

# Location-Verification and Network Planning via Machine Learning Approaches

Alessandro Brighente, Francesco Formaggio, Marco Centenaro, Giorgio Maria Di Nunzio, and Stefano Tomasin  
Department of Information Engineering, University of Padova, via G. Gradenigo 6/B, Padova, Italy. first.lastname@dei.unipd.it

**Abstract**—In-region location verification (IRLV) in wireless networks is the problem of deciding if user equipment (UE) is transmitting from inside or outside a specific physical region (e.g., a safe room). The decision process exploits the features of the channel between the UE and a set of network access points (APs). We propose a solution based on machine learning (ML) implemented by a neural network (NN) trained with the channel features (in particular, noisy attenuation values) collected by the APs for various positions both inside and outside the specific region. The output is a decision on the UE position (inside or outside the region). By seeing IRLV as an hypothesis testing problem, we address the optimal positioning of the APs for minimizing either the area under the curve (AUC) of the receiver operating characteristic (ROC) or the cross entropy (CE) between the NN output and ground truth (available during the training). In order to solve the minimization problem we propose a two-stage particle swarm optimization (PSO) algorithm. We show that for a long training and a NN with enough neurons the proposed solution achieves the performance of the Neyman-Pearson (N-P) lemma.

**Index Terms**—Physical layer security, location verification, neural network, network planning.

## I. INTRODUCTION

Applications using information on the user location are rapidly spreading, also to ensure that some services are obtained only in pre-determined areas. In order to establish the user position we can rely on the device itself, requested to report the position provided by its GPS module. However, tempering with the GPS module or its software interface is relatively easy [1]. Thus, more reliable solutions must be explored. Location verification systems aim at verifying the position of devices [2], [3], possibly using distance measures obtained, for example, through the received signal strength (RSS) at anchor nodes for signals transmitted by the terminal under verification. This problem is closely related to *user authentication* at the physical layer, where wireless channel features are exploited to verify the sender of a message [4].

We focus here on in-region location verification (IRLV), the problem of deciding whether a message coming from a terminal over a wireless network has been originated from a specific physical region, e.g., a safe room, or not [2]. IRLV can be seen as an hypothesis testing problem between two alternatives, namely being inside or outside the specific region. Among proposed solutions, we recall distance bounding techniques with rapid exchanges of packets between the verifier and the prover [5], also using radio-frequency and ultrasound signals [6], and solutions based on anchor nodes and increasing transmit power by the sender [7]. More recently,

a delay-based verification technique has been proposed in [8], leveraging geometric properties of triangles, which prevent an adversary from manipulating measured delays.

In this paper, we consider the IRLV problem for a user equipment (UE) connected to a set of network access points (APs). The decision on the user position is taken on the basis of observed features of the channel over which communication occurs. For exemplary purposes we focus here on the observation of the attenuation of the channels between the UE and the APs. We propose a machine learning (ML) approach where i) channel measurements are collected by trusted nodes both inside and outside the region of interest (ROI), ii) a machine is trained to take decisions between the two hypotheses, iii) the machine is exploited to take decisions on the unknown UEs in the exploitation phase. ML techniques have already found application in user authentication (see [9] and references therein), however never in IRLV, to the best of authors' knowledge. The neural network (NN) training is based on the cross entropy (CE), and, framing IRLV problem into an hypothesis testing problem, we establish the optimality of this criterion according to the Neyman-Pearson (N-P) lemma, in asymptotic conditions.

Then, we address the problem of optimum positioning of the APs (network planning) for IRLV. Two metrics are considered for this optimization: a) the CE of NN training, and b) the area under the curve (AUC) of the receiver operating characteristic (ROC) of the hypothesis test. While CE is directly related to the NN training, the AUC is more connected to the final performance that we expect from IRLV. A limited number of neurons, as well as a limited size of the training set may provide different results for the two metrics. For the optimization of APs position we propose a two-stage particle swarm optimization (PSO) algorithm, minimizing either the CE or the ROC AUC. Simulation results over channels with shadow fading show the merits of the proposed solution, and its effectiveness in providing reliable IRLV.

## II. SYSTEM MODEL

We consider a cellular system with  $N_{AP}$  APs covering a region  $\mathcal{A}$  over a plane. We propose a IRLV system able to determine if a UE is transmitting from inside an *authorized* sub-region  $\mathcal{A}_0 \subset \mathcal{A}$ . The dependency on location of the UE-APs channels is exploited to distinguish between transmissions from inside and outside  $\mathcal{A}_0$ . Transmissions are assumed to be narrowband and the channel feature used for IRLV is its attenuation.

The IRLV procedure comprises two phases. In the first phase, named identification or training, a trusted UE transmits a training signal (known at the AP) from various points inside region  $\mathcal{A}_0$ . The APs estimate the channel attenuations and store them in association with  $\mathcal{A}_0$ . Some external authentication technique on the transmitted packet ensures that the received signal upon which the attenuation is estimated is actually transmitted by the trusted UE. Similarly, attenuation values are collected when the trusted UE transmits from the complementary region  $\mathcal{A}_1$  and stored by the APs in association to  $\mathcal{A}_1 = \mathcal{A} \setminus \mathcal{A}_0$ . In the second phase, named verification or exploitation, the UE transmits a known training sequence from any point in  $\mathcal{A}$  and the IRLV system must decide whether the UE is in region  $\mathcal{A}_0$  or  $\mathcal{A}_1$ .

#### A. Channel Model

Let  $\mathbf{x}_{\text{ap}}^{(n)} = (X_{\text{ap}}^{(n)}, Y_{\text{ap}}^{(n)})$  be the position of the  $n$ -th AP. For a UE located at  $\mathbf{x}_{\text{ue}} = (X_u, Y_u)$ , its distance from AP  $n$  is denoted as  $L(\mathbf{x}_{\text{ue}}, \mathbf{x}_{\text{ap}}^{(n)})$ . We assume that the UE transmits with constant power so that AP  $n$  can estimate the attenuation  $a(n)$  incurred over the channel, including the effects of path-loss and shadowing. Let  $\mathbf{a} = [a(1), a(2), \dots, a(N_{\text{AP}})]$  collect attenuation values from all AP.

Denoting the path-loss coefficient as  $a_{\text{PL}}(n)$  the shadowing component is log-normally distributed, i.e.,  $(a_{\text{S}}(n))_{\text{dB}} \sim \mathcal{N}(0, \sigma_s^2)$ , and we have

$$(a(n))_{\text{dB}} = (a_{\text{PL}}(n))_{\text{dB}} + (a_{\text{S}}(n))_{\text{dB}}. \quad (1)$$

The channel model for path-loss and shadowing is derived from [10]. For a line of sight (LOS) link the path loss coefficient in dB is modelled as

$$a_{\text{PL-LOS}}(n) = 20 \log_{10} \left( \frac{f_0 4\pi L(\mathbf{x}_{\text{UE}}, \mathbf{x}_{\text{AP}}^{(n)})}{c} \right), \quad (2)$$

where  $f_0$  is the carrier frequency, and  $c$  is the speed of light. For a non-LOS link the path loss coefficient in dB is defined as

$$\begin{aligned} a_{\text{PL-NLOS}}(n) &= 40 \left( 1 - 4 \cdot 10^{-3} h_{\text{AP}}|_{\text{m}} \right) \times \\ &\log_{10} \left( \frac{L(\mathbf{x}_{\text{UE}}, \mathbf{x}_{\text{AP}}^{(n)})}{10^3} \Big|_{\text{m}} \right) - 18 \log_{10} h_{\text{AP}}|_{\text{m}} + \\ &+ 21 \log_{10} \left( \frac{f_0|_{\text{MHz}}}{10^6} \right) + 80, \end{aligned} \quad (3)$$

where  $h_{\text{AP}}|_{\text{m}}$  is the AP antenna elevation in meters,  $f_0|_{\text{MHz}}$  is the carrier frequency in MHz, and  $L(\mathbf{x}_{\text{UE}}, \mathbf{x}_{\text{AP}}^{(n)})|_{\text{m}}$  is the UE-AP  $n$  distance in meters. We assume that correlation between shadowing coefficients  $(a_{\text{S}}(i))_{\text{dB}}$  and  $(a_{\text{S}}(j))_{\text{dB}}$  for two APs located in  $\mathbf{x}_i$  and  $\mathbf{x}_j$  when the UE is transmitting is

$$\mathbb{E}_{\mathbf{x}} [a_{\text{S}}(i)a_{\text{S}}(j)] = \sigma_s^2 \exp \left( -\frac{L(\mathbf{x}_i, \mathbf{x}_j)}{d_c} \right), \quad (4)$$

where  $d_c$  is the shadowing decorrelation distance and  $\mathbb{E}_{\mathbf{x}}[\cdot]$  is the expected value with respect to the distribution of  $\mathbf{x}$ .

### III. IN-REGION LOCATION VERIFICATION

Let us define the two hypotheses of the IRLV problem as

$$\mathcal{H}_0 : \text{the UE is in } \mathcal{A}_0, \quad \mathcal{H}_1 : \text{the UE is in } \mathcal{A}_1. \quad (5)$$

In the training phase the UE transmits from  $S$  locations. For transmission  $i = 1, \dots, S$ , let  $\mathbf{a}^{(i)} = [a^{(i)}(1), \dots, a^{(i)}(N_{\text{AP}})]$  be the vector of measured attenuations. We associate to  $\mathbf{a}^{(i)}$  the label  $t_i = 0$  if UE is transmitting from inside  $\mathcal{A}_0$  (hypothesis  $\mathcal{H}_0$ ), and  $t_i = 1$  otherwise. Let also define  $\mathcal{T} = \{\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(S)}\}$ . By using these attenuation training vectors and labels, we aim at building a function

$$\hat{t} = g(\mathbf{a}) \in \{0, 1\} \quad (6)$$

that maps any attenuation vector  $\mathbf{a}$  into a decision on the location of the UE. We would like to have  $\hat{t} = 0$  if  $\mathbf{a}$  was obtained when the UE was inside  $\mathcal{A}_0$  and  $\hat{t} = 1$  otherwise.

The performance of the IRLV system is assessed in terms of two error probabilities: the false alarm (FA) probability, i.e., the probability that a UE in  $\mathcal{A}_0$  is declared outside this region, and the misdetection (MD) probability, i.e., the probability that a UE outside  $\mathcal{A}_0$  is declared inside the region. In formulas, denoting with  $\mathbb{P}[\cdot]$  the probability function,

$$P_{\text{FA}} = \mathbb{P}[\hat{t} = 1 | \mathcal{H}_0], \quad P_{\text{MD}} = \mathbb{P}[\hat{t} = 0 | \mathcal{H}_1]. \quad (7)$$

#### A. Test for Known Attenuation Statistics

The IRLV problem can be seen as an hypothesis testing problem between the two hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . When the statistics of the attenuation vectors are known under the two hypotheses, the most powerful test for the IRLV problem is provided by the N-P lemma. In particular, let us define the log likelihood-ratio (LLR)

$$\mathcal{L}(\mathbf{a}) = \log \left( \frac{p_{\mathbf{a}}(\mathbf{a} | \mathcal{H}_0)}{p_{\mathbf{a}}(\mathbf{a} | \mathcal{H}_1)} \right), \quad (8)$$

where  $p_{\mathbf{a}|\mathcal{H}_i}(\mathbf{a} | \mathcal{H}_i)$  is the probability density function (PDF) of the random vector  $\mathbf{a}$  modelling all attenuation values  $\mathbf{a}$ , given that hypothesis  $\mathcal{H}_i$  is verified, and  $\log$  denotes the base-2 logarithm. The N-P test function is

$$\hat{t} = g(\mathbf{a}) = \begin{cases} 0 & \mathcal{L}(\mathbf{a}) \geq \theta, \\ 1 & \mathcal{L}(\mathbf{a}) < \theta, \end{cases} \quad (9)$$

where  $\theta$  is a threshold to be chosen in order to ensure the desired FA probability. This test ensures that for the given FA probability the MD probability is minimized.

#### B. Example of N-P Test

We now describe an example of application of the N-P test, where we can easily obtain a close-form expression for  $f(\mathbf{a})$ .

Let us define the overall network region as a circle  $\mathcal{A}_c$  with radius  $R_{\text{out}}$  and consider a single AP located at the center of  $\mathcal{A}_c$ . Consider  $\mathcal{A}_0$  as a rectangle with nearest point to the center of  $\mathcal{A}_c$  at a distance  $R_{\text{min}}$ . The outside region is  $\mathcal{A}_1 = \mathcal{A}_c \setminus \mathcal{A}_0$ . In the LOS scenario the scalar attenuation  $a$  incurred by a UE is given by path loss, which only depends on its relative

distance to the AP. Considering an attenuation value  $a$ , the UE-AP distance is given by

$$R = \frac{ca}{4\pi f_0}. \quad (10)$$

Therefore, instead of considering  $p_{a|\mathcal{H}_i}(a|\mathcal{H}_i)$  we consider  $p_{R|\mathcal{H}_i}(r|\mathcal{H}_i)$ , where distance  $R$  corresponds to attenuation  $a$ . We first derive the cumulative distribution function (CDF) of  $R$  in  $\mathcal{A}_0$ , i.e., the probability that the UE is located in  $\mathcal{A}_0$  at a distance  $R \leq r$  from the AP. This is

$$\mathbb{P}[R \leq r|\mathcal{H}_0] = \frac{1}{|\mathcal{A}_0|} \int_{R_{\min}}^r \rho \alpha(\rho) d\rho, \quad (11)$$

where  $\alpha(R)$  denotes the angle of the circular sector measured from a distance  $R$  and intersecting region  $\mathcal{A}_0$  and  $|\mathcal{A}_0|$  is the area of region  $\mathcal{A}_0$ . Then by taking the derivative of the CDF (11) with respect to  $r$  we obtain the PDF

$$p_{R|\mathcal{H}_0}(r|\mathcal{H}_0) = \frac{1}{|\mathcal{A}_0|} r \alpha(r). \quad (12)$$

Following the same reasoning and considering that the length of the circular sector with radius  $r$  located in  $\mathcal{A}_1$  is  $2\pi - \alpha(r)$ , we obtain the PDF of transmission from a distance  $r$  in  $\mathcal{A}_1$  as

$$p_{R|\mathcal{H}_1}(r|\mathcal{H}_1) = \frac{1}{|\mathcal{A}_1|} r (2\pi - \alpha(r)). \quad (13)$$

From (12) and (13) we obtain the LLR as a function of the UE's distance from the AP as

$$\mathcal{L}(a) = \log \left[ \frac{|\mathcal{A}_1| \alpha\left(\frac{ca}{f_0 4\pi}\right)}{|\mathcal{A}_0| \left(2\pi - \alpha\left(\frac{ca}{f_0 4\pi}\right)\right)} \right]. \quad (14)$$

### C. Neural Network Implementation

Under more complicated scenarios, it becomes hard to obtain close-form expressions for the LLR. Therefore, we resort to a ML approach, using a NN trained with attenuation vectors  $\mathbf{a}^{(i)}$  and labels  $t_i \in \{0, 1\}$ . In the verification phase the trained NN is used on the test attenuation vectors  $\mathbf{a}$  to provide the decision  $\hat{t} \in \{0, 1\}$ . Now,  $g(\cdot)$  is the function implemented by the NN.

We now provide a short description of a NN. A feed-forward NN processes the input in stages, named layers, where the output of one layer is the input of the next layer. The input of the NN is  $\mathbf{y}^{(0)} = \mathbf{a}$ , and layer  $\ell - 1$  has  $N^{(\ell-1)}$  outputs obtained by processing the inputs with  $N^{(\ell-1)}$  scalar functions named neurons. The output of the  $n$ -th neuron of the  $\ell$ -th layer is

$$y_n^{(\ell)} = \sigma \left( \mathbf{w}_n^{(\ell-1)} \mathbf{y}^{(\ell-1)} + b_n^{(\ell)} \right), \quad (15)$$

where  $\mathbf{w}_n^{(\ell-1)}$  and  $b_n^{(\ell)}$  are coefficients to be determined in the training phase, and  $\sigma(\cdot)$  is the sigmoid activation function. The last layer comprises only one neuron,  $y^{(L)}$ , and the final output of the NN is the scalar

$$\tilde{t}(\mathbf{a}) \triangleq \sigma(y^{(L)}), \quad (16)$$

where  $L$  is the total number of layers. Finally, the test function is obtained by thresholding  $\tilde{t}(\mathbf{a})$ , i.e.,

$$g(\mathbf{a}) = \begin{cases} 1 & \tilde{t}(\mathbf{a}) > \lambda \\ 0 & \tilde{t}(\mathbf{a}) \leq \lambda. \end{cases} \quad (17)$$

By varying  $\lambda$  we obtain different values of  $P_{\text{FA}}$  and  $P_{\text{MD}}$  for this IRLV test.

Various options have been proposed in the literature for NN training. We consider here as objective function the empirical CE between the NN output and the labels  $t_i$ , defined as

$$\hat{H}_{p_{\mathcal{H}_i|\mathbf{a}}}(g) \triangleq - \frac{1}{S} \sum_{i=1}^S \left[ t_i \log \tilde{t}(\mathbf{a}^{(i)}) + (1 - t_i) \log (1 - \tilde{t}(\mathbf{a}^{(i)})) \right]. \quad (18)$$

Training is performed with the gradient descent algorithm minimizing  $\hat{H}_{p_{\mathcal{H}_i|\mathbf{a}}}(g)$ .

In the following we show that a CE based NN is equivalent, in probability and for perfect training, to the N-P solution. First, we prove that the output of the NN can be interpreted as the class conditional probability.

**Theorem 1.** *Let  $g(\mathbf{a}) \in [0, 1]$  be the output of a NN obtained with perfect training, i.e., with infinite number of training points, layers and neurons. Let the training be performed with the CE metric. Then*

$$g(\mathbf{a}) = p_{\mathcal{H}_1|\mathbf{a}}(\mathcal{H}_1|\mathbf{a}) \quad (19)$$

almost surely.

*Proof.* See Appendix A.  $\square$

Note that this approach does not require the knowledge of the statistics of  $\mathbf{a}$  under the two hypotheses, while, instead, it requires a large enough set of training points to converge. However, at convergence, the NN achieves the same performance of the NP approach. This holds since (17) is equivalent to (9), i.e., they provide the same ROC, with

$$\tilde{t} = p_{\mathcal{H}_1|\mathbf{a}}(\mathcal{H}_1|\mathbf{a}), \quad \theta = \frac{1 - \lambda}{\lambda} \frac{p_{\mathcal{H}_1}(\mathcal{H}_0)}{p_{\mathcal{H}_1}(\mathcal{H}_1)}, \quad (20)$$

where (20) is a direct consequence of the Bayes rule applied as follows

$$p_{\mathcal{H}_1|\mathbf{a}}(\mathcal{H}_1|\mathbf{a}) = \left[ 1 + \frac{p_{\mathcal{H}_1}(\mathcal{H}_1)}{p_{\mathcal{H}_1}(\mathcal{H}_0)} 2^{\mathcal{L}(\mathbf{a})} \right]^{-1}. \quad (21)$$

## IV. NETWORK PLANNING

As the attenuation depends on the position of the APs and on the surrounding environment, the performance of the authentication system depends on the number of APs and on their location. In this Section, we derive an approach to optimally locate APs (*network planning*) so that the authentication system attains the best performance.

For APs positioning we consider as performance metric a suitable trade-off between FA and MD probabilities. In particular, the ROC curve associates the  $P_{\text{MD}}$  with the corresponding  $P_{\text{FA}}$ , for all possible values of thresholds  $\lambda$ .

However, as we aim at using a single performance measure without setting a priori  $P_{\text{FA}}$ , we resort to the ROC AUC [11], defined as

$$C(\{\mathbf{x}_{\text{AP}}^{(n)}\}) = \int_0^1 P_{\text{MD}}(P_{\text{FA}}) dP_{\text{FA}}, \quad (22)$$

where  $P_{\text{MD}}(P_{\text{FA}})$  is the  $P_{\text{MD}}$  value as a function of  $P_{\text{FA}}$ . In (22) we have highlighted the dependency of the AUC on the AP positions. Note that  $C(\{\mathbf{x}_{\text{AP}}^{(n)}\})$  is the integral of the ROC function. Therefore, the AP position optimization aims at minimizing the AUC, i.e.

$$\text{argmin}_{\{\mathbf{x}_{\text{AP}}^{(n)}\}} C(\{\mathbf{x}_{\text{AP}}^{(n)}\}). \quad (23)$$

Note that minimizing the AUC is equivalent to minimizing the average  $P_{\text{MD}}$  under the assumption of a uniformly distributed  $P_{\text{FA}}$ . In practice, in order to compute the AUC we must run the NN over the training set multiple times, with different thresholds and find the corresponding ROC curve, before performing its integral by numerical methods.

We hence propose to exploit the training process of the NN and use the CE, readily provided at the end of training, as a proxy of the system performance, avoiding an explicit estimation of ROC AUC. This is also motivated by Theorem 1, as the lower CE, the more a NN approaches N-P, which is the optimal solution. Recalling that the training minimization problem is non-convex, the same training set can lead to NNs with different classification performance and hence different AUCs. However, we select the one minimizing the CE, which is expected to have the minimum AUC.

### A. Particle Swarm Optimization

In order to solve the network planning problem (23) we resort to the PSO method [12], which is an iterative algorithm performing the simultaneous optimization of different points. This is similar to the multi-start solution for non-convex optimization, where local minima are avoided by selecting among different descent paths the one providing the minimum solution.

The PSO method is briefly recalled here. PSO is an iterative optimization algorithm based on social behavior of animals, e.g., birds flocking and fish schools. Consider  $P$  particles, where particle  $p = 1, \dots, P$ , is described by a vector of APs positions  $\mathbf{x}_p = [\mathbf{x}_{\text{AP}}^{(1)}(p), \dots, \mathbf{x}_{\text{AP}}^{(N_{\text{ap}})}(p)]$ , and by its velocity  $\mathbf{v}_p$ . Each particle is a candidate solution of the optimization problem. Starting from particles at random positions and velocities, at each iteration both positions and velocities are updated. Two optimal values are defined in each iteration: the global optimum found so far in the entire particle population, and a local optimum for each particle, i.e., the optimal value found by the individual  $p$  up to the current iteration. We define as  $\mathbf{o}_G$  the position of the the global optimal values and as  $\mathbf{o}_p$  the position of the optimal value found by particle  $p$  at the current iteration. The optimal values are those minimizing the selected objective function.

The position and velocity of the particles are updated at iteration  $\ell$  as [12]

$$\begin{aligned} \mathbf{v}_p(\ell) &= \omega \mathbf{v}_p(\ell - 1) + \phi_1(\ell)(\mathbf{o}_p(\ell - 1) - \\ & - \mathbf{x}_p(\ell - 1)) + \phi_2(\ell)(\mathbf{o}_G(\ell - 1) - \mathbf{x}_p(\ell - 1)); \end{aligned} \quad (24)$$

$$\mathbf{x}_p(\ell) = \mathbf{x}_p(\ell - 1) + \mathbf{v}_p(\ell), \quad (25)$$

where  $\omega$  is the inertia coefficient, and  $\phi_1$  ( $\phi_2$ ) is a random variable uniformly distributed in  $[0, c_1]$  ( $[0, c_2]$ ), where  $c_1$  ( $c_2$ ) is named the *acceleration constant*. The inertia coefficients and acceleration constants are parameters to be properly chosen.

### B. PSO-Based Network Planning

As we have seen the ROC AUC well describes the overall behaviour of the ROC and is hence widely recognized as a valid synthetic metric for hypothesis testing. On the other hand, AUC computation is complicated by the need of performing extensive testing, while the CE is immediately provided after the NN training process.

In particular, the testing needed to compute AUC has an additional complexity (with respect to training that must be performed anyway), of

$$\mathcal{C}_{\text{test}} = P(\mathcal{C}_{\text{out}} + \mathcal{C}_{\text{ROC}} + \mathcal{C}_{\text{AUC}}), \quad (26)$$

where  $\mathcal{C}_{\text{out}}$  denotes the complexity associated to running the NN on the test points,  $\mathcal{C}_{\text{ROC}}$  denotes the complexity of building the ROC function, and  $\mathcal{C}_{\text{AUC}}$  denotes the complexity of integrating the ROC. The NN running cost  $\mathcal{C}_{\text{out}}$  is given by the total number of multiplications and additions needed to compute the output value  $y^{(L-1)}$  for all testing vectors, i.e.,

$$\mathcal{C}_{\text{out}} = (2N_{\text{AP}}N_{\text{h}} + 2N_{\text{h}}^2N_{\text{L}} + 2N_{\text{h}}) \tau, \quad (27)$$

where  $N_{\text{h}}$  is the number of neurons in the hidden layer,  $N_{\text{L}}$  is the number of hidden layers, and  $\tau$  is the size of the testing set. The computation of the ROC curve requires the estimation of the  $P_{\text{FA}}$  and  $P_{\text{MD}}$  values for each threshold value  $\lambda$ , whereas the computation of the AUC requires the numerical integration of the ROC curve over  $P_{\text{FA}}$  values.

The proposed PSO-based network planning algorithm is reported in Algorithm 1. We denote as  $\mathcal{B}$  the optimization metric and we initialize  $P$  particles with random positions for each of the  $N_{\text{AP}}$  APs in each particle. For each particle we train the NN and compute  $\mathcal{B}_p^{(0)}$ . The global optimum value  $\mathcal{B}_g$  is set to the minimum among all  $\mathcal{B}_p^{(0)}$  values. Then, positions and velocities of the particles are updated via (24) and (25), and both the local and global optima are updated according to the obtained values at the current iteration. The algorithm stops when the global optimum converges.

Notice that, as the optimization problem is non-convex, PSO is similar to a multi-start optimization with  $P$  different starting points, which is a standard method used to avoid local minima. As  $P$  increases, the probability of finding only a local solution is reduced.

## V. NUMERICAL RESULTS

The considered scenario is depicted in Fig. 1. We consider a region represented by a square with side length 525 m, where four buildings with side length 255 m are located at the map corners, and separated by a road with width 15 m. The ROI  $\mathcal{A}_0$  is located inside the lower-left building, delimited by the dash-dotted line. Roads are considered as LOS paths, whereas transmissions from UEs located in any other map position are in non-LOS conditions.  $N_{\text{AP}} = 5$  APs are deployed, one for each street and one at the map center, to collect attenuation values. Each AP sees the LOS path of the street it is located in. For each AP we generate a shadowing map with standard deviation  $\sigma_s = 8$  dB and decorrelation distance  $d_c = 75$  m. The UE transmits with average unitary power at the frequency  $f_0 = 2.12$  GHz.

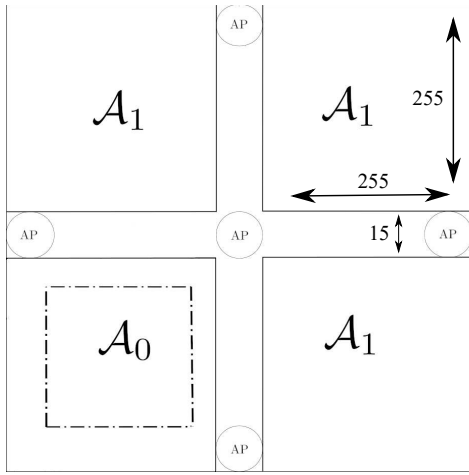


Fig. 1. IRLV deployment scenario.  $N_{\text{AP}} = 5$  APs are located in the streets separating 4 buildings. The ROI is located inside the lower-left building, delimited by the dash-dotted line.

---

### Algorithm 1: Proposed CE-based APs positioning algorithm.

---

**Data:** number of particles  $P$ ,  $N_{\text{AP}}$   
**Result:** optimal position

- 1 Initialize particles;
- 2 train the NN algorithm for each particle;
- 3  $\mathcal{B}_p^{(0)}$ ,  $p = 1, \dots, N_p$ ;
- 4  $\mathcal{B}_g = \min_{p=1, \dots, N_p} \mathcal{B}_p^{(0)}$ ;
- 5  $i = 0$ ;
- 6 **repeat**
- 7      $i = i + 1$ ;
- 8     **for**  $p = 1, \dots, P$  **do**
- 9         update velocity and position vector of particle via (24) and (25);
- 10         train the NN for each particle  $\rightarrow \mathcal{B}_p^{(i)}$ ;
- 11         **if**  $\mathcal{B}_p^{(i)} < \mathcal{B}_g$  **then**
- 12              $\mathcal{B}_g = \mathcal{B}_p^{(i)}$ ;
- 13         **end**
- 14     **end**
- 15 **until** convergence of  $\mathcal{B}_g$ ;

---

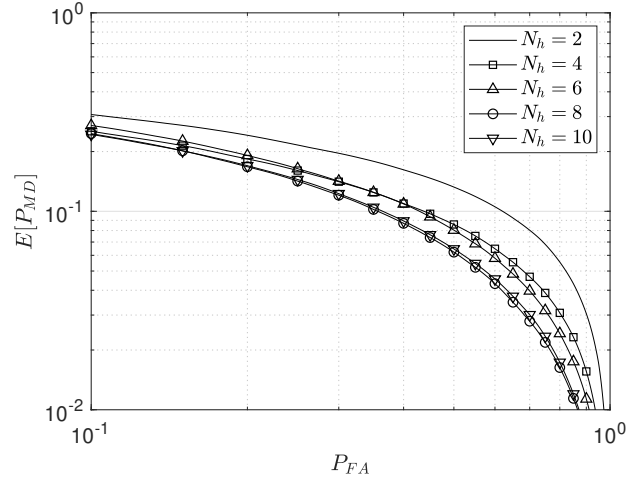


Fig. 2. ROC of the NN IRLV system with different numbers of neurons in the hidden layer  $N_h$ .

Results are averaged over different shadowing realizations. In particular, for each FA probability value we compute the average MD probability over different shadowing maps.

#### A. In-region Location Verification Results

Fig. 2 shows the average (over shadowing realizations)  $P_{\text{MD}}$  vs.  $P_{\text{FA}}$  of the proposed NN IRLV system, with different numbers of neurons in the hidden layer  $N_h$ . Results have been obtained for a NN with  $N_L = 3$  layers and with a training set of size  $S = 10^5$ . We notice that, as the number of neurons at the hidden layer increases, the average FA probability decreases. When the number of neurons  $N_h$  is higher than 8 however we notice that results converge, meaning that increasing the network size does not lead to a performance improvement. Therefore, in the following we set  $N_h = 8$ .

Fig. 3 shows the average (over shadowing realizations)  $P_{\text{MD}}$  vs.  $P_{\text{FA}}$  of the proposed NN IRLV system trained with different numbers of training points  $S$ . Results have been obtained for a NN with  $L = 3$  layers and  $N_h = 8$  neurons in the hidden layer. We see that the AUC decreases when increasing the number of training points and that, starting from  $S = 10^5$ , the ROC does not significantly improve. This is due to the fact that, for the selected NN architecture, training reaches convergence and hence adding further training points does not improve the NN performance.

Fig. 4 shows part of the ROC obtained with the N-P test and with the NN, using the model of Section III-B. In particular, we consider an overall circular region  $\mathcal{A}$  with radius  $R_{\text{out}} = 40$  m, a square authentic region  $\mathcal{A}_0$  of  $L = H = 25$  m located inside  $\mathcal{A}$ , with upper left corner at a distance of  $R_{\text{min}} = 4$  m from the center of  $\mathcal{A}$ . We also report the results of N-P theorem, that can be computed in close form for this simple scenario. We see that, even with a small number of neurons, in this simple problem, the NN achieves the same ROC of the N-P test, thus confirming Theorem 1.

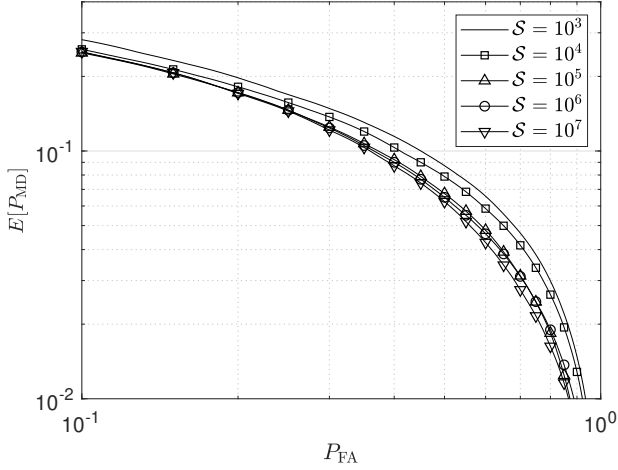


Fig. 3. ROC of the NN IRLV system trained with  $S$  training points.

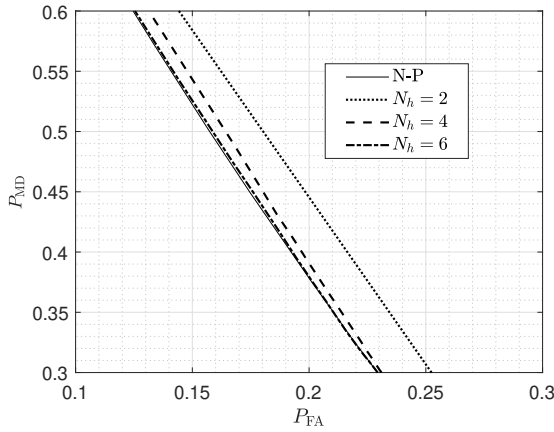


Fig. 4. ROC of the N-P test and the the proposed ML test, with different number of neurons in the hidden layer of the NN.

### B. Network Planning Performance

We now consider the network planning problem, using the proposed Algorithm 1. We consider a PSO with  $P = 6$  particles, each composed by a set of  $N_{AP} = 5$  APs initialized with random positions. There exists a variety of implementations of the PSO, but the most general case for the parameter initialization is given by [13], where it is suggested to set  $\omega = 0.7298$ , and  $c_1 = c_2 = 1.4961$ . Results are averaged over different shadowing realizations. The used NNs are implemented with  $L = 3$  layers and  $N_h = 8$  neurons at the hidden layer. In order to validate the use of the CE as a proxy for the AUC, we compare the performance of Algorithm 1 using either the CE or the AUC as objective function  $\mathcal{B}$ . For a performance assessment we show the AUC obtained with the various optimization approaches.

Fig. 5 shows the average AUC vs. the number of PSO iterations for the two implementations of Algorithm 1, and for different training set size  $S$ . We recall that the result

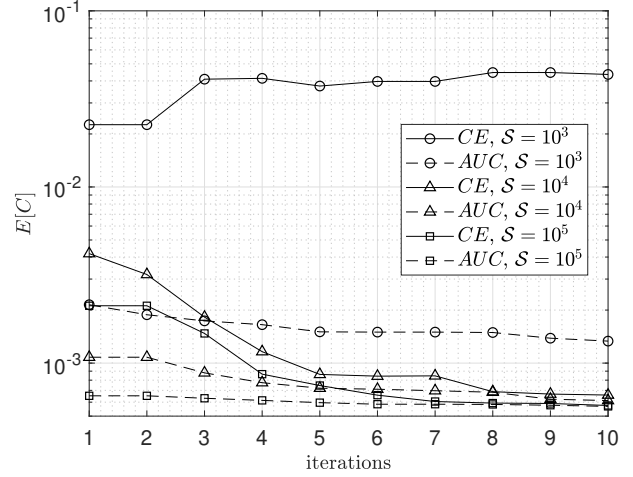


Fig. 5. Mean AUC vs. the number of PSO iterations for Algorithm 1, and two PSO algorithms using only the AUC and the CE as objective functions.

of Theorem 1 holds asymptotically and with perfect training (in terms of training set, number of layers and neurons), which are conditions that our experiments do not satisfy. Two observations are in place here.

First, the training set size should be sufficiently large so that CE becomes a proxy of AUC: in fact we notice that for  $S = 10^3$ , the mean AUC for PSO with CE objective function increases with the number of iterations. This is due to the fact that convergence has not been reached because of the limited size of the training set, and hence the obtained CE value is not representative of the IRLV system performance. This is solved by augmenting the training set size, as we see when  $S = 10^4$  or  $10^5$ .

Second, the AUC-based algorithm converges earlier than the CE-based algorithm. As stated earlier, the NN converges to the optimal N-P solution only asymptotically. Therefore the obtained CE is an approximation of the selected objective function, whose optimization does not necessarily entail the minimization of the AUC. Although requiring a larger number of iterations, the CE-based solution is convenient as it does not require the explicit estimation of the ROC curve.

## VI. CONCLUSIONS

In this paper we formulated the IRLV problem as an hypothesis testing problem and proposed a ML solution. We proved that the NN implementation achieves the same performance of the optimal N-P test with CE as training objective function, and verified numerically this claim for a simple scenario. We also assessed the effects of the training set size over the ROC for a more realistic scenario. We then proposed a PSO algorithm for optimal APs positioning, showing that there is a minimum training set size which allows to use the CE as a proxy of the AUC.

APPENDIX A  
PROOF OF THEOREM 1

Consider

$$\hat{H}_{p_{\mathcal{H}|a}}(g) = - \left[ \frac{n_0}{S} \frac{1}{n_0} \sum_{\mathbf{a} \in \mathcal{T}_0} \log(g(\mathbf{a})) + \frac{n_1}{S} \frac{1}{n_1} \sum_{\mathbf{a} \in \mathcal{T}_1} \log(1 - g(\mathbf{a})) \right], \quad (28)$$

where  $\mathcal{T}_k = \{\mathbf{a}_i \in \mathcal{T} : t_i = k\}$ ,  $|\mathcal{T}_k| = n_k$ ,  $k \in \{0, 1\}$ .

Let  $S \rightarrow \infty$ . By the strong law of large numbers

$$\lim_{S \rightarrow \infty} \hat{H}_{p_{\mathcal{H}|a}}(g) \simeq - \left[ p_{\mathcal{H}}(\mathcal{H}_0) \int_{\mathcal{Y}} p_{\mathbf{a}|\mathcal{H}}(\mathbf{a}|\mathcal{H}_0) \log(g(\mathbf{a})) d\mathbf{a} + p_{\mathcal{H}}(\mathcal{H}_1) \int_{\mathcal{Y}} p_{\mathbf{a}|\mathcal{H}}(\mathbf{a}|\mathcal{H}_1) \log(1 - g(\mathbf{a})) d\mathbf{a} \right], \quad (29)$$

where equality  $\simeq$  holds in probability, i.e., *almost surely*, as per the strong law of large numbers. Rearranging terms with the Bayes rule we get

$$\lim_{S \rightarrow \infty} \hat{H}_{p_{\mathcal{H}|a}}(g) \simeq - \left\{ \int_{\mathcal{Y}} [p_{\mathcal{H}|a}(\mathcal{H}_0|\mathbf{a}) \log g(\mathbf{a}) + (1 - p_{\mathcal{H}|a}(\mathcal{H}_0|\mathbf{a})) \log(1 - g(\mathbf{a}))] p(\mathbf{a}) d\mathbf{a} \right\}. \quad (30)$$

By definition of expected value we can rewrite (30) as

$$\lim_{S \rightarrow \infty} \hat{H}_{p_{\mathcal{H}|a}}(g) \simeq - \mathbb{E}_{\mathbf{a}} [p_{\mathcal{H}|a}(\mathcal{H}_0|\mathbf{a}) \log g(\mathbf{a}) + (1 - p_{\mathcal{H}|a}(\mathcal{H}_0|\mathbf{a})) \log(1 - g(\mathbf{a}))]. \quad (31)$$

We introduce the Bernoulli random variable  $\xi$  with alphabet  $\{0, 1\}$  and PDF

$$p_{\xi}(0) = g(\mathbf{a}), \quad p_{\xi}(1) = 1 - g(\mathbf{a}). \quad (32)$$

Note that  $p_{\xi}$  is a valid PDF since it sums to 1 and  $g(\mathbf{a}) \in [0, 1]$  by hypothesis. Recall now that the cross entropy between two discrete PDFs  $p_{W_1}(w)$  and  $p_{W_2}(w)$  having the same alphabet  $\mathcal{W}$  is defined as

$$H_{p_{W_1}}(W_2) = - \sum_{w \in \mathcal{W}} p_{W_1}(w) \log p_{W_2}(w), \quad (33)$$

and an equivalent definition is

$$H_{p_{W_1}}(W_2) = H(p_{W_1}) + D(p_{W_1}||p_{W_2}), \quad (34)$$

where  $D(\cdot||\cdot)$  is the Kullback-Leibler (K-L) divergence and  $H(\cdot)$  is the entropy function. Then, from (31) and (33), we have

$$\lim_{S \rightarrow \infty} \hat{H}_{p_{\mathcal{H}|a}}(g) \simeq \mathbb{E}_{\mathbf{a}} [H_{p_{\mathcal{H}|a}}(\xi)], \quad (35)$$

which from (34) yields

$$\lim_{S \rightarrow \infty} \hat{H}_{p_{\mathcal{H}|a}}(g) \simeq \mathbb{E}_{\mathbf{a}} [H(p_{\mathcal{H}|a}) + D(p_{\mathcal{H}|a}||p_{\xi})] \quad (36)$$

Recall that NN training is performed by minimizing the left hand side of (36) with respect to NN parameters. In the right hand side of (36) the only quantity depending on NN parameters, through  $g(\mathbf{a})$  in (32), is  $D(p_{\mathcal{H}|a}||p_{\xi})$ . Then, with a infinite number of neurons (i.e., with the possibility of choosing any PDF  $p_{\xi}$ , the minimum of the K-L divergence is attained for  $D(p_{\mathcal{H}|a}||p_{\xi}) = 0$ , that is when  $p_{\xi}(i) = p_{\mathcal{H}|a}(i|\mathbf{a})$ .

REFERENCES

- [1] S. Ceccato, F. Formaggio, G. Caparra, N. Laurenti, and S. Tomasin, "Exploiting side-information for resilient gnss positioning in mobile phones," in *Position, Location and Navigation Symposium (PLANS), 2018 IEEE/ION*, IEEE, 2018, pp. 1515–1524.
- [2] Y. Zeng, J. Cao, J. Hong, S. Zhang, and L. Xie, "Secure localization and location verification in wireless sensor networks: A survey," *The Journal of Supercomputing*, vol. 64, no. 3, pp. 685–701, 2013.
- [3] G. Caparra, M. Centenaro, N. Laurenti, and S. Tomasin, "Optimization of anchor nodes' usage for location verification systems," in *2017 International Conf. on Localization and GNSS (ICL-GNSS)*, Jun. 2017, pp. 1–6.
- [4] E. Jorswieck, S. Tomasin, and A. Sezgin, "Broadcasting into the uncertainty: Authentication and confidentiality by physical-layer processing," *Proc. of the IEEE*, vol. 103, no. 10, pp. 1702–1724, Oct. 2015.
- [5] S. Brands and D. Chaum, "Distance-bounding protocols," in *Workshop on the Theory and Application of Cryptographic Techniques*, Springer, 1993, pp. 344–359.
- [6] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proc. of the 2nd ACM workshop on Wireless security*, ACM, 2003, pp. 1–10.
- [7] A. Vora and M. Nesterenko, "Secure location verification using radio broadcast," *IEEE Trans. on Dependable and Secure Computing*, vol. 3, no. 4, pp. 377–385, Oct. 2006.
- [8] A. Abdou, A. Matrawy, and P. C. van Oorschot, "CPV: Delay-based location verification for the internet," *IEEE Trans. on Dependable and Secure Computing*, vol. 14, no. 2, pp. 130–144, Mar. 2017.
- [9] L. Xiao, X. Wan, and Z. Han, "PHY-layer authentication with multiple landmarks with reduced overhead," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 3, pp. 1676–1687, 2018.
- [10] "LTE; evolved universal terrestrial radio access (E-UTRA); radio frequency (RF) system scenarios," 3GPP, TR 36.942 version 15.0.0 Release 15, Jul 2018.
- [11] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [12] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, Springer, 2011, pp. 760–766.
- [13] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.