.it, m.baldi@univpm.it, e.gambi@univpm.it).

nearest neighbor (NN) and support vector machine (SVM) algorithms, examining both their binary and one-class versions. The former have a low computational complexity with a reduced training set, which makes them attractive in resource-constrained applications, while the latter allow avoiding the search of optimal parameters in the initial phase. We perform the training phase considering both the case of an authenticator who knows exactly the source of the received setup packets and that of an uninformed authenticator. We therefore also consider clustering algorithms to establish the identity of the message sender during the initial phase. One of the points raised in [17] is the difficulty in pre-designing a precise authentication model. By exploiting predefined algorithms we can assess the performance of generic channel models, not specifically designed for the examined scheme. In order to have a shared and reproducible benchmark, we consider the use of off-the-shelf machine learning algorithms. A possible follow-up of our analysis is the design of specific classifiers for the considered application, but this is out of the scope of the present work and is left for future works.

A. Related work

Blind authentication schemes have been proposed in the last years, as in [13], where techniques of blind known-interference cancellation and differential processing are combined to implement authentication, but they rely on the sharing of a secret key between legitimate participants. Effects of time-varying fading have been studied in literature considering schemes based both on the sharing of a secret key [13] and key-less approaches [14], [15]. To the best of our knowledge, machine learning has been applied to PLA only very recently and under different conditions, for instance not considering time-varying fading channels (see [16], [12]). A review of machine-learning-aided intelligent authentication techniques proposed for 5G communications, with the definition of the requirements needed for new generation networks is provided in [17]. In [18], k-NN algorithms are used to identify electronic devices through their radio-frequency emissions, and the presence of an attacker is not explicitly considered. One-class SVM and k-means clustering have been successfully applied in a context different from the one we consider, such as in multiple input multiple output (MIMO) stationary systems [19], or for the detection of eavesdropping attacks in unmanned aerial vehicle (UAV)-aided wireless systems [20]. Besides considering scenarios different from ours, all the mentioned works also consider the use of high level cryptographic schemes during the first phase of the authentication process. Authors of [21] propose the application of an adaptive algorithm for PLA in a dynamically changing wireless environment, considering a single channel model and a series of physical layer attributes which may include CSI. Very recently, the use of adaptive neural networks to achieve adaptive authentication has been proposed in [22], also including a prototype implementation on a universal software radio peripheral (USRP) platform in presence of multipath fading. However, in [22] it is assumed that a pre-shared secret key is available during the training phase, while we exploit clustering algorithms to perform authentication without either shared secrets or previous knowledge on the legitimate transmitter. Furthermore, although the methodology proposed in [22] can be implemented on USRP platforms, the use of neural networks is not suitable for resource-constrained scenarios, as the one of interest in this work.

B. Contribution

We start from the analysis developed in [9], which considers only flat fading, and extend it to the time-varying fading case. We compare the performance obtained for the flat fading case with that achieved with time-varying fading, showing the effect of fading on authentication. Then, we consider the use of machine learning techniques to accomplish the same task in the same setting, and compare their performance with that of the statistical techniques used in [9]. We consider a training phase corresponding to the initial state of the channel, and the existence of some correlation between the main channel and the adversary's one. This models a more realistic training phase with respect to [21]. In addition, differently from [21], we also consider the use of non-supervised techniques to discriminate authentic from forged messages during the training phase, letting a clustering algorithm to decide about the nature of the received packets. We show that even in such an unfavorable condition it is possible to achieve good security performance using machine learning algorithms, especially when the attacker's channel has a low spatial correlation with the main one.

This work consolidates and extends our previous analyses reported in [23], [24] by:

- Exploiting clustering algorithms to establish the source of a message during the initial training phase, in order to avoid the need for higher layer cryptographic protocols or manual procedures.
- Comparing different machine learning techniques based on different classification principles, also considering their binary and one-class versions.
- Providing a more extensive comparative assessment between statistical and machine techniques applied to PLA under different conditions. Introducing a hybrid approach that embeds statistical metrics into machine learning algorithms. In fact, we show that statistical metrics can be used with NN algorithms to improve the ability to detect forged messages.
- Considering the more general case in which fading also affects the training phase, showing that this leads to a loss in terms of missed detection with both the analyzed approaches.

The paper is organized as follows. In Section II we introduce the authentication protocol and the channel model considered, along with the relevant performance metrics. In Section III two different decision methods based on statistical techniques are introduced. Section IV describes the attacker models, while Section V describes the one-class classification algorithms we apply. Numerical evaluations and the relevant results are reported in Section VI, after which Section VII provides some conclusive remarks.

II. System model and metrics

The following channel model is considered. A peer (Alice) has to be authenticated by an authenticator (Bob) at the presence of a malicious attacker (Eve), who aims at impersonating Alice by forging her messages. Through PLA, Bob should be able to recognize the messages coming from Alice as legitimate and to refuse the ones coming from Eve.

Transmission of a message is performed over a set of N parallel channels, which can model multi-carrier or orthogonal frequency division multiplexing (OFDM) transmission. Each channel is corrupted by additive white Gaussian noise (AWGN) and time-varying fading, which represents a standard model used in the literature to assess and compare the performance of different transmission techniques. The rapidity of channel variations clearly influences the quality of authentication, since it makes more difficult for Bob to recognize channel estimates coming from the same source. If we wish to take into account the possible occurrence of interference, we must broaden the study and include network aspects, e.g., concerning the number of nodes, the network topology and the adopted medium access control protocols. Although this may be interesting in principle, it opens the way to many degrees of freedom, and makes benchmarks susceptible to the chosen network parameters and protocols. This is clearly out of the scope of this paper, thus we leave it for future works.

Channel samples representing the channel impulse response (CIR) are collected into a vector $\mathbf{h}$ of complex numbers, whose entries are zero-mean correlated circularly symmetric complex Gaussian variables. Each channel estimate is then written as

$$\mathbf{h}^{(\mathrm{XY})} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \mathbf{R}^{(\mathrm{XY})}), \tag{1}$$

where X and Y represent a general couple of transmitter and receiver, respectively, and $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ denotes the distribution of circularly symmetric complex Gaussian random vectors (with zero mean) having covariance matrix $\mathbf{R}$.

The authentication procedure we consider is based only on channel estimates and comprises two phases, that are summarized next.

a) Phase I (training): Bob observes one or more packets with known content transmitted from Alice during a fixed interval of time. The source of the messages can be guaranteed either by exploiting higher layer protocols or through physical measures (e.g., by manually executing the setup phase). By exploiting these setup packets, Bob obtains a set of M time-correlated channel estimates which, depending on the duration of the time interval, may be subject to time-varying fading. The m-th estimate can be written as

$$\hat{\mathbf{h}}^{(\mathrm{AB})}(m) = \boldsymbol{\alpha}^{(\mathrm{I})}(m)\mathbf{h}^{(\mathrm{AB})} + \sqrt{1-\boldsymbol{\alpha}^{(\mathrm{I})2}(m)}\mathbf{w}(m) + \mathbf{w}^{(\mathrm{I})}(m), \tag{2}$$

where $\mathbf{w}^{(\mathrm{I})} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_{\mathrm{I}}^2\mathbf{I}_N)$ is a noise vector. The contribution of time-varying fading is represented by a $[1\times N]$ vector $\boldsymbol{\alpha}$, whose elements correspond to real numbers taking values in $[0, 1]$, and by a random variable $\mathbf{w}$ generated according to a Rayleigh distribution with unitary variance. We work under the hypothesis of slowly time-varying fading channels, meaning that the fading coefficient is assumed constant during transmission of each packet [25].

If during Phase I Bob collects $M > 1$ reference estimates, he can average over them in order to reduce the noise level and to obtain the average value of the fading parameter $\boldsymbol{\alpha}^{(\mathrm{I})}$, which may vary from one estimate to another. Hence he obtains the average estimate

$$\bar{\mathbf{h}}^{(\mathrm{AB})} = \frac{1}{M}\sum_{m=1}^{M}\hat{\mathbf{h}}^{(\mathrm{AB})}(m) = \bar{\boldsymbol{\alpha}}^{(\mathrm{I})}\mathbf{h}^{(\mathrm{AB})} + \sqrt{1-\bar{\boldsymbol{\alpha}}^{(\mathrm{I})2}}\bar{\mathbf{w}} + \bar{\mathbf{w}}^{(\mathrm{I})}, \tag{3}$$

where $\bar{\boldsymbol{\alpha}}^{(\mathrm{I})}$, $\bar{\mathbf{w}}$ and $\bar{\mathbf{w}}^{(\mathrm{I})}$ represent the average value of the time-varying fading and noise vectors, respectively. $\bar{\mathbf{h}}^{(\mathrm{AB})}$ can be used as a reference to classify new estimates coming from unknown sources.

b) Phase II (classification): After the training phase has been completed, Bob receives further packets without assurance that they come from Alice. Bob estimates the channel through which these new packets arrive, and exploits such an estimate to decide their source. This is done by comparing any of these estimates with the reference one obtained during the training phase. As done in [9], we suppose that in this phase Eve can forge packets on which Bob's estimate is forced to be equal to a vector $\mathbf{g}$ (plus noise).

In order for Bob to decide if a packet comes from Alice or from Eve during the classification phase, we consider and compare two approaches:

- classical statistical methods based on hypothesis testing;
- modern techniques based on machine learning.

When Bob resorts to statistical methods, it is convenient for him to use the average estimate in (3) as a benchmark, while when machine learning-based decision criteria are used he can exploit the whole training set deriving from the M channel estimates. The two approaches are described next.

A. Performance metrics

The performances achievable through the considered approaches are evaluated by measuring their probability of false alarm (FA) and of missed detection (MD). By *false alarm* we mean the event that occurs when Bob rejects a message coming from Alice, while there is a *missed detection* when he accepts a message forged by Eve. If we take a measure to reduce the probability of FA, normally this increases the probability of MD. Therefore, a trade-off between these two effects has to be found. It is important to observe that while time-varying fading negatively affects the correct authentication of legitimate signals, it plays a positive role from the attacker's point of view. In fact, as will be shown in the following, it directly influences the probability of FA, forcing Bob to accept a larger range of inputs, and thus increasing the chances for Eve that one of her forged messages is accepted as authentic.

As regards the statistical methods, analytical formulations of the two mentioned probabilities can be found and depend on the performed test. Probabilities of FA and MD resulting from the application of machine learning algorithms can instead be evaluated by exploiting the so-called *Confusion Matrix* [26]. The confusion matrix provides a comprehensive overview of

classification results. Columns represent the predicted values, while rows represent the actual values. It contains 4 kinds of entries, the true positives, the false positives, the true negatives and the false negatives. The total number of samples in the test data corresponds to the sum of these entries.

False positives (FP) represent samples coming from Eve incorrectly classified as positives, i.e. considered as authentic by Bob. This event corresponds to a MD. A large number of FPs can be due to a noisy training set and to the impact of fading in Phase I. If we define the number of negative samples classified as negative (or true negatives) as TN, the probability of MD of a classifier can be written as

$$P_{\text{MD}} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{TNR}, \tag{4}$$

where $\text{TNR} = \frac{\text{TN}}{\text{TN+FP}}$ represents the true negative rate.

False negatives (FN) are instead messages coming from Alice refused by Bob (positive samples classified as negative). This event corresponds to a FA. If we denote as TP the number of true positives (positive samples classified as positive), the probability of FA corresponds to

$$P_{\text{FA}} = \frac{\text{FN}}{\text{TP} + \text{FN}} = 1 - \text{TPR}, \tag{5}$$

where $\text{TPR} = \frac{\text{TP}}{\text{TP+FN}}$ represents the true negative rate.

Another common metric used to evaluate the performance of a classification algorithm is the accuracy. It is defined as the ratio between the number of correct predictions and the total number of instances classified or, in formulas

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{6}$$

Accuracy and other widespread performance metrics of supervised classification, however, are not always suitable for the OCC scenario, since negative data have a highly skewed distribution with regard to the target data, and could lead to misleading values. To overcome this issue, we resort to the *Geometric Mean* (g_{mean}) of accuracy (introduced in [27]), measured separately on each class; by combining the true positive rate (TPR) and the true negative rate (TNR), it is defined as

$$g_{mean} = \sqrt{\text{TPR} \cdot \text{TNR}} = \sqrt{(1 - P_{\text{FA}})(1 - P_{\text{MD}})}. \tag{7}$$

An important property of g_{mean} is that it is independent of the distribution of positive and negative samples in the test data. This allows assessing the classifier performance not only on the basis of the predominant class (as happens for the accuracy), but on both classes.

III. Statistical methods

According to classical statistical decision methods, Bob resorts to hypothesis testing [28] to decide whether the transmission was performed by Alice or not. Hypothesis testing can be applied only when the authenticator has no information about the statistics of the attacker's channel. The considered channel model is in fact focused on determining if the source of the received message is the authentic one or not, and not on recognizing who is the transmitter. The authenticator, Bob, bases the authentication only on the knowledge of Alice's channel statistics, and the event of Eve sending a message is therefore noticed only when the statistics of the new estimates differ from (2). Denoting by $\hat{\mathbf{h}}$ the channel estimated by Bob, the two hypotheses hence are

- $\mathcal{H}_0$: the message is coming from Alice. The new measured channel estimate is subject to time-varying fading and its correlation with the reference estimates collected during Phase I depends on how severely the fading affects the channel. The hypothesis $\mathcal{H}_0$ at time t can therefore be written as

$$\hat{\mathbf{h}}(t) = \boldsymbol{\alpha}^{(\text{II})}(t)\mathbf{h}^{(\text{AB})} + \sqrt{1 - \boldsymbol{\alpha}^{(\text{II})2}(t)}\mathbf{w}_F + \mathbf{w}^{(\text{II})}(t), \tag{8}$$

  where the noise vector is written as $\mathbf{w}^{(\text{II})} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma^2_{\text{II}}\mathbf{I}_N)$. In order to distinguish between the impact of time-varying fading in Phase I and Phase II, in this second case it is represented by means of the vector $\boldsymbol{\alpha}^{(\text{II})}$ and of a random variable $\mathbf{w}_F$.
- $\mathcal{H}_1$: the message is coming from Eve, and

$$\hat{\mathbf{h}}(t) = \mathbf{g} + \mathbf{w}^{(\text{II})}(t). \tag{9}$$

  Since information about Eve's channel is not available, the hypothesis $\mathcal{H}_1$ does not have its own statistical model, but it represents the complement of the null hypothesis. We in fact suppose that Eve can forge messages through which she can modify the channel estimation obtained by Bob for the Eve-Bob channel into any possible vector $\mathbf{g}$.

When Bob does not know the statistical distribution of the attacker's channel, the presence of multiple attackers always boils down to the case of a single attacker. The authenticator in fact must refuse all new messages that differ from Alice's channel estimate known to him, regardless of which attacker has forged the message. The general case with more than one legitimate transmitter is not considered here, but it can be derived from the model presented. We omit it for brevity.

Two different statistical criteria for Bob to decide between the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are considered next. The performance achievable through these tests is then evaluated by measuring their probability of FA and of MD.

A. Logarithm of likelihood ratio test

Let us first consider the generalized likelihood ratio test (GLRT) [28], as done in [9], where flat fading channels were considered and $\mathbf{g}$ is replaced by its maximum likelihood (ML) estimate. We take into account the more general case in which channel variations occur during the authentication.

The logarithm of the likelihood ratio (LLR) of a channel estimate $\hat{\mathbf{h}}$ over its N components can be written as [9]:

$$\Psi \propto 2\sum_{n=1}^{N} \frac{1}{\sigma_n^2} \left| \hat{h}_n - \bar{h}_n^{(\text{AB})} \right|^2, \tag{10}$$

where σ_n^2 represents the per-dimension variance, evaluated as $\sigma_n^2 = \sigma_{\text{I}}^2 + \sigma_{\text{II}}^2 + (1 - \bar{\alpha}_n^{(\text{I})2}) + (1 - \alpha_n^{(\text{II})2})$.

By substituting (8) in (10), we obtain that under the hypothesis $\mathcal{H}_0$, Ψ is a non-central chi-square random variable as in [9], with non centrality parameter

$$\mu = \sum_{n=1}^{N} \frac{2}{\sigma_n^2} \left| (\alpha_n^{(\mathrm{II})} - \bar{\alpha}_n^{(\mathrm{I})}) h_n^{(\mathrm{AB})} \right|^2. \tag{11}$$

We note that μ is strictly dependent on $\bar{\boldsymbol{\alpha}}^{(\mathrm{I})}$, and becomes zero in the limit case of $\bar{\alpha}_n^{(\mathrm{I})} = \alpha_n^{(\mathrm{II})} = 1$ on each channel (absence of fading during both phases), which boils down to the case considered in [9].

The GLRT consists in comparing the LLR with a threshold $\theta > 0$, i.e.

$$\begin{cases} \Psi \leq \theta: & \text{decide for } \mathcal{H}_0, \\ \Psi > \theta: & \text{decide for } \mathcal{H}_1. \end{cases} \tag{12}$$

We can evaluate the probability of FA P_{FA}, i.e., the probability that Bob refuses a message coming from Alice, as

$$P_{\mathrm{FA}} = P[\Psi > \theta | \mathcal{H}_0] = 1 - F_{\chi^2,\mu}(\theta), \tag{13}$$

where $F_{\chi^2,\mu}(\cdot)$ denotes the cumulative density function (c.d.f.) of a chi-square random variable with $2N$ degrees of freedom and noncentrality parameter μ.

By substituting the hypothesis $\mathcal{H}_1$ in (10), we observe that Ψ is again a non-central chi-square random variable, but the noncentrality parameter in this case is given by

$$\beta = \sum_{n=1}^{N} \frac{2}{\sigma_n^2} \left| g_n - \bar{\alpha}_n^{(\mathrm{I})} h_n^{(\mathrm{AB})} \right|^2. \tag{14}$$

We can then calculate the probability of MD as

$$P_{\mathrm{MD}} = P[\Psi \leq \theta | \mathcal{H}_1] = F_{\chi^2,\beta}(\theta). \tag{15}$$

By imposing a target P_{FA}, the threshold is set as

$$\theta = F_{\chi^2,\mu}^{-1}(1 - P_{\mathrm{FA}}). \tag{16}$$

In the case we consider, in which fading can also affect the training phase, the authentication performance is always worse than in the case without fading during the training phase. In fact, according to the properties of the non-central chi-squared distribution, the presence of fading (measured by means of the parameter $\alpha^{(\mathrm{I})}$) leads to an increase of both the non-centrality parameters μ and β with respect to their corresponding values in absence of time-varying fading. As a consequence, given the same P_{FA}, also the value of the threshold θ increases, with an inevitable worsening of the probability of MD. This is also proved by the numerical results shown in Sec. VIB.

B. Ideal performance

In order to assess the authentication performance in an ideal setting, let us suppose that in Phase 1 Bob is able to collect messages that surely come from Eve, and thus are in the form

$$\hat{\mathbf{h}}^{(\mathrm{E})} = \mathbf{g} + \mathbf{w}^{(\mathrm{I})}. \tag{17}$$

In this case we do not need to substitute Eve's forged vector $\mathbf{g}$ with its ML estimate, since Bob knows it exactly except for the presence of AWGN noise. In this case, we can no longer resort to hypothesis testing, since we now have two different hypotheses about the source of the message. However, with a slight abuse of notation, we still use the labels $\mathcal{H}_0$ and $\mathcal{H}_1$ for identifying the two events corresponding to Alice's and Eve's transmissions, respectively. Thus the LLR on the new estimated channel $\hat{\mathbf{h}}$ becomes

$$\bar{\Psi} = \ln \frac{f_{\hat{\mathbf{h}}|\mathcal{H}_1}(\hat{\mathbf{h}})}{f_{\hat{\mathbf{h}}|\mathcal{H}_0}(\hat{\mathbf{h}})}, \tag{18}$$

where $f_{(\hat{\mathbf{h}}|\cdot)}(\hat{\mathbf{h}})$ represents the probability density function (p.d.f.) of $\hat{\mathbf{h}}$ under hypothesis $\mathcal{H}_1$ at the numerator and under hypothesis $\mathcal{H}_0$ at the denominator.

Considering that under hypothesis $\mathcal{H}_0$ $\hat{\mathbf{h}}$ is Gaussian distributed around $\hat{\mathbf{h}}^{(\mathrm{AB})}$ with per-dimension variance σ^2, while under hypothesis $\mathcal{H}_1$ is Gaussian distributed around $\hat{\mathbf{h}}^{(\mathrm{E})}$ with per-dimension variance σ_{E}^2 (the value of the variance depends on how Eve forges $\mathbf{g}$, which will be described in Section IV), eq. (18) on the n-th sub-carrier can be written as

$$\bar{\Psi} = N \ln \frac{\sigma}{\sigma_{\mathrm{E}}} + \frac{1}{2\sigma^2} \sum_{n=1}^{N} |\hat{h}_n - \bar{h}_n^{(\mathrm{AB})}|^2 - \frac{1}{2\sigma_{\mathrm{E}}^2} \sum_{n=1}^{N} |\hat{h}_n - \hat{h}_n^{(\mathrm{E})}|^2. \tag{19}$$

We now exploit the following decision criterion:

$$\begin{cases} \bar{\Psi} \leq \bar{\theta}: & \text{accept}, \\ \bar{\Psi} > \bar{\theta}: & \text{refuse}, \end{cases} \tag{20}$$

where $\bar{\theta}$ represents the minimum value of the threshold needed to achieve a given probability of false alarm.

Differently from (13) and (15), in this case no closed form expression is available for P_{FA} and P_{MD} and they are estimated through Monte Carlo simulations. The achievable probabilities of FA and MD represent a lower bound on the test performance obtainable when applying the LLR test and will be used as a benchmark in some examples presented in Section VI.

C. Combined test

As it results from our numerical simulations, reported in Section VI, the LLR test alone however is not sufficient to guarantee a correct authentication (and a small probability of MD) for values of $\boldsymbol{\alpha}^{(\mathrm{II})}$ that are not next to 1. In order to address this issue and improve performance, we also consider a slightly modified decision strategy for Bob, based on a double verification. For this purpose, let us consider that Bob still exploits the LLR test, but followed by a modulus comparison, to decide whether a message is coming from Alice or Eve. The additional test based on the modulus is performed by comparing the modulus of the reference estimate $\bar{\mathbf{h}}^{(\mathrm{AB})}$ and the current estimate $\hat{\mathbf{h}}$. Thus we define

$$\Gamma = \sum_{n=1}^{N} \left(\left| \bar{h}_n^{(\mathrm{AB})} \right| - \left| \hat{h}_n \right| \right). \tag{21}$$

Using such a simple modulus comparison alone results in a poor performance. However, we consider that Bob uses both the criterion based on the LLR and that based on the modulus: only if both these conditions are met, then Bob

accepts the message as authentic. The verification condition can be therefore written as

$$\begin{cases} \Phi \leq \theta, -\epsilon \leq \Gamma \leq \epsilon : & \text{decide for } \mathcal{H}_0, \\ \text{else} \ : & \text{decide for } \mathcal{H}_1, \end{cases} \tag{22}$$

where ϵ is a sufficiently small threshold. In the ideal case of absence of noise and fading during the training phase, ϵ should be zero. Since we are considering a realistic scenario affected by both disturbances, we must allow ϵ to be greater than zero in order to allow Bob to accept messages coming from Alice.

The probability of false alarm can be therefore defined as the probability that at least one of the two conditions is not verified when the sender is Alice (hypothesis $\mathcal{H}_0$), i.e.

$$P_{\text{FA}} = 1 - P\left[\Psi \leq \theta, -\epsilon \leq \Gamma \leq \epsilon | \mathcal{H}_0\right], \tag{23}$$

while the probability of missed detection can be written as

$$P_{\text{MD}} = P\left[\Psi \leq \theta, -\epsilon \leq \Gamma \leq \epsilon | \mathcal{H}_1\right]. \tag{24}$$

Being Γ computed as the difference of the modulus of two complex normal random variables, which follow an Hoyt distribution [29], its c.d.f. can be evaluated according to [30, eq. (8)]. However, a closed form expression for the joint probability distribution of Ψ and Γ is not known, thus for its estimation we resort to Monte Carlo simulations.

Thresholds optimization: In order to find the optimal values θ^* and ϵ^* of the thresholds θ and ϵ, we look for their joint values which minimize the probability of MD, ieeei.e.

$$(\theta^*, \epsilon^*) = \arg\min_{\theta,\epsilon} P\left[\Psi \leq \theta, -\epsilon \leq \Gamma \leq \epsilon | \mathcal{H}_1\right], \tag{25}$$

under the constraint of a fixed P_{FA}.

We exploit a two-steps optimization procedure, which computes the couples (θ, ϵ) that satisfy the constraint imposed on P_{FA} and then select the one that gives the minimum P_{MD}.

IV. Attacker model

We assume that Eve aims at performing a tailored attack based on Bob's decision strategy. According to well-known Kerckhoffs's principle, we build our system model not relying on the principle of "security through obscurity", meaning that the attacker has all the information available. In this sense, the assumption that Eve knows exactly which test is used by Bob is made to model the worst case scenario and consequently adapt the system parameters to prevent possible authentication errors (for example by adjusting the transmission power and the signal-to-noise ratio (SNR) on the different channels).

We suppose that she has partial CSI, that is, she knows the statistics of all transmission channels, but not the exact channel realizations. We assume that Eve cannot know the channel estimates made by Bob exactly (this would be possible only if they were in the same position), but she can obtain a slightly modified version of them by eavesdropping the communication channels used by Alice and Bob. We also suppose that Eve can observe transmissions from Alice to Bob and vice versa, thus estimating $\mathbf{h}^{(\text{AE})}$ and $\mathbf{h}^{(\text{EB})}$. We denote the m-th channel estimates obtained by Eve as

$$\hat{\mathbf{h}}^{(\text{AE})}(m) = \rho_{\text{AE}}\mathbf{h}^{(\text{AB})} + \sqrt{1-\rho_{\text{AE}}^2}\mathbf{r}(m) + \mathbf{w}^{(\text{AE})}(m), \tag{26}$$

$$\hat{\mathbf{h}}^{(\text{EB})}(m) = \rho_{\text{EB}}\mathbf{h}^{(\text{AB})} + \sqrt{1-\rho_{\text{EB}}^2}\mathbf{r}(m) + \mathbf{w}^{(\text{EB})}(m), \tag{27}$$

where $\mathbf{w}^{(\text{AE})} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_{\text{AE}}^2\mathbf{I}_N)$ and $\mathbf{w}^{(\text{EB})} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_{\text{EB}}^2\mathbf{I}_N)$ represent the noise vectors and $\mathbf{r}$ is a complex normal random vector with unitary variance. The coefficient $\rho_{\text{XY}} \in [0;1]$ denotes the spatial correlation between two channels linking a generic node Z to two distinct nodes X and Y.

Opposed to the correlation of several realizations of the same channel in time, we denote this correlation as spatial correlation of two different channels at some fixed time. The correlation of the estimates $\hat{\mathbf{h}}^{(\text{AE})}(m)$ and $\hat{\mathbf{h}}^{(\text{EB})}(m)$ obtained through (26) and (27), respectively, with the main channel coefficient $\mathbf{h}^{(\text{AB})}$ strictly depends on Eve's position with respect to Bob's one, and is represented by means of the parameters ρ_{AE} and ρ_{EB}. If $\rho_{\text{AE}} = 1$, for example, Eve and Bob are in the same position and the channels estimated by them with respect to messages transmitted by Alice are identical. On the contrary, when ρ_{AE} tends to zero the channels Alice-Bob and Alice-Eve are completely uncorrelated.

Eve's attack is supposed to be based on the average of the estimates collected in Phase 1, since Bob relies on this for the subsequent authentication phase. As already done by Bob, if she can retrieve $M > 1$ observations of $\mathbf{h}^{(\text{AE})}$ and $\mathbf{h}^{(\text{EB})}$, she can try to refine her attack by averaging over them in order to reduce the noise level. This is opposed to a differential authentication approach [31], in which Bob progressively updates its reference estimate. The latter, however, is suitable in the case of correlated fading over time, which is not the case we consider. As a consequence, the time coefficient vector $\boldsymbol{\alpha}^{(\text{II})}(t)$ in Phase 2 does not affect her forged vector $\mathbf{g}$.

A. Attack to the LLR test

When Bob uses the LLR test, Eve's best attack strategy is represented by the ML estimate of $\hat{\mathbf{h}}^{(\text{AB})}$ based on her observations $\hat{\mathbf{h}}^{(\text{AE})}$ and $\hat{\mathbf{h}}^{(\text{EB})}$. According to [9, eq. (45)], the components of the forged vector $\mathbf{g}$ can be written as

$$g_n = \hat{h}_n^{(\text{EB})} C_n + \hat{h}_n^{(\text{AE})} D_n, \tag{28}$$

with C_n and D_n defined as

$$C_n = \frac{\rho_{\text{EB}}\omega_n^{(\text{EB})} - \rho_{\text{AB}}\rho_{\text{AE}}}{\omega_n^{(\text{AE})}\omega_n^{(\text{EB})} - \rho_{\text{AB}}^2}, \tag{29a}$$

$$D_n = \frac{\rho_{\text{AE}}\omega_n^{(\text{AE})} - \rho_{\text{AB}}\rho_{\text{EB}}}{\omega_n^{(\text{AE})}\omega_n^{(\text{EB})} - \rho_{\text{AB}}^2}, \tag{29b}$$

where $\omega_n^{(\text{AE})} = 1 + \frac{\sigma_{\text{AE}}^2}{\lambda_n}$, $\omega_n^{(\text{EB})} = 1 + \frac{\sigma_{\text{EB}}^2}{\lambda_n}$. According to the previous definition, ρ_{AB} represents the spatial correlation between the two channels observed by Eve, i.e. the Alice-Eve channel and the Eve-Bob one. λ_n is the power delay, and we suppose that this value is always known to the attacker, thus considering a worst case scenario, in which Eve is in an advantageous condition to perform her attack. The case of Eve not knowing exactly the power delay profile is more realistic, but puts Eve in a less favorable situation and is out of the scope of this paper.

B. Attack to the combined test

In order to find the optimal attack strategy for Eve when Bob uses the combined test described in Section III-C, denoted as *modulus attack* for brevity, we consider the worst case scenario, where Eve is able to perfectly estimate $\mathbf{h}^{(\mathrm{AE})}$ and $\mathbf{h}^{(\mathrm{EB})}$, i.e. $\sigma^2_{\mathrm{AE}} = \sigma^2_{\mathrm{EB}} = 0$. For the sake of simplicity we also suppose that no correlation exists between the two channels observed by her or, in other words, $\rho_{\mathrm{AB}} \to 0$.

Under these hypotheses (28) boils down to

$$g_n = \rho_{\mathrm{AE}}\hat{h}_{\mathrm{AE}} + \rho_{\mathrm{EB}}\hat{h}_{\mathrm{EB}}. \tag{30}$$

On the other hand, in order to make $|\mathbf{g}|$ more similar to $\left|\hat{\mathbf{h}}^{(\mathrm{AB})}\right|$, in the modulus attack Eve can choose

$$g_n = \frac{\hat{h}_n^{(\mathrm{AE})}}{\rho_{\mathrm{AE}}} + \frac{\hat{h}_n^{(\mathrm{EB})}}{\rho_{\mathrm{EB}}}, \text{ for } \rho_{\mathrm{AE}}, \rho_{\mathrm{EB}} \neq 0. \tag{31}$$

When we consider a decision strategy based on both LLR and modulus comparison, however, attacks based on (30) and (31) are no longer optimal. In this case, the best attack strategy in fact consists in forging

$$g_n = \rho_{\mathrm{AE}}^{x}\hat{h}_n^{(\mathrm{AE})} + \rho_{\mathrm{EB}}^{y}\hat{h}_n^{(\mathrm{EB})}, \tag{32}$$

where $x, y \in [-1, 1]$. This requires finding a trade-off between the modulus attack and the LLR attack. In fact, $x = y = -1$ corresponds to the best attack to the modulus comparison method, while $x = y = 1$ corresponds to the best attack strategy to the LLR. Finding the optimal values of the couple (x, y) corresponds to solve the problem

$$(x, y) = \arg \max_{x,y\in[-1,1]} P\left[\Psi(x) \leq \theta, \Gamma(x)^2 \leq \epsilon^2 | \mathbf{g} = (32)\right], \tag{33}$$

i.e. to find the values of (x, y) that give the highest P_{MD}.

In Table I we report the optimal values of (x, y) obtained by solving (33) through numerical methods, with $P_{\mathrm{FA}} = 10^{-4}$ and different values of sub-carriers N and the time-varying fading $\boldsymbol{\alpha}^{(\mathrm{II})}$, considering the absence of time-varying fading in Phase 1 ($\boldsymbol{\alpha}^{(\mathrm{I})} = 1$). We do not report values corresponding to $\rho_{\mathrm{AE})} \geq 0.7$, which are always equal to 1. We observe that for high values of the spatial correlation the attack to the LLR test results to be most convenient strategy to adopt for Eve even when Bob chooses the combined test. The case with $\rho_{\mathrm{AE}} = 1$ is of special interest: Eve is exactly is Alice's same position, so each kind of attack results successful regardless of the choice of (x, y). Since the attacker does not know exactly the values of $\boldsymbol{\alpha}^{(\mathrm{II})}$, her most conservative choice is to suppose that all the entries of $\boldsymbol{\alpha}^{(\mathrm{II})}$ are equal to 1. In fact, Eve is in the worst condition in absence of fading, since Bob lets fewer messages to be accepted as authentic.

C. Mismatched Attacker

Let us consider an attacker who does not exactly know which is the decision criterion used by Bob. In such a case, Eve has no option but to try to design $\mathbf{g}$ such as it gives the best probability of MD in each case. In particular, Eve's problem here is to decide a general strategy which leads to the highest possible probability of missed detection whatever is the decision method adopted by Bob.

TABLE I
OPTIMAL VALUES OF THE EXPONENTS (x, y) IN EQ. (32) FOR DIFFERENT VALUES OF $\boldsymbol{\alpha}^{(\mathrm{II})}$ AND N, WITH $\boldsymbol{\alpha}^{(\mathrm{I})} = 1$ AND $\rho_{\mathrm{AE}} = \rho_{\mathrm{EB}} = \rho$.

$\boldsymbol{\alpha}^{(\mathrm{II})}$	N	ρ					
		0.1	0.2	0.3	0.4	0.5	0.6
1	1	$(0.7, 0.7)$	$(0.8, 0.8)$	$(0.9, 1)$	$(1, 1)$	$(1, 1)$	$(1, 1)$
	3	$(0.7, 0.7)$	$(0.8, 0.9)$	$(1, 1)$	$(1, 1)$	$(1, 1)$	$(1, 1)$
	6	$(0.8, 0.6)$	$(0.9, 0.9)$	$(1, 1)$	$(1, 1)$	$(1, 1)$	$(1, 1)$
0.8	1	$(0.4, 0.5)$	$(0.5, 0.6)$	$(0.6, 0.6)$	$(0.7, 0.6)$	$(0.8, 0.8)$	$(0.9, 1)$
	3	$(0.4, 0.5)$	$(0.5, 0.6)$	$(0.6, 0.6)$	$(0.8, 0.6)$	$(0.8, 0.8)$	$(1, 1)$
	6	$(0.5, 0.3)$	$(0.6, 0.4)$	$(0.5, 0.6)$	$(0.7, 0.6)$	$(0.7, 0.8)$	$(0.9, 0.9)$

In Figs. 1(a) and 1(b) we compare the average probability of MD obtained when Eve's attack strategy follows (30), (31) and (32) with the values of (x, y) found in Table I for $\rho_{\mathrm{AE}} = \rho_{\mathrm{EB}} = 0.1$ and $\boldsymbol{\alpha}^{(\mathrm{II})} = [0.8, 1]$ on each sub-carrier. It is possible to note that using the wrong attack strategy is penalizing especially when Bob exploits a combined test and when $\boldsymbol{\alpha}^{(\mathrm{II})} < 1$, and this is particularly evident for a large value of N. In case of stationary channels, i.e. $\boldsymbol{\alpha}^{(\mathrm{II})} = 1$, performances are almost equivalent for both decision strategies.

V. METHODS BASED ON MACHINE LEARNING

Let us now consider the application of machine learning techniques to the authentication protocol presented in Sec. II. These methods have the advantage of not requiring any knowledge about the receiver's CSI, not even the variance σ^2 needed by the statistical tests illustrated in Sec. III. They are hence able to perform authentication in a blind way.

We consider two different scenarios. In the ideal one, the authenticator has some information about the attacker, thus Phase I corresponds to the training phase of a binary classifier. In the second one, more realistic, Bob is able to collect only samples belonging to the legitimate transmitter during Phase I and we must consider one class classification (OCC), where only one class (referred to as positive class or target class) is present during the training phase, while the others (negative classes or non-target classes) are not known [32].

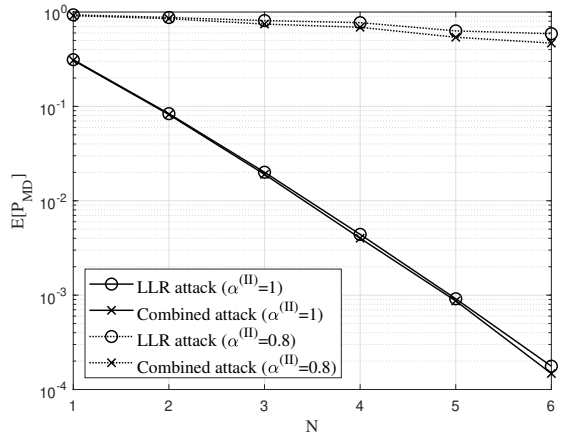
OCC methods are mainly based on two parameters [33]: the distance $d(x)$ between the sample to classify and the target class, and the threshold θ_d on the value of the distance. Formally, a new instance is recognized as belonging to the positive class if its distance from the target class is below the threshold. In formulas

$$f_{oc}(x) = I\left(d(x) < \theta_d\right) \tag{34}$$

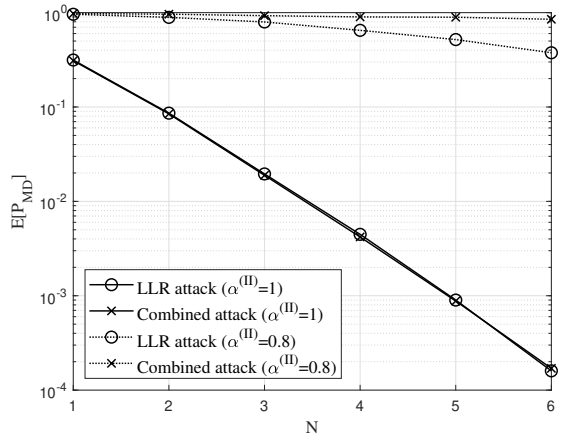
where $I(\cdot)$ represents an indicator function[1] and $f_{oc}(x)$ is the decision function, i.e. a binary function expressing acceptance of the object x into the target class. There are several metrics which can be used to evaluate $d(x)$, among which the most common choice is the Euclidean distance.

OCC approaches differ in the evaluation of the distance and in the optimization of the threshold, on the basis of

[1]An indicator function is a function defined on a set *X* that indicates membership of an element in a subset *A* of *X*. Its value is 1 for all elements of *A* and 0 for all elements of *X* not in *A*.



(a)



(b)

Fig. 1. Performance of the matched versus mismatched attacker in terms of average P_{MD}, when Bob applies the a) LLR test and b) combined test, with $\boldsymbol{\alpha}^{(\text{I})} = 1$, $\rho_{\text{AE}} = \rho_{\text{EB}} = 0.1$.

the available training set. In the following, we consider two different algorithms, i.e. the NN and the SVM techniques.

A. Nearest Neighbors

NN algorithms classify a sample x by assigning it the most frequent label among its k nearest neighbors. What makes NN algorithms particularly suitable for an authentication scheme to be implemented in resource-constrained devices is their low complexity, which grows with the training set dimension. They operate classification only on instances of the training set and not on statistical assumptions. Since in authentication schemes it is reasonable to assume that phase I is short and Bob receives a limited number of setup packets, it is also reasonable to assume to work with a small training set, thus making the choice of classification based on NN algorithms a good trade-off between complexity and performance.

One-class Nearest Neighbor (OCNN) techniques are derived from traditional binary NN algorithms. In detail, a OCNN algorithm works as follows [34]: it finds the *j* nearest neighbors $\{y_1, \cdots, y_j\}$ of the test sample x in the target class, and the k nearest neighbors $\{z_{i1}, \cdots, z_{ik}\}$ of the first j neighbors; it evaluates the average distance $\bar{D}_{xy}$ over $\left\{D_{xy_1}, \cdots, D_{xy_j}\right\}$ and the average distance $\bar{D}_{yz}$ over $\{D_{y_1 z_{11}}, \cdots, D_{y_1 z_{1k}}, \cdots, D_{y_1 z_{k1}}, \cdots, D_{y_k z_{kk}}\}$; x is then considered as a member of the target class if

$$\frac{\bar{D}_{xy}}{\bar{D}_{yz}} < \theta_d. \tag{35}$$

OCNN methods can be grouped into four main categories (11NN, 1KNN, J1NN, JKNN), depending on which of the parameters j and k is fixed to 1. They differ in the number of neighbors used to compute the decision threshold [34].

When the training set is not affected by time-varying fading while the test set is, the probability of FA increases with the number of features. As an example, let us consider a 11NN algorithm, which accepts a new packet if (35) is satisfied. y and z both belong to the training set, thus D_{yz} is not influenced by the time-varying fading, differently from D_{xy}. Both distances increase with the number of features $2N$, whether they are computed using a Euclidean distance or the LLR considered in Section V-D. In the stationary case, the increase in D_{xy} following from an increased number of features is balanced by an increase in D_{yz}, and their ratio does not increase with the number of features. The presence of time-varying fading instead makes D_{xy} increase more than in the stationary case, and in this case such an effect is only partly counterbalanced by an increase in D_{yz}. The threshold θ_d remains the same in both cases, being computed only on the basis of the training phase, which does not include time-varying fading. Therefore, by increasing the number of features, it is more likely that the ratio D_{xy}/D_{yz} overcomes θ_d, causing the rejection of an authentic message and the occurrence of a FA. These considerations can be easily extended to all the OCNN methods examined in this paper.

B. Support Vector Machines

Despite their low complexity, NN algorithms require to determine some optimal parameters, such as j, k and θ_d, which can become computationally expensive. For this reason, we also consider a second kind of classifiers, known as SVM.

In general, SVMs can create a non-linear decision boundary to separate different classes by projecting the data through a non-linear function to a space with a higher dimension, lifting them from their original space to a feature space, which can be of unlimited dimension. Their one-class version, also known as single-class classification or *novelty detection*, was introduced in [35]. The main concept behind the OC-SVM algorithm consists in obtaining a spherical boundary, in feature space, around the data. The volume of this hyper sphere is minimized, to minimize the effect of incorporating outliers in the solution [36]. In particular, the goal of OC-SVM is to estimate a function $f_{oc}(x)$ that encloses the most of training data into a hyper sphere $R_x = \left\{x \in R^N | f_{oc}(x) > 0\right\}$, where N is the size of feature vector. The decision function $f_{oc}(x)$ is written as

$$f_{oc}(x) = sgn\left\{\sum_{i=1}^{m} \lambda_i K(x, x_i) - \xi\right\}, \tag{36}$$

where m represents the number of training samples, ξ is the distance of the hyper sphere from the origin and $K(\cdot,\cdot)$ defines the OC-SVM kernel that allows projecting data from the original space to the feature space. λ_i are the Lagrange multipliers computed by optimizing the following equations

$$\min_{\lambda}\left\{\frac{1}{2}\sum_{i,j}\lambda_i\lambda_j K(x_i,x_j)\right\}, \tag{37}$$

subject to $0 \leq \lambda_i \leq \frac{1}{\nu m}$ and $\sum_{i=1}^{m}\lambda_i = 1$, where ν is the percentage of data considered as outliers. Modifying the parameter ν allows to optimize the value of g_{mean} or to minimize one of the two probabilities of FA and MD.

A pattern x is accepted if $f_{oc}(x) > 0$ and rejected otherwise. Different functions can be used, such as linear, polynomial or Gaussian kernels. Usually, the Gaussian is the most used kernel, which allows determining the radius of the hyper sphere according the parameter $1/2\sigma^2_{\text{SVM}}$. It is defined as

$$K(x,x_i) = \exp\left(-\frac{||x-x_i||^2}{2\sigma^2_{\text{SVM}}}\right), \tag{38}$$

where the numerator represents the squared Euclidean distance between two general feature vectors, while σ_{SVM} is a free parameter.

C. Clustering

All previous classification mechanisms require that Bob exactly knows the source of one or more packets during the training phase. This assumption requires the use of higher layer cryptographic protocols or manual solutions, which may result unpractical in some cases. To address this issue, we also consider the use of clustering algorithms [37] for Bob to establish the origin of packets during the training phase and to assign them a possibly correct label. In particular, we focus on the k-means approach [38], which guarantees good performance and a fast convergence; moreover, with a small input set and when the number of clusters k is known and small, it also has a reduced complexity.

Given a set of m instances $(x_1, x_2, \cdots, x_m)$, the goal of k-means clustering is to partition them into $k \leq m$ sets (or *clusters*) $\mathbf{S} = \{S_1, S_2, \cdots, S_k\}$ such as to minimize the within-cluster sum of squares (WCSS). More formally, the objective is to find

$$\arg\min_{S}\sum_{i=1}^{k}\sum_{x\in\mathcal{S}_i}\|x-\mu_i\|^2, \tag{39}$$

where μ_i is the mean of points in S_i. Classes labels in the training phase will be determined by results got from the algorithm and not assigned a priori by the authenticator.

D. Hybrid approach

Let us consider a hybrid approach that exploits the statistical metrics considered in Section III-A along with OCNN classifiers described in Section V-A. This is done by using in (35) the LLR defined in (10) instead of the Euclidean distance. With respect to the Euclidean distance, computing the LLR requires knowledge of the statistical distribution through the value of σ^2. When the LLR is used along with a JKNN classifier, we select the j samples from the training set that minimize the LLR evaluated on the new sample to classify, and the k samples that minimize the LLR computed on the j samples. Analogously, the LLR can be used with all the considered NN methods.

VI. Results and discussion

In this section, we assess and compare the performance of statistical and machine learning-based decision methods under different system conditions and assumptions. For the sake of simplicity, and without loss of generality, we consider examples in which time-varying fading affects all channels in the same way during both authentication phases, i.e. $\alpha_n = \alpha$ for $n = 1, \cdots, N$. We also suppose that Eve can estimate the channel between Alice and herself, but not the channel between her and Bob. Bob in fact receives messages from Alice but is not expected to transmit enough messages to allow the attacker to extract useful information on $\mathbf{h}^{(\text{EB})}$, thus imposing $\rho_{\text{AB}} \to 0$, $\rho_{\text{EB}} \to 0$, $\rho_{\text{AE}} > 0$. Moreover, we assume to give Eve the maximum advantage, allowing her to perfectly estimate $\mathbf{h}^{(\text{AE})}$, i.e. considering $\sigma^2_{\text{AE}} = 0$. The average SNR on channel estimates during both phases is $\text{SNR}^{(\text{I})} = 1/\sigma^2_{\text{I}}$ and $\text{SNR}^{(\text{II})} = 1/\sigma^2_{\text{II}}$.

We compare the performance of statistical and machine learning-based decision methods by giving them the same inputs. To this end, the features we use for machine learning are represented by the real and the imaginary parts of the channel coefficient measured on each sub-carrier for both NN and SVM algorithms. Therefore the number of features corresponds to $2N$.

A. Parameter setting and OCC performance evaluation

We first examine the security performance achievable by OCC techniques. As already said in Sec. V, the definition of the parameters of NN algorithms has an important impact on the precision of the results. In order to find a trade-off between false alarm and missed detection, the parameters j, k and θ_d have been optimized in such a way as to maximize g_{mean}, by using a g-fold cross validation. As regards binary k-NN, common values for k are usually in the range $[3, \sqrt{M}]$, given M as the number of samples in the training set [39], and k must be an odd number in order to avoid parity issues. As an example, the results obtained for $\alpha^{(\text{I})} = \alpha^{(\text{II})} = 1$ and $N = 3$ are reported in Tab. II, choosing g equal to 5 and considering two training sets of 100 and 1000 samples (with 50% of samples belonging to the positive class and 50% to the negative class for the k-NN).

OCNN algorithms have been implemented in Matlab 2019a. The presented case studies were run on a machine equipped with an Intel Core i7-8565U 3.9 Ghz Quad-core processor and 16 GB of RAM. The number of features apparently has an almost irrelevant impact on the computational times, as shown in Fig. 2. In this kind of classifiers, in fact, the number of features, i.e. the length of the input vectors, only affects the initial computation of the distance (in our case the Euclidean

distance) between samples. For the relatively small numbers of features considered in Fig. 2, the variation in the computation time of the Euclidean distance for vectors of increasing lengths is almost negligible and has a small effect on the training time. We note that, as predictable, JKNN requires a larger time than the other methods, especially compared to 11NN, which needs only a threshold optimization.

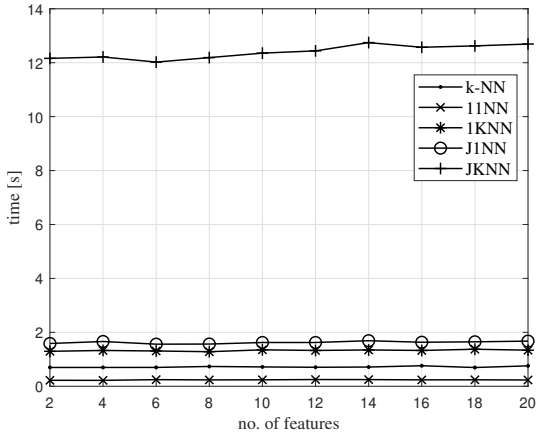


Fig. 2. Training times required for different NN algorithms with $\alpha^{(\mathrm{I})} = 0.8$, $\alpha^{(\mathrm{II})} = 0.9$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB and $M = 100$.

TABLE II
OPTIMIZED PARAMETERS FOR NN METHODS, WITH $\alpha^{(\mathrm{I})} = \alpha^{(\mathrm{II})} = 1$ AND $N = 3$, CONSIDERING TWO TRAINING SETS OF 100 AND 1000 SAMPLES (INTO BRACKETS).

		KNN	11NN	1KNN	J1NN	JKNN
$\rho_{\mathrm{AE}} = 0.1$	k	9 (31)	-	3 (3)	-	2 (2)
	j	-	-	-	2 (2.5)	2 (2)
	θ_d	-	3 (3.5)	2.5 (3)	3 (3.5)	2.5 (2.5)
$\rho_{\mathrm{AE}} = 0.8$	k	3 (7)	-	5 (27)	-	6 (22)
	j	-	-	-	7 (25)	6 (22)
	θ_d	-	2 (2)	1.5 (1)	2 (2.5)	2 (1.5)

As for SVM algorithms, we exploit a modified version of the tool in [40]. We use a Gaussian kernel, which proved to be more effective than linear and polynomial kernels in our scenario, as it is evident from the results in Tab. III.

Results achieved by applying different OCC algorithms to the same data set are then compared. In the following examples, we consider $\mathrm{SNR}^{(\mathrm{I})} = 15$dB and $\mathrm{SNR}^{(\mathrm{II})} = 20$dB, fading coefficients $\alpha^{(\mathrm{I})} = 1$ and $\alpha^{(\mathrm{II})} = 0.9$, a training set of $M = 100$ samples and a classification set composed by 5 subsets of $4 \cdot 10^5$ elements. Each subset contains 50% of positive samples and 50% of negative samples. The values of P_{MD} and P_{FA} are averaged over 100 different datasets randomly generated.

In Fig. 3 the performance obtained for different values of the spatial correlation coefficient ρ_{AE} is shown, with the meaning of Eve being at decreasing distances from Bob, while in Fig. 4a results are obtained for different numbers of sub-carriers. We can note that, in general, there is no OCNN algorithm that results to be the best one for any value of ρ_{AE}. However, for a given similar performance, 11NN requires much smaller

TABLE III
RESULTS OBTAINED BY SVM WITH DIFFERENT KERNELS, WITH $M = 100$, $\rho_{\mathrm{AE}} = 0.8$, $\alpha^{(\mathrm{I})} = \alpha^{(\mathrm{II})} = 1$ AND A DATASET DIMENSION EQUAL TO 1000.

	N	1	2	3	4	5	6
Gaussian	P_{FA}	0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	P_{MD}	0.08	0.024	0.05	0.01	0.001	< 0.001
Poly	P_{FA}	0.019	0.007	0.022	0.002	0.03	0.005
	P_{MD}	0.119	0.064	0.058	0.04	0.02	0.017
Linear	P_{FA}	0.019	0.007	0.02	0.002	0.029	0.005
	P_{MD}	0.118	0.064	0.058	0.04	0.02	0.017

training and classification times, especially in comparison with JKNN. Under the considered assumptions, SVM obtains the lowest probability of MD, but at the expenses of a higher FA. The SVM in fact obtains a P_{FA} equal to 0.944, while for J1NN it is equal to 0.716 (similar values are obtained by the other NN algorithms and not reported here for brevity). For both the examined methods we consider the parameters, i.e. the values of j, k and θ_d for the OCNN classifiers and ν as regards the SVM, that achieve the best value of g_{mean} during the training phase. By looking at the curves it is possible to observe that until $\rho_{\mathrm{AE}} \leq 0.4$ all the algorithms exhibit a MD probability lower than 10^{-6}, meaning that no MD event has been observed over the entire classification set, of dimension 10^6. Moreover, as expected, Fig. 4(a) testifies the fact that an increasing number of features is beneficial from a security point of view, even if the probability of FA slightly increases with the number of features, as discussed in Section V-A. According to the considerations reported at the end of Section V-A, the presence of fading, by means of the parameter $\alpha^{(\mathrm{II})}$, negatively influences the probability of false alarm achieved by OCNN classifiers with an increasing number of subcarriers, and therefore of features. It is reasonable to suppose that a similar behavior also affects the SVM, leading to the reduction of performance shown in Fig. 4(a). The worsening of the false alarm probability, together with an improvement of the missed detection probability, inevitably deteriorates the performance in terms of g_{mean}, which gives a measure of the balance between the values of P_{FA} and P_{MD} (see Eq. (7)). Fig. 4(b) reports the optimal values of g_{mean} for increasing values of N, showing that NN algorithms are able to achieve a better g_{mean} than SVMs. For the sake of comparison, we also show the performance in terms of accuracy, which exhibits less sensitivity to the number of features with respect to g_{mean}.

B. Impact of the training phase

The training phase, or Phase I, has clearly a relevant impact on the subsequent classification phase performance. Let us assess how different training parameters and conditions influence the system performance.

Firstly, reliability of the training set must be taken into account. In the ideal case, instances labels forming the training set are exactly known. In real cases, instead, they may be unknown, and must be hence inferred through clustering techniques. A comparison of these two cases in terms of probability of MD is illustrated in Fig. 5, where a training

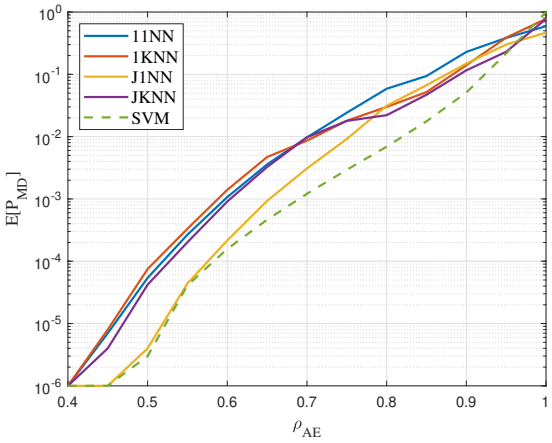


Fig. 3. Probability of MD as a function of the spatial correlation ρ_{AE}, comparing NN and SVM algorithms, with $\alpha^{(\text{I})} = 1$, $\alpha^{(\text{II})} = 0.9$, $N = 3$, $\text{SNR}^{(\text{I})} = 15$dB, $\text{SNR}^{(\text{II})} = 20$dB and $M = 100$ (training set dimension).

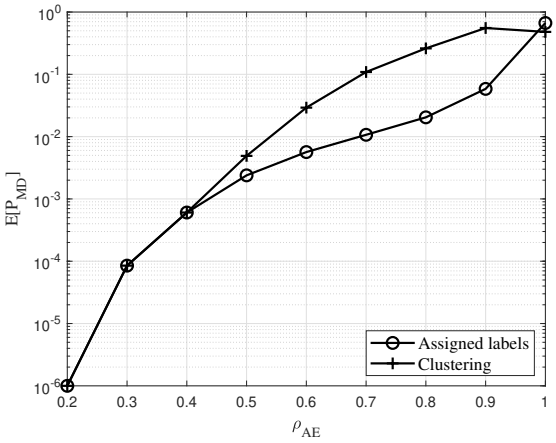
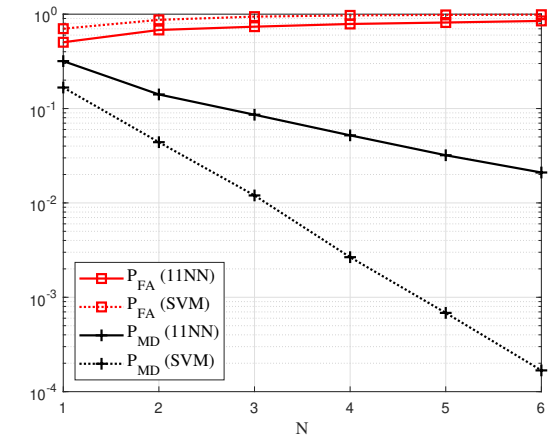
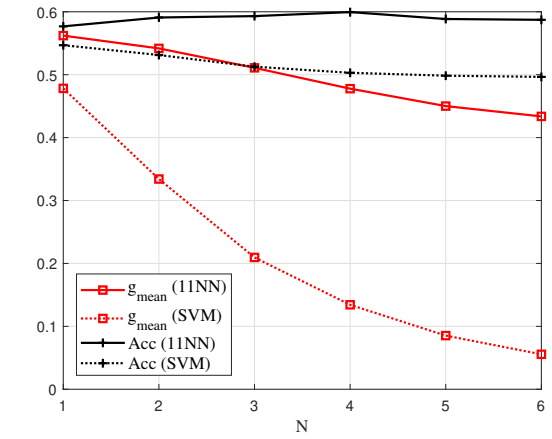


Fig. 5. Average probability of MD versus ρ_{AE} using clustering with respect to known instance labels, with $N = 6$, $M = 200$, $\alpha^{(\text{I})} = \alpha^{(\text{II})} = 1$, $\text{SNR}^{(\text{I})} = 15$dB, $\text{SNR}^{(\text{II})} = 20$dB, classification performed using a SVM.



(a)



(b)

Fig. 4. a) Probabilities of FA and MD and b) g_{MEAN} and ACC as function of the number of sub-carriers N, comparing 11NN and SVM algorithms, with $\alpha^{(\text{I})} = 1$ and $\alpha^{(\text{II})} = 0.9$, $\rho_{\text{AE}} = 0.8$, $\text{SNR}^{(\text{I})} = 15$dB, $\text{SNR}^{(\text{II})} = 20$dB and $M = 300$.

procedure based on clustering, and in particular on k-means algorithm, is compared with one in which instance labels are

TABLE IV
RESULTS FOR DIFFERENT TRAINING SET DIMENSIONS CHOOSING $N = 3$, $\rho_{\text{AE}} = 0.8$ AND $\alpha^{(\text{I})} = 1$.

		dim	10	100	1,000	10,000
$\alpha^{(\text{II})} = 1$	SVM	P_{FA}	$7.4 \cdot 10^{-3}$	$1.93 \cdot 10^{-4}$	$4 \cdot 10^{-6}$	$4 \cdot 10^{-6}$
		P_{MD}	$3.45 \cdot 10^{-3}$	$7.07 \cdot 10^{-3}$	$6.47 \cdot 10^{-3}$	$5.73 \cdot 10^{-3}$
	11NN	P_{FA}	$3.39 \cdot 10^{-3}$	$6.36 \cdot 10^{-3}$	0.055	0.03
		P_{FA}	0.093	0.159	0.085	0.099
$\alpha^{(\text{II})} = 0.8$	SVM	P_{MD}	0.994	0.991	0.992	0.992
		P_{MD}	$3.45 \cdot 10^{-3}$	$7.07 \cdot 10^{-4}$	$6.47 \cdot 10^{-3}$	$5.73 \cdot 10^{-3}$
	11NN	P_{FA}	0.880	0.890	0.904	0.907
		P_{MD}	0.130	0.128	0.114	0.101

exactly known. In both cases, a two-class SVM algorithm with Gaussian kernel has been applied over a training set of 200 samples, 100 from Alice and 100 from Eve. k-means requires as initial choice the knowledge of the number of cluster, which in our case is known by hypothesis and is equal to 2, and the initial partitions. Initial cluster centers, or *centroids*, have been chosen by selecting observations uniformly at random from the dataset. Different initial centroids, however, may produce different final clusters. In order to address this problem, results have been averaged over 50 possible different initial choices. Despite being disadvantaged, clustering methods are able to achieve the same performance of methods with assigned labels for low values of ρ_{AE} but, as expected, they achieve larger P_{MD} as soon as ρ_{AE} increases. The probability of FA remains fixed and lower than 10^{-6}, except for $\rho_{\text{AE}} = 1$.

In Tab. IV we compare the results obtained with training sets of different dimension by using a 11NN (chosen among the other OCNN methods thanks to its low complexity, which makes it able to train a set of dimension 10,000 in reasonable times) and a SVM algorithm. In both cases, we observe that the value of M has a small impact on the results, and the only relevant improvement is given on the P_{FA} of the SVM for $\alpha^{(\text{II})} = 1$. This implies that we can use a small training set without incurring in any significant performance degradation.

Let us now assess the impact of the fading coefficient α in Phase 1 on the authentication performances. In Fig. 6

we exploit a 11NN algorithm and we show that best performances are always obtained in absence of time-varying fading ($\alpha^{(\mathrm{I})} = 1$) during the training phase, outperforming even the case where the value of the time coefficient $\alpha^{(\mathrm{I})}$ remains the same during both phases. Nevertheless this affects the false alarm probability, which experiences a slight worsening. In Fig. 7 we observe the same behavior described above for the LLR test, where we always obtain the lowest missed detection when considering $\alpha^{(\mathrm{I})} = 1$, regardless of value of $\alpha^{(\mathrm{II})}$.

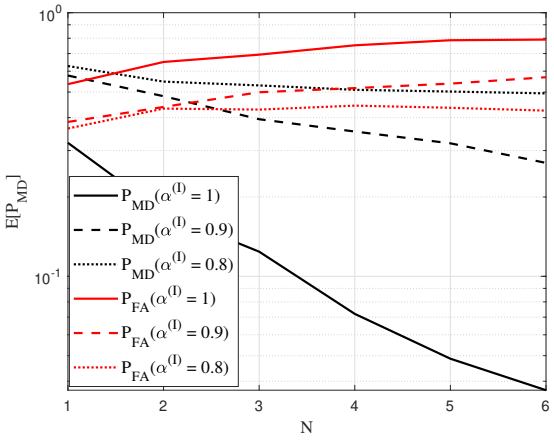


Fig. 6. Probabilities of FA and MD as function of the number of sub-carriers N, 11NN, with $\alpha^{(\mathrm{II})} = 0.9$, $\rho_{\mathrm{AE}} = 0.8$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB and $M = 100$.

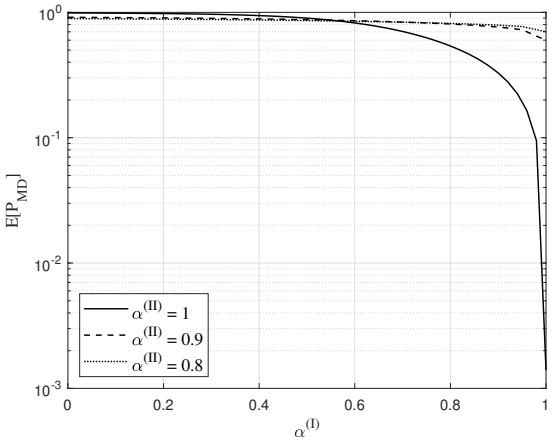


Fig. 7. Probability of MD as function of $\alpha^{(\mathrm{I})}$, LLR test, with different values of $\alpha^{(\mathrm{II})}$, $\rho_{\mathrm{AE}} = 0.1$, $P_{\mathrm{FA}} = 10^{-1}$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB.

C. Comparison of statistical and machine learning methods

In order to compare statistical and machine learning-based methods, let us assess the performance achieved by the statistical decision criteria presented in Section III in terms of average probability of MD. In Fig. 8 the probability of FA has been fixed equal to 10^{-4} for both the examined methods, in absence of fading in the training phase ($\alpha^{(\mathrm{I})} = 1$) and with the spatial correlation coefficient ρ_{AE} set to 0.1, with the meaning of Eve being very far from Bob's position. As a benchmark, we also show the curves obtained for the limit case presented in Sec III-B, which represents a lower bound on the performance achievable. From the results it is evident how much the channel variability (represented by decreasing values of α) degrades the performance of the system, with a significant increase in the probability of MD with respect to the flat fading case (i.e., $\alpha^{(\mathrm{II})} = 1$). Looking at the figure, we note that, with respect to the single LLR test, the combined test helps Bob to enhance the performance of the scheme, and this becomes more evident for increasing numbers of sub-carriers. In addition, we observe that with $\alpha^{(\mathrm{II})} = 0.8$ the performance of the LLR test is very close to the bound, but both of them are outperformed by the combined test.

A similar assessment is reported in Fig. 9, considering binary and one-class versions of SVM and NN algorithms besides LLR test and its bound. The LLR test is equivalent to a one-class statistical method, while the bound corresponds to its binary version, since it considers the presence of Eve's samples in the training set. As it results from the figure, in case of a high spatial correlation machine learning methods exhibit an opposite behavior with respect to the LLR test with or without availability of Eve's samples: given almost the same P_{FA}, in fact, OCC based on SVM and 1KNN achieves a lower probability of MD with respect to their binary counterparts and the LLR test. This is probably due to the presence of negative samples highly correlated with the positive ones in the training set of the binary version, which "confuses" the algorithm and leads it to misclassify new instances with a high occurrence.

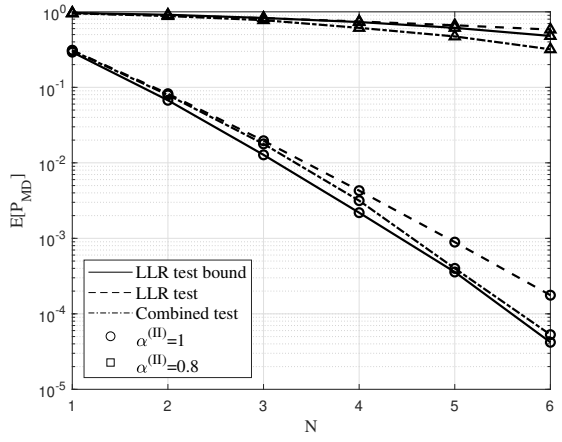


Fig. 8. Average MD probability $\mathbb{E}[P_{\mathrm{MD}}]$ versus number of sub-carriers N, comparing statistical-based test methods, for different values of $\alpha^{(\mathrm{II})}$, $\alpha^{(\mathrm{I})} = 1$, $\rho_{\mathrm{AE}} = 0.1$, with $\mathrm{SNR}^{(\mathrm{I})} = 15$dB and $\mathrm{SNR}^{(\mathrm{II})} = 20$dB.

Further comparisons between statistical and machine learning based methods are discussed in the following. In Fig. 10 we show how the SNR in Phase I affects the authentication performance, exploiting both a LLR test and a SVM classifier. For a fair comparison, we fix as FA target for the LLR test the value of P_{FA} achieved by the SVM classifier. It is possible to observe that the occurrence of MDs increases with the decreasing of the SNR (and the increasing of the variance); this can be explained by considering that for a low SNR Bob is forced to accept a larger set of data in order to guarantee

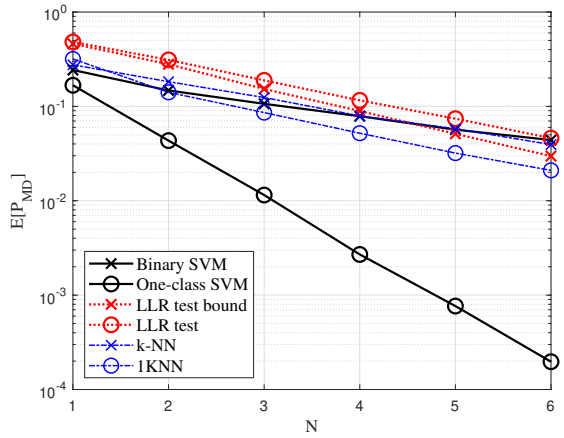


Fig. 9. Average probability of MD as function of the number of subcarriers N, with $\alpha^{(\mathrm{I})} = \alpha^{(\mathrm{II})} = 1$, $\rho_{\mathrm{AE}} = 0.8$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB and $M = 100$, considering $P_{\mathrm{FA}}(\mathrm{SVM}) = \left[1.39 \cdot 10^{-3}, 2.12 \cdot 10^{-4}, 6.85 \cdot 10^{-5}, 1.68 \cdot 10^{-5}, 5.55 \cdot 10^{-6}, 1.62 \cdot 10^{-6}\right]$, $P_{\mathrm{FA}}(1\mathrm{KNN}) = \left[6.76 \cdot 10^{-4}, 4.21 \cdot 10^{-4}, < 10^{-6}, < 10^{-6}, < 10^{-6}, < 10^{-6}\right]$ and $P_{\mathrm{FA}}(\mathrm{LLR}) = 10^{-4}$.

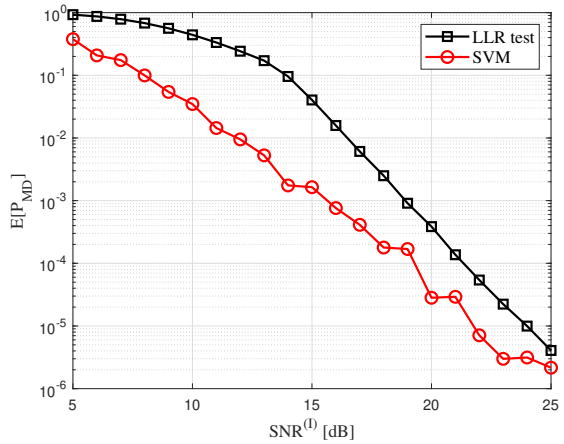


Fig. 10. Average probability of MD using LLR test and SVM algorithm, with $N = 3$, $\rho_{\mathrm{AE}} = 0.5$, $M = 100$, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB, $\alpha^{(\mathrm{I})} = \alpha^{(\mathrm{II})} = 1$. $P_{\mathrm{FA}} = [< 10^{-6}, < 10^{-6}, < 10^{-6}, < 10^{-6}, < 10^{-6}, < 10^{-6}, < 10^{-6}, 4.92 \cdot 10^{-6}, 7.73 \cdot 10^{-5}, 7.11 \cdot 10^{-4}, 4.48 \cdot 10^{-3}, 0.015, 0.0515, 0.117, 0.227, 0.3667, 0.512, 0.637, 0.761]$.

the desired level of FA. In any case, SVM outperforms the LLR test for all the considered values of $\mathrm{SNR}^{(\mathrm{I})}$.

In Fig. 11 we assess performance achieved by different methods varying the attacker's spatial correlation ρ_{AE}, with time-varying fading $\alpha^{(\mathrm{II})} = 0.9$. It is evident, and will be confirmed in the following by looking at Tabs. V and VI, that the use of machine learning approaches is convenient for small values of ρ_{AE}, while for larger values it is advisable to exploit statistical techniques.

In Tables V and VI[2] we show the performance achieved by OCC, considering both NN and SVM algorithms, and we compare them with statistical methods applied to the same data

[2]Probability less than 10^{-6} means that no error has been found over the entire data set. In order to perform an analytical evaluation, these values have been considered equal to 10^{-6}.

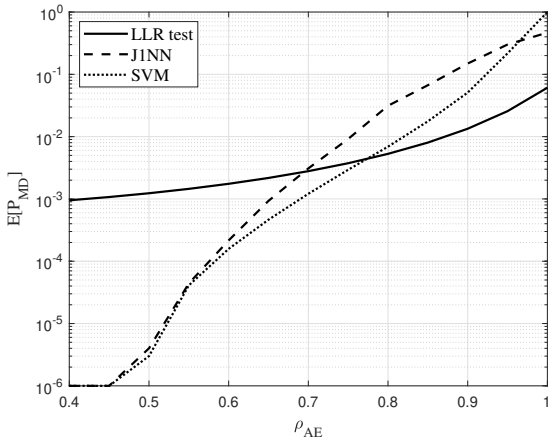


Fig. 11. Average probability of MD using LLR test, J1NN and SVM algorithm, with $N = 3$, $M = 100$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB, $\alpha^{(\mathrm{I})} = 1$,$\alpha^{(\mathrm{II})} = 0.9$. $P_{\mathrm{FA}} = 0.94$.

set (for the sake of brevity we only report the results obtained by one among the four NN classifiers) for increasing number of subcarriers (and thus for increasing number of features), with two different values of the spatial correlation parameter ρ_{AE}. In order to perform a fair comparison between different decision techniques, we choose as a target the probabilities of FA achieved by both the considered OCC algorithms separately, and we compare the resulting probabilities of MD. The lowest probabilities of MD for each case are highlighted in the tables. We observe that in general with low values of ρ_{AE} NN algorithms outperform SVM and statistical methods, in terms of both P_{FA} and P_{MD}. Excellent results are achieved especially in conditions of flat fading and with a large number of channels. When Eve is closer to Bob however, i.e. when the value of ρ_{AE} is large, and with $\alpha^{(\mathrm{II})}$ different from 1, statistical methods (and the LLR test in particular) maintain some advantage over the machine learning techniques. In Tab. VI, where $\rho_{\mathrm{AE}} = 0.8$, SVM algorithms achieve lower probabilities of MD with respect to NN, although they exhibit a higher P_{FA} and therefore a worse balance between the two probabilities, as already shown in Fig. 4(a) and through g_{mean} in Fig. 4(b). We also note that, when the value of the spatial correlation remains constant, we obtain the same probability of MD for both SVM and NN algorithms, probably because ρ_{AE} has no impact on the training phase. The same assertion holds true as regards the probability of FA achieved by SVM algorithms when $\alpha^{(\mathrm{II})}$ does not vary, but not for NN methods.

We finally test the proposed hybrid approach, comparing it with statistical and machine learning methods. In Fig. 12(a) we report the results obtained by considering two different metrics in presence of time-varying fading $\alpha^{(\mathrm{II})} = 0.9$. It is evident that using the LLR metric within a OCNN algorithm (11NN in this particular case) improves the capability of the system to recognize forged messages coming from Eve, while it leads to a small loss in terms of false alarms raised, which becomes negligible with the increase of the number of subcarriers. In Fig. 12(b) we compare the probability of MD achieved by applying a statistical method which exploits a

TABLE V
AVERAGE MD AND FA PROBABILITIES OBTAINED BY DIFFERENT TEST METHODS, FOR DIFFERENT VALUES OF $\alpha^{(\mathrm{II})}$, WITH $\alpha^{(\mathrm{I})} = 1$, $\rho_{\mathrm{AE}} = 0.1$, $\mathrm{SNR}^{(\mathrm{I})} = 15$DB AND $\mathrm{SNR}^{(\mathrm{II})} = 20$DB, $M = 1000$.

	N	1	2	3	4	5	6
$\alpha^{(\mathrm{II})} = 1$	P_{FA} (1KNN)	$6.76{\cdot}10^{-4}$	$4.21 \cdot 10^{-5}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (1KNN)	0.055	$4.74 \cdot 10^{-4}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (LLR)	0.255	0.094	0.044	0.012	0.0029	$6.9 \cdot 10^{-4}$
	P_{MD} (comb)	0.254	0.089	0.036	0.010	0.0019	$5.5 \cdot 10^{-4}$
	P_{FA} (SVM)	$1.39 \cdot 10^{-3}$	$2.12 \cdot 10^{-4}$	$6.85 \cdot 10^{-4}$	$1.68 \cdot 10^{-5}$	$5.55 \cdot 10^{-6}$	$1.62 \cdot 10^{-6}$
	P_{MD} (SVM)	0.059	$2.46 \cdot 10^{-3}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (LLR)	0.233	0.073	0.021	$6.6 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$6.09 \cdot 10^{-4}$
	P_{MD} (comb)	0.233	0.072	0.045	0.009	0.001	$7.7 \cdot 10^{-4}$
$\alpha^{(\mathrm{II})} = 0.8$	P_{FA}	0.525	0.632	0.745	0.847	0.930	0.949
	P_{MD} (1KNN)	0.055	$4.74 \cdot 10^{-4}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (LLR)	0.176	0.058	0.016	0.0037	$5.78 \cdot 10^{-4}$	$1.71 \cdot 10^{-4}$
	P_{MD} (comb)	0.228	0.011	0.0027	$3.9 \cdot 10^{-5}$	$7 \cdot 10^{-6}$	$< 10^{-6}$
	P_{FA} (SVM)	0.830	0.953	0.981	0.988	0.989	0.990
	P_{MD} (SVM)	0.059	$2.46 \cdot 10^{-3}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (LLR)	0.054	$4.9 \cdot 10^{-3}$	$6.75 \cdot 10^{-4}$	$5.57 \cdot 10^{-5}$	$2.05 \cdot 10^{-5}$	$< 10^{-6}$
	P_{MD} (comb)	0.043	$1.07 \cdot 10^{-3}$	$5.7 \cdot 10^{-5}$	10^{-6}	$< 10^{-6}$	$< 10^{-6}$

TABLE VI
AVERAGE MD AND FA PROBABILITIES OBTAINED BY DIFFERENT TEST METHODS, FOR DIFFERENT VALUES OF $\alpha^{(\mathrm{II})}$, WITH $\alpha^{(\mathrm{I})} = 1$, $\rho_{\mathrm{AE}} = 0.8$, $\mathrm{SNR}^{(\mathrm{I})} = 15$DB AND $\mathrm{SNR}^{(\mathrm{II})} = 20$DB, $M = 1000$.

	N	1	2	3	4	5	6
$\alpha^{(\mathrm{II})} = 1$	P_{FA}	$6.76 \cdot 10^{-4}$	$4.21 \cdot 10^{-5}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
	P_{MD} (1KNN)	0.318	0.141	0.086	0.052	0.032	0.021
	P_{MD} (LLR)	0.531	0.378	0.319	0.183	0.099	0.052
	P_{MD} (comb)	0.592	0.380	0.279	0.168	0.077	0.047
	P_{FA} (SVM)	$1.39 \cdot 10^{-3}$	$2.12 \cdot 10^{-4}$	$6.85 \cdot 10^{-5}$	$1.68 \cdot 10^{-5}$	$5.55 \cdot 10^{-6}$	$1.62 \cdot 10^{-6}$
	P_{MD} (SVM)	0.167	0.044	0.012	$2.66 \cdot 10^{-3}$	$6.83 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$
	P_{MD} (LLR)	0.494	0.313	0.189	0.120	0.074	0.047
	P_{MD} (comb)	0.495	0.316	0.188	0.113	0.068	0.044
$\alpha^{(\mathrm{II})} = 0.8$	P_{FA}	0.684	0.867	0.918	0.955	0.974	0.983
	P_{MD} (1KNN)	0.318	0.141	0.086	0.052	0.032	0.021
	P_{MD} (LLR)	0.196	0.052	0.021	0.007	0.003	0.001
	P_{MD} (comb)	0.319	0.124	0.073	0.040	0.023	0.016
	P_{FA} (SVM)	0.830	0.953	0.981	0.988	0.989	0.990
	P_{MD} (SVM)	0.167	0.044	0.012	$2.66 \cdot 10^{-3}$	$6.83 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$
	P_{MD} (LLR)	0.102	0.017	0.004	$1.8 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$6.84 \cdot 10^{-4}$
	P_{MD} (comb)	0.166	0.043	0.018	$9.82 \cdot 10^{-3}$	$9.61 \cdot 10^{-3}$	$7.75 \cdot 10^{-3}$

LLR test and the proposed hybrid approach, given the same probability of FA. Also in this case, the hybrid method has some advantage, especially with a large number of subcarriers. We have therefore shown that the proposed approach improves authentication performance of both statistical techniques and machine learning methods based on NN algorithms. Nevertheless, statistical approaches maintain the advantage of having a very low complexity with respect to machine learning techniques.

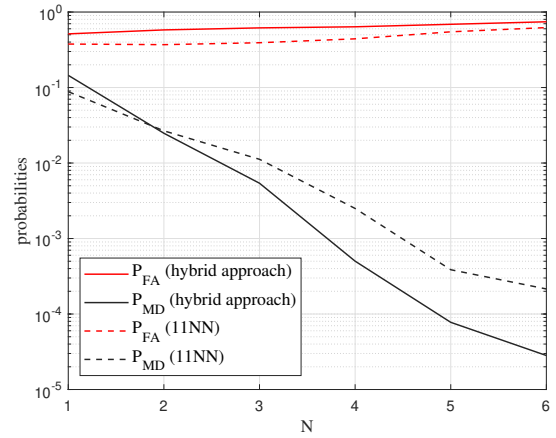
VII. CONCLUSION

We have assessed the performance achieved by different decision techniques in a physical layer authentication scenario with time-varying fading and in presence of an attacker. We have considered different methods based on both statistical criteria and machine learning algorithms. We have shown that using a large training set that includes different realizations of the time-varying fading is not beneficial from the security point of view. We have also shown how clustering algorithms can help to avoid the use of higher layer authentication techniques in the initial phase, with only a small loss in terms of performance when a medium-large spatial correlation with the attacker channel exists. Somehow unexpectedly, in the same conditions the use of binary classification algorithms does not bring any advantage over their OCC counterparts. Our results demonstrate that NN techniques are able to achieve a better trade-off between the FA and MD probabilities than the other considered classification methods. Moreover, they always result to be the best choice with low values of the spatial correlation, while in the other cases the application of statistical techniques leads to better performance.

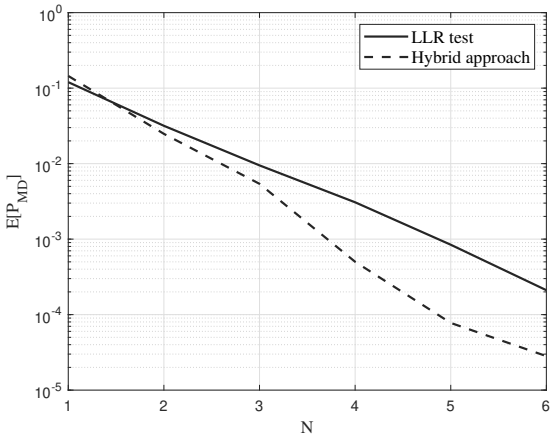
REFERENCES

[1] A. Mukherjee, "Physical-layer security in the internet of things: Sensing and communication confidentiality under resource constraints," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1747–1761, Oct 2015.

[2] L. Sun and Q. Du, "A review of physical layer security techniques for internet of things: Challenges and solutions," *Entropy*, vol. 20, no. 10, p. 730, 2018.



(a)



(b)

Fig. 12. a) Probabilities of FA and MD obtained by using a hybrid approach and a 11NN with Euclidean metric, and b) Probability of MD achieved by LLR test and hybrid approach, with $M = 100$, $\alpha^{(\mathrm{I})} = 1$, $\alpha^{(\mathrm{II})} = 0.9$, $\rho_{\mathrm{AE}} = 0.1$, $\mathrm{SNR}^{(\mathrm{I})} = 15$dB, $\mathrm{SNR}^{(\mathrm{II})} = 20$dB.

[3] N. Zhang, D. Chen, F. Ye, T. Zheng, and Z. Wei, "Physical layer security for internet of things," *Wireless Communications and Mobile Computing*, vol. 2019, 2019. [Online]. Available: https://doi.org/10.1155/2019/2627938

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

[5] J. Zhang, S. Rajendran, Z. Sun, R. Woods, and L. Hanzo, "Physical layer security for the internet of things: Authentication and key generation," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 92–98, October 2019.

[6] H. Wang, T. Zheng, J. Yuan, D. Towsley, and M. H. Lee, "Physical layer security in heterogeneous cellular networks," *IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1204–1219, March 2016.

[7] M. Baldi, L. Senigagliesi, and F. Chiaraluce, "On the security of transmissions over fading wiretap channels in realistic conditions," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[8] G. Caparra, M. Centenaro, N. Laurenti, S. Tomasin, and L. Vangelista, *Wireless Physical-Layer Authentication for the Internet of Things*. Cambridge University Press, 2017, pp. 390—418.

[9] P. Baracca, N. Laurenti, and S. Tomasin, "Physical layer authentication over mimo fading wiretap channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2564–2573, July 2012.

[10] U. Maurer, "Authentication theory and hypothesis testing," in *Proc. International Symposium on Information Theory (ISIT 2000)*, vol. 46, no. 4, jul 2000, pp. 1350–1356.

[11] A. Weinand, M. Karrenbauer, R. Sattiraju, and H. Schotten, "Application of machine learning for channel based message authentication in mission critical machine type communication," in *European Wireless 2017; 23th European Wireless Conference*, May 2017, pp. 1–5.

[12] N. Wang, T. Jiang, S. Lv, and L. Xiao, "Physical-layer authentication based on extreme learning machine," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1557–1560, July 2017.

[13] N. Xie and S. Zhang, "Blind authentication at the physical layer under time-varying fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 7, pp. 1465–1479, July 2018.

[14] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2571–2579, July 2008.

[15] J. K. Tugnait, "Using artificial noise to improve detection performance for wireless user authentication in time-variant channels," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 377–380, Aug 2014.

[16] F. Pan, H. Wen, R. Liao, Y. Jiang, A. Xu, K. Ouyang, and X. Zhu, "Physical layer authentication based on channel information and machine learning," in *2017 IEEE Conference on Communications and Network Security (CNS)*, Oct 2017, pp. 364–365.

[17] H. Fang, X. Wang, and S. Tomasin, "Machine learning for intelligent authentication in 5G and beyond wireless networks," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 55–61, October 2019.

[18] G. Baldini, R. Giuliani, and G. Steri, "Physical layer authentication and identification of wireless devices using the synchrosqueezing transform," *MDPI Applied Sciences*, vol. 8, Nov 2018.

[19] J. Yoon, Y. Lee, and E. Hwang, "Machine learning-based physical layer authentication using neighborhood component analysis in mimo wireless communications," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 63–65.

[20] T. M. Hoang, N. M. Nguyen, and T. Q. Duong, "Detection of eavesdropping attack in uav-aided wireless systems: Unsupervised learning with one-class svm and k-means clustering," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 139–142, 2020.

[21] H. Fang, X. Wang, and L. Hanzo, "Learning-aided physical layer authentication as an intelligent process," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2260–2273, March 2019.

[22] X. Qiu, J. Dai, and M. Hayes, "A learning approach for physical layer authentication using adaptive neural network," *IEEE Access*, vol. 8, pp. 26 139–26 149, 2020.

[23] L. Senigagliesi, M. Baldi, and E. Gambi, "Statistical and machine learning-based decision techniques for physical layer authentication," in *2019 IEEE Global Communications Conference (Globecom)*, Dec 2019.

[24] L. Senigagliesi, L. Cintioni, M. Baldi, and E. Gambi, "Blind physical layer authentication over fading wireless channels through machine learning," in *2019 IEEE Workshop on Information Forensics and Security (WIFS)*, Dec 2019.

[25] A. Wiesel, J. Goldberg, and H. Messer-Yaron, "SNR estimation in time-varying fading channels," *IEEE Transactions on Communications*, vol. 54, no. 5, pp. 841–848, May 2006.

[26] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.

[27] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *In Proceedings of the Fourteenth International Conference on Machine Learning*. Morgan Kaufmann, 1997, pp. 179–186.

[28] S.Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1993.

[29] R. S. Hoyt, "Probability functions for the modulus and angle of the normal complex variate," *The Bell System Technical Journal*, vol. 26, no. 2, pp. 318–359, April 1947.

[30] J. F. Paris, "Nakagami-q (hoyt) distribution function with applications," *Electronics Letters*, vol. 45, no. 4, pp. 210–211, February 2009.

[31] S. Tomasin, "Analysis of channel-based user authentication by key-less and key-based approaches," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 5700–5712, Sep. 2018.

[32] S. S. Khan and M. G. Madden, "A survey of recent trends in one class classification," in *Artificial Intelligence and Cognitive Science*, L. Coyle and J. Freyne, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 188–197.

[33] D. M. J. Tax, "One-class classification: Concept learning in the absence of counter-examples," Ph.D. dissertation, Technische Universiteit Delft, 2001.

[34] S. S. Khan and A. Ahmad, "Relationship between variants of one-class nearest neighbors and creating their accurate ensembles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1796–1809, Sep. 2018.
[35] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
[36] D. M. Tax and R. P. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.
[37] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, pp. 264–323, Sep 1999.
[38] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *5-th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967, pp. 281–297.
[39] P. E. H. R. O. Duda and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2000.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.